Lectures on
Network Systems

Francesco Bullo

With contributions by
Jorge Cortés
Florian Dörfler
Sonia Martínez

Lectures on Network Systems
Francesco Bullo
Version v0.96(a) (Jan 3, 2018).
With contributions by J. Cortés, F. Dörfler, and S. Martínez

The latest version of this document is available at
http://motion.me.ucsb.edu/book-lns/

To Marcello, Gabriella, and Lily

Contents

Preface

Books which try to digest, coordinate, get rid of the duplication, get rid of the less fruitful methods and present the underlying ideas clearly of what we know now, will be the things the future generations will value. Richard Hamming (1915-1998)

Topics   These lecture notes are intended for first-year graduate students interested in network systems, distributed algorithms, and cooperative control. The objective is to answer basic questions such as: What are fundamental dynamical models of interconnected systems? What are the essential dynamical properties of these models and how are they related to network properties? What are basic estimation, control, design, and optimization problems for these dynamical models?

The book is organized in three parts: Linear Systems, Topics in Averaging Systems, and Nonlinear Systems. The Linear Systems part, together with part on the Topics in Averaging Systems, includes

(i) several key motivating examples systems drawn from social, sensor, and compartmental networks, as well as additional ones from robotics,

(ii) basic concepts and results in matrix and graph theory, with an emphasis on Perron–Frobenius theory, algebraic graph theory and linear dynamical systems,

(iii) averaging systems in discrete and continuous time, described by static, time-varying and random matrices, and

(iv) positive and compartmental systems, described by Metzler matrices, with examples from ecology, epidemiology and chemical kinetics.

The Nonlinear Systems part includes

(v) formation control and coordination problems for relative sensing networks,

(vi) networks of phase oscillator systems with an emphasis on the Kuramoto model and models of power networks, and

(vii) virus propagation models, including lumped and network models as well as stochastic and deterministic models, and

(viii) population dynamic models, describing mutualism, competition and cooperation in multi-species systems.

Teaching instructions   These lecture notes are meant to be taught over a quarter-long course with a total 35 to 40 hours of contact time. On average, each chapter should require approximately 2 hours of lecture time. Indeed, these lecture notes are an outgrowth of an introductory graduate course that I taught at UC Santa Barbara over the last several years.

    The intended audience is 1st year graduate students in Engineering, Sciences, and Applied Mathematics programs. For the first part on Linear Systems, the required background includes competency in linear algebra and only very basic notions of dynamical systems. For the second part on Nonlinear Systems (including coupled oscillators and virus propagation), the required background includes a calculus course. The treatment is self-contained and does not require a nonlinear systems course.

    For the benefit of instructors, these lecture notes are supplemented by three documents:

- a solution manual, available upon request by instructors at accredited institutions;

- an abbreviated version of these notes in *slides/landscape format*, especially suited for displaying on a projector for classroom teaching, and

- an abbreviated version of these notes in *classnotes format* (with large sans-serif fonts, small margins), especially suited as markup copy for classroom teaching.

The book, in its three formats, are available for download at: http://motion.me.ucsb.edu/book-lns.

Acknowledgments   I am extremely grateful to Florian Dörfler for his extensive contributions to

  (i) Chapter 17 "Networks of Coupled Oscillators,"

 (ii) Chapter 18 "Robotic Coordination and Formation Control,"

(iii) Sections 5.4, 6.4, 7.6, 8.3 and 10.3, as well as

(iv) a large number of exercises.

I am extremely grateful to Jorge Cortés and Sonia Martínez for their fundamental contribution to my understanding and our joint work on distributed algorithms and robotic networks; their scientific contribution is most obviously present in

  (i) Chapter 1 "Motivating Problems and Systems,"

 (ii) Chapter 3 "Elements of Graph Theory," and

(iii) Chapter 4 "The Adjacency Matrix."

I am extremely grateful to Wenjun Mei, Shadi Mohaghegi and Sandro Zampieri for their extensive contributions to

  (i) Chapter 16 "Virus Propagation in Contact Networks."

I am extremely grateful to Noah Friedkin, Alessandro Giua, and Roberto Tempo for numerous detailed comments and insightful suggestions; their inputs helped shape the numerous chapters, especially the treatment of averaging and social influence networks. I wish to thank Stacy Patterson and Victor Preciado for adopting an early version of these notes and providing me with detailed feedback. I wish to thank Jason Marden and Lucy Pao for their invitation to visit the University of Colorado at Boulder and deliver an early version of these lecture notes.

I wish to thank the maintainers and the contributing authors at openclipart.org and wikipedia.org.

A special thank you goes to all students who took this course and all scientists who read these notes and send me feedback. Particular thanks go to Alex Olshevsky, Ashish Cherukuri, Bala Kameshwar Poolla, Basilio Gentile, Catalin Arghir, Deepti Kannapan, Fabio Pasqualetti, Francesca Parise, John W. Simpson-Porco, Luca Furieri, Marcello Colombino, Paolo Frasca, Pedro Cisneros-Velarde, Peng Jia, Saber Jafarpour, Sepehr Seifi, Shadi Mohagheghi, Tyler Summers, Vaibhav Srivastava, and Xiaoming Duan for their contributions to these lecture notes and related homework.

Finally, I wish to thank Marcello, Gabriella, Lily, and my whole family for their loving support.

Santa Barbara, California, USA   *Francesco Bullo*
29 Mar 2012 — 4 Jan 2018

Part I

Linear Systems

Motivating Problems and Systems

In this chapter, we introduce some example problems and systems from multiple disciplines to motivate our treatment of linear network systems in the following chapters. We look at the following examples:

(i) In the context of social influence networks, we discuss a classic model on how opinions evolve and possibly reach a consensus opinion in groups of individuals.

(ii) In the context of wireless sensor networks, we discuss a simple distributed averaging algorithms and, in the appendix, two advanced design problems for parameter estimation and hypothesis testing.

(iii) In the context of compartmental networks, we discuss dynamical flows among compartments with a classic example for water in desert ecosystems.

(iv) Finally, we discuss simple robotic behaviors for cyclic pursuit and balancing.

In all cases we are interested in presenting the basic models and motivating interest in understanding their dynamic behaviors, such as the existence and attractivity of equilibria.

1.1 Social influence networks: opinion dynamics

We consider a group of n individuals who must act together as a team. Each individual has his own subjective probability density function (pdf) p_i for the unknown value of some parameter (or more simply an estimate of the parameter). We assume now that individual i is appraised of the pdf p_j of each other member $j \neq i$ of the group. Then the model by (French, 1956; Harary, 1959), see also the later (DeGroot, 1974), predicts that the individual will revise its pdf to be:

$$p_i^+ = \sum_{j=1}^n a_{ij} p_j,$$

Figure 1.1: Interactions in a social influence network

where a_{ij} denotes the weight that individual i assigns to the pdf of individual j when carrying out this revision. More precisely, the coefficient a_{ii} describes the attachment of

3

individual i to its own opinion and a_{ij}, $j \neq i$, is an interpersonal influence weight that individual i accords to individual j.

In this model, the coefficients a_{ij} satisfy the following constraints: they are non-negative, that is, $a_{ij} \geq 0$, and, for each individual, the sum of self-weight and accorded weights equals 1, that is, $\sum_{j=1}^{m} a_{ij} = 1$ for all i. In mathematical terms, the matrix

$$A = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{bmatrix}$$

has non-negative entries and each of its rows has unit sum. Such matrices are said to be *row-stochastic*.

Scientific questions of interest include:

   (i) Is this model of human opinion dynamics believable? Is there empirical evidence in its support?

  (ii) How does one measure the coefficients a_{ij}?

 (iii) Under what conditions do the pdfs converge to consensus? What is this value?

 (iv) What are more realistic, empirically-motivated models, possibly including stubborn individuals or antagonistic interactions?

1.2   Wireless sensor networks: averaging algorithms
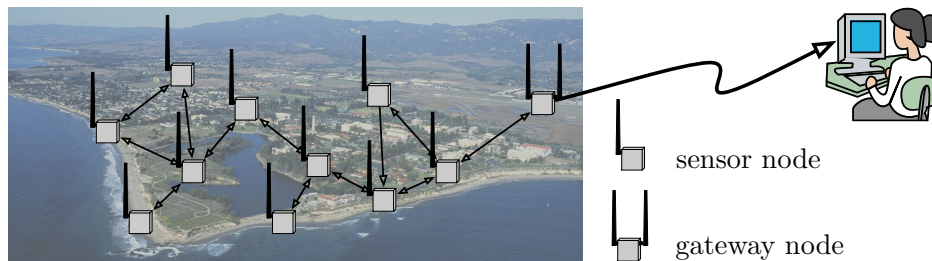
Figure 1.2: A wireless sensor network composed of a collection of spatially-distributed sensors in a field and a gateway node to carry information to an operator. The nodes are meant to measure environmental variables, such as temperature, sound, pressure, and cooperatively filter and transmit the information to an operator.

A wireless sensor network is a collection of spatially-distributed devices capable of measuring physical and environmental variables (e.g., temperature, vibrations, sound, light, etc), performing local computations, and transmitting information to neighboring devices and, in turn, throughout the network (including, possibly, an external operator).

Suppose that each node in a wireless sensor network has measured a scalar environmental quantity, say x_i. Consider the following simple distributed algorithm, based on the concepts of linear averaging: each node repeatedly executes

$$x_i^+ := \text{average}\left(x_i, \{x_j, \text{ for all neighbor nodes } j\}\right), \tag{1.1}$$

where x_i^+ denotes the new value of x_i.   For example, for the graph in Figure 1.3, one

can easily write $x_1^+ := (x_1 + x_2)/2$, $x_2^+ := (x_1 + x_2 + x_3 + x_4)/4$, and so forth. In summary, the algorithm's behavior is described by

$$x^+ = \begin{bmatrix} 1/2 & 1/2 & 0 & 0 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 0 & 1/3 & 1/3 & 1/3 \\ 0 & 1/3 & 1/3 & 1/3 \end{bmatrix} x = A_{\text{wsn}} x,$$
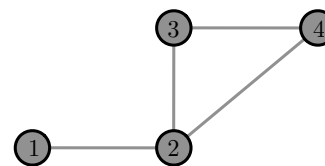
Figure 1.3: Example graph

where the matrix A_{wsn} in equation is again row-stochastic.

Motivated by these examples from social influence networks and wireless sensor networks, we define the *averaging model* to be the dynamical system

$$x(k + 1) = Ax(k), \tag{1.2}$$

where A has non-negative entries and unit row sums. Here, k is the discrete-time variable. We will discuss the continuous-time analogue of this discrete-time model later the book.

Scientific questions of interest for the averaging model include:

(i) Does each node converge to a value? Is this value the same for all nodes?

(ii) Is this value equal to the average of the initial conditions?

(iii) What properties do the graph and the corresponding matrix need to have in order for the algorithm to converge?

(iv) How quick is the convergence?

1.3 Compartmental systems: dynamical flows among compartments

Compartmental systems model dynamical processes characterized by conservation laws (e.g., mass, fluid, energy) and by the flow of material between units known as compartments. For example, the flow of energy and nutrients (water, nitrates, phosphates, etc) in *ecosystems* is typically studied using compartmental modelling; Figure 1.4 illustrates a widely-cited water flow model for a desert ecosystem (Noy-Meir, 1973).
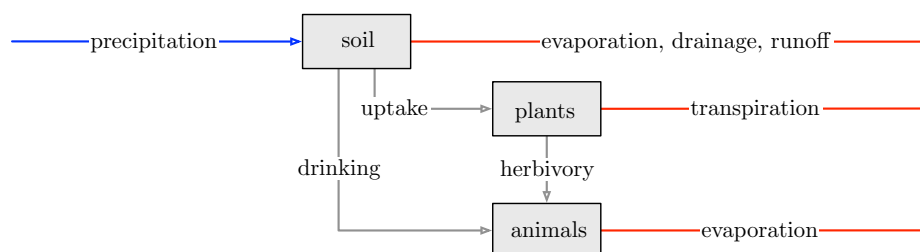
Figure 1.4: Noy-Meir water flow model for a desert ecosystem. The blue line denotes an inflow from the outside environment. The red lines denote outflows into the outside environment.

Given a collection of interconnected compartments, we let q_i denote the amount of material in compartment i, for $i \in \{1, \ldots, n\}$, and write the mass balance equation for the ith compartment as:

$$\dot{q}_i = \sum_{j=1, j \neq i}^{n} (F_{j \to i} - F_{i \to j}) - F_{i \to 0} + u_i, \tag{1.3}$$

where u_i is the inflow from the environment and $F_{i \to 0}$ is the outflow into the environment. We refer to equation (1.3) as a *compartmental system*. We next assume linear flows, that is, we assume that the flow $F_{i \to j}$ from node i to node j (as well as to the environment) is proportional to the mass quantity at i, that is, $F_{i \to j} = f_{ij} q_i$ for a positive *flow rate* constant f_{ij}. Therefore we can write the dynamics of a *linear compartmental system* as

$$\dot{q}_i(t) = \sum_{j=1, j \neq i}^{n} (f_{ji} q_j(t) - f_{ij} q_i(t)) - f_{i0} q_i(t) + u_i. \tag{1.4}$$

Here, t is the continuous-time variable. Equivalently, in vector notation, for an appropriate *compartmental matrix* C,

$$\dot{q}(t) = C q(t) + u. \tag{1.5}$$

For example, let us write down the compartmental matrix C for the water flow model in figure. We let q_1, q_2, q_3 denote the water mass in soil, plants and animals, respectively. Moreover, as in figure, we let $f_{\text{e-d-r}}, f_{\text{trnsp}}, f_{\text{evap}}, f_{\text{drnk}}, f_{\text{uptk}}, f_{\text{herb}}$, denote respectively the evaporation-drainage-runoff, transpiration, evaporation, drinking, uptake, and herbivory rate. With these notations, we can write

$$C = \begin{bmatrix} -f_{\text{e-d-r}} - f_{\text{uptk}} - f_{\text{drnk}} & 0 & 0 \\ f_{\text{uptk}} & -f_{\text{trnsp}} - f_{\text{herb}} & 0 \\ f_{\text{drnk}} & f_{\text{herb}} & -f_{\text{evap}} \end{bmatrix}.$$

Scientific questions of interest include:

(i) for constant inflows u, does the total mass in the system remain bounded?

(ii) is there an equilibrium solution? do all evolutions converge to it?

(iii) which compartments become empty asymptotically?

1.4   Appendix: Robotic networks in cyclic pursuit and balancing

In this section we consider two simple examples of coordination motion in robotic networks. The standing assumption is that n robots, amicably referred to as "bugs," are placed and restricted to move on a circle of unit radius. Because of this bio-inspiration and because this language is common in the literature (Klamkin and Newman, 1971; Bruckstein et al., 1991), we refer to the following two problems as n-*bugs problems*.

On this unit circle the bugs' positions are angles measured counterclockwise from the positive horizontal axis. We let angles take value in $[0, 2\pi)$, that is, an arbitrary position θ satisfies $0 \leq \theta < 2\pi$. The bugs are numbered counterclockwise with identities $i \in \{1, \ldots, n\}$ and are at positions $\theta_1, \ldots, \theta_n$. It is convenient

to identify $n + 1$ with 1. We assume the bugs move in discrete times k in a counterclockwise direction by a controllable amount u_i (i.e., a control signal), that is:

$$\theta_i(k + 1) = \text{mod}(\theta_i(k) + u_i(k),\, 2\pi).$$

where $\text{mod}(\vartheta,\, 2\pi)$ is the remainder of the division of ϑ by 2π and its introduction is required to ensure that $\theta_i(k + 1)$ remains inside $[0, 2\pi)$.

Objective: optimal patrolling of a perimeter. Approach: Cyclic pursuit

We now suppose that each bug feels an attraction and moves towards the closest counterclockwise neighbor, as illustrated in Figure 1.5. Recall that the *counterclockwise distance from θ_i and θ_{i+1}* is the length of the counterclockwise arc from θ_i and θ_{i+1} and satisfies:

$$\text{dist}_{\mathsf{cc}}(\theta_i, \theta_{i+1}) = \text{mod}(\theta_{i+1} - \theta_i,\, 2\pi),$$

In short, given a *control gain* $\kappa \in [0, 1]$, we assume that the ith bug sets its control signal to

$$u_{\text{pursuit},i}(k) = \kappa\, \text{dist}_{\mathsf{cc}}(\theta_i(k), \theta_{i+1}(k)).$$

(a) Cyclic pursuit control law

(b) Cyclic balancing control law
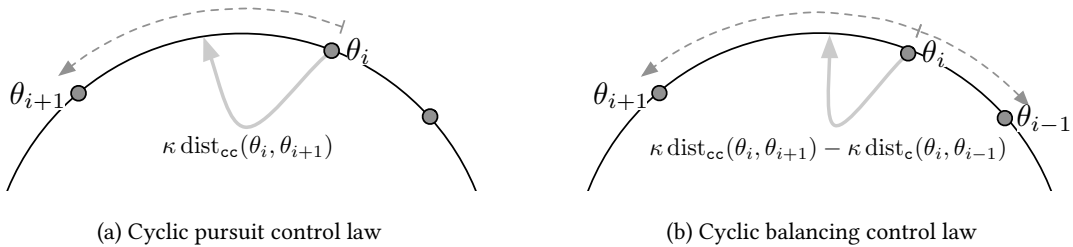
Figure 1.5: Cyclic pursuit and balancing are prototypical n-bug problems.

Scientific questions of interest include:
- (i) Does this system have any equilibrium?
- (ii) Is a rotating equally-spaced configuration a solution? An equally-spaced angle configuration is one for which $\text{mod}(\theta_{i+1} - \theta_i,\, 2\pi) = \text{mod}(\theta_i - \theta_{i-1},\, 2\pi)$ for all $i \in \{1, \ldots, n\}$. Such configurations are sometimes called *splay states*.
- (iii) For which values of κ do the bugs converge to an equally-spaced configuration and with what pairwise distance?

Objective: optimal sensor placement. Approach: Cyclic balancing

Next, we suppose that each bug feels an attraction towards both the closest counterclockwise and the closest clockwise neighbor, as illustrated in Figure 1.5. Given a "control gain" $\kappa \in [0, 1/2]$ and the natural notion of clockwise distance, the ith bug sets its control signal to

$$u_{\text{balancing},i}(k) = \kappa\, \text{dist}_{\mathsf{cc}}(\theta_i(k), \theta_{i+1}(k)) - \kappa\, \text{dist}_{\mathsf{c}}(\theta_i(k), \theta_{i-1}(k)),$$

where $\mathrm{dist}_{\mathsf{c}}(\theta_i(k), \theta_{i-1}(k)) = \mathrm{dist}_{\mathsf{cc}}(\theta_{i-1}(k), \theta_i(k))$.

Questions of interest are:

(i) Is a static equally-spaced configuration a solution?

(ii) For which values of κ do the bugs converge to a static equally-spaced configuration?

(iii) Is it true that the bugs will approach an equally-spaced configuration and that each of them will converge to a stationary position on the circle?

A preliminary analysis

It is unrealistic (among other aspects of this setup) to assume that the bugs know the absolute position of themselves and of their neighbors. Therefore, it is interesting to rewrite the dynamical system in terms of pairwise distances between nearby bugs.

For $i \in \{1, \dots, n\}$, we define the relative angular distances (the lengths of the counterclockwise arcs) $d_i = \mathrm{dist}_{\mathsf{cc}}(\theta_i, \theta_{i+1}) \geq 0$. (We also adopt the usual convention that $d_{n+1} = d_1$ and that $d_0 = d_n$). The change of coordinates from $(\theta_1, \dots, \theta_n)$ to $(d_1, \dots, d_n)$ leads us to rewrite the cyclic pursuit and the cyclic balancing laws as:

$$u_{\mathrm{pursuit},i}(k) = \kappa d_i,$$
$$u_{\mathrm{balancing},i}(k) = \kappa d_i - \kappa d_{i-1}.$$

In this new set of coordinates, one can show that the cyclic pursuit and cyclic balancing systems are, respectively,

$$d_i(k+1) = (1 - \kappa)d_i(k) + \kappa d_{i+1}(k), \tag{1.6}$$
$$d_i(k+1) = \kappa d_{i+1}(k) + (1 - 2\kappa)d_i(k) + \kappa d_{i-1}(k). \tag{1.7}$$

These are two linear time-invariant dynamical systems with state $d = (d_1, \dots, d_n)$ and governing equation described by the two $n \times n$ matrices:

$$
A_{\mathrm{pursuit}} = \begin{bmatrix}
1 - \kappa & \kappa & \cdots & 0 & 0 \\
0 & 1 - \kappa & \ddots & \ddots & 0 \\
\vdots & \ddots & \ddots & \ddots & 0 \\
0 & \ddots & \ddots & 1 - \kappa & \kappa \\
\kappa & 0 & \cdots & 0 & 1 - \kappa
\end{bmatrix}, \quad
A_{\mathrm{balancing}} = \begin{bmatrix}
1 - 2\kappa & \kappa & \cdots & 0 & \kappa \\
\kappa & 1 - 2\kappa & \ddots & \ddots & 0 \\
\vdots & \ddots & \ddots & \ddots & 0 \\
0 & \ddots & \ddots & 1 - 2\kappa & \kappa \\
\kappa & 0 & \cdots & \kappa & 1 - 2\kappa
\end{bmatrix}.
$$

We conclude with the following remarks.

(i) Equations (1.6) and (1.7) are correct if the counterclockwise order of the bugs is never violated. One can show that this is true for $\kappa < 1$ in the pursuit case and $\kappa < 1/2$ in the balancing case; we leave this proof to the reader in Exercise E1.4.

(ii) The matrices A_{pursuit} and $A_{\mathrm{balancing}}$, for varying n and κ, are called Toeplitz and circulant based on the nonzero/zero patterns of their entries; we study the properties of such matrices in later chapters. Moreover, they have non-negative entries for the stated ranges of κ and are row-stochastic.

(iii) If one defines the agreement space, i.e., $\{(\alpha, \alpha, \ldots, \alpha) \in \mathbb{R}^n \mid \alpha \in \mathbb{R}\}$, then each point in this set is an equilibrium for both systems.

(iv) It must be true for all times that $(d_1, \ldots, d_n) \in \{x \in \mathbb{R}^n \mid x_i \geq 0, \sum_{i=1}^n x_i = 2\pi\}$. This property is indeed the consequence of the non-negative matrices A_{pursuit} and $A_{\text{balancing}}$ being *doubly-stochastic*, i.e., each row-sum and each column-sum is equal to 1.

(v) We will later study for which values of κ the system converges to the agreement space.

1.5 Appendix: Design problems in wireless sensor networks

In this appendix we show how averaging algorithms are relevant in wireless sensor network problems and can be used to tackle more sophisticated than what shown in Section 1.2.

1.5.1 Wireless sensor networks: distributed parameter estimation

The next two examples are also drawn from the field of wireless sensor network, but they feature a more advanced setup and require a basic background in estimation and detection theory, respectively. The key lessons to be learnt from these examples is that it is useful to have algorithms that compute the average of distributed quantities.

Following ideas from (Xiao et al., 2005; Garin and Schenato, 2010), we aim to estimate an unknown parameter $\theta \in \mathbb{R}^m$ via the measurements taken by a sensor network. Each node $i \in \{1, \ldots, n\}$ measures

$$y_i = B_i \theta + v_i,$$

where $y_i \in \mathbb{R}^{m_i}$, B_i is a known matrix and v_i is random measurement noise. We assume that

(A1) the noise vectors $v_1, \ldots, v_n$ are independent jointly-Gaussian variables with zero-mean $\mathbb{E}[v_i] = \mathbb{0}_{m_i}$ and positive-definite covariance $\mathbb{E}[v_i v_i^\mathsf{T}] = \Sigma_i = \Sigma_i^\mathsf{T}$, for $i \in \{1, \ldots, n\}$; and

(A2) the measurement parameters satisfy the following two properties: $\sum_i m_i \geq m$ and $\begin{bmatrix} B_1 \\ \vdots \\ B_n \end{bmatrix}$ is full rank.

Given the measurements $y_1, \ldots, y_n$, it is of interest to compute a least-square estimate of θ, that is, an estimate of θ that minimizes a least-square error. Specifically, we aim to minimize the following *weighted least-square error*:

$$\min_{\widehat{\theta}} \sum_{i=1}^n \left\| y_i - B_i \widehat{\theta} \right\|_{\Sigma_i^{-1}}^2 = \sum_{i=1}^n \left(y_i - B_i \widehat{\theta} \right)^\mathsf{T} \Sigma_i^{-1} \left(y_i - B_i \widehat{\theta} \right).$$

In this weighted least-square error, individual errors are weighted by their corresponding inverse covariance matrices so that an accurate (respectively, inaccurate) measurement corresponds to a high (respectively, low) error weight. With this particular choice of weights, the least-square estimate coincides with the

so-called maximum-likelihood estimate; see (Poor, 1998) for more details. Under assumptions (A1) and (A2), the optimal solution is

$$\widehat{\theta}^* = \Big(\sum_{i=1}^{n} B_i^\mathsf{T} \Sigma_i^{-1} B_i \Big)^{-1} \sum_{i=1}^{n} B_i^\mathsf{T} \Sigma_i^{-1} y_i.$$

This formula is easy to implement by a single processor with all the information about the problem, i.e., the parameters and the measurements.

To compute $\widehat{\theta}^*$ in the sensor (and processor) network, we perform two steps:
[Step 1:] we run two distributed algorithms in parallel to compute the average of the quantities $B_i^\mathsf{T} \Sigma_i^{-1} B_i$ and $B_i^\mathsf{T} \Sigma_i^{-1} y_i$.
[Step 2:] we compute the optimal estimate via

$$\widehat{\theta}^* = \text{average} \left(B_1^\mathsf{T} \Sigma_1^{-1} B_1, \dots, B_n^\mathsf{T} \Sigma_n^{-1} B_n \right)^{-1} \text{average} \left(B_1^\mathsf{T} \Sigma_1^{-1} y_1, \dots, B_n^\mathsf{T} \Sigma_n^{-1} y_n \right).$$

Questions of interest are:

(i) How do we design algorithms to compute the average of distributed quantities?

(ii) What properties does the graph need to have in order for such an algorithm to exist?

(iii) How do we design an algorithm with fastest convergence?

1.5.2   Wireless sensor networks: distributed hypothesis testing

We consider a distributed hypothesis testing problem; these ideas appeared in (Rao and Durrant-Whyte, 1993; Olfati-Saber et al., 2006). Let h_γ, for $\gamma \in \Gamma$ in a finite set Γ, be a set of two or more hypotheses about an uncertain event. For example, given a certain area of interest, we could have $h_0 = $ "no target is present", $h_1 = $ "one target is present" and $h_2 = $ "two or more targets are present".

Suppose that we know the *a priori probabilities* $p(h_\gamma)$ of the hypotheses and that n nodes of a sensor network take measurements y_i, for $i \in \{1, \dots, n\}$, related to the event. Independently of the type of measurements, assume you can compute

$$p(y_i | h_\gamma) = \text{probability of measuring } y_i \text{ given that } h_\gamma \text{ is the true hypothesis.}$$

Also, assume that each observation is conditionally independent of all other observations, given any hypothesis.

(i) We wish to compute the *maximum a posteriori estimate*, that is, we want to identify which one is the most likely hypothesis, given the measurements. Note that, under the independence assumption, Bayes' Theorem implies that the *a posteriori probabilities* satisfy

$$p(h_\gamma | y_1, \dots, y_n) = \frac{p(h_\gamma)}{p(y_1, \dots, y_n)} \prod_{i=1}^{n} p(y_i | h_\gamma).$$

(ii) Observe that $p(h_\gamma)$ is known, and $p(y_1, \ldots, y_n)$ is a constant normalization factor scaling all posteriori probabilities equally. Therefore, for each hypothesis $\gamma \in \Gamma$, we need to compute

$$\prod_{i=1}^{n} p(y_i|h_\gamma),$$

or equivalently, we aim to exchange data among the sensors in order to compute:

$$\exp\left(\sum_{i=1}^{n}\log(p(y_i|h_\gamma))\right) = \exp\left(n \operatorname{average}\left(\log p(y_1|h_\gamma), \ldots, \log p(y_n|h_\gamma)\right).\right)$$

(iii) In summary, even in this hypothesis testing problem, we need algorithms to compute the average of the n numbers $\log p(y_1|h_\gamma), \ldots, \log p(y_n|h_\gamma)$, for each hypothesis γ.

Questions of interest here are the same as in the previous section.

1.6 Historical notes and further reading

Numerous other examples of multi-agent systems and applications can be found in the recent texts (Ren and Beard, 2008; Bullo et al., 2009; Mesbahi and Egerstedt, 2010; Bai et al., 2011; Cristiani et al., 2014; Fuhrmann and Helmke, 2015; Francis and Maggiore, 2016; Arcak et al., 2016; Porter and Gleeson, 2016). Other, related, and instructive examples are presented in recent surveys such as (Martínez et al., 2007; Ren et al., 2007; Garin and Schenato, 2010; Cao et al., 2013; Oh et al., 2015). Textbooks, monographs and surveys on the broader and different theme of network science include (Newman, 2003; Boccaletti et al., 2006; Castellano et al., 2009; Easley and Kleinberg, 2010; Jackson, 2010; Newman, 2010; Spielman, 2017).

The opinion dynamics example in Section 1.1 is an illustration of the rich literature on social influence networks, starting with the early works by French (1956), Harary (1959), Abelson (1964), and DeGroot (1974). While the linear averaging model is by now known as the DeGroot model, the key ideas were already present in French (1956) and the main results (e.g., average consensus for doubly stochastic matrices) were already obtained by (Harary, 1959). Empirical evidence in support of the averaging model (including its variations) is described in (Friedkin and Johnsen, 2011; Chandrasekhar et al., 2015; Friedkin et al., 2016). An outstanding tutorial and survey on dynamic social networks is (Proskurnikov and Tempo, 2017). We postpone to Chapter 9 the literature review on compartmental systems.

The n-bugs problem is related to the study of pursuit curves and inquires about what the paths of n bugs are when they chase one another. We refer to (Klamkin and Newman, 1971; Watton and Kydon, 1969; Bruckstein et al., 1991; Marshall et al., 2004; Smith et al., 2005) for some classic works, surveys, and recent results.

1.7 Exercises

E1.1 **Bounded evolution for averaging systems.** Given a matrix $A \in \mathbb{R}^{n \times n}$ with non-negative entries and unit row sums, consider the *averaging model* (1.2)

$$x(k+1) = Ax(k).$$

Show that, for all initial conditions $x(0)$ and times k,

$$\min_i x_i(0) \leq \min_i x_i(k) \leq \min_i x_i(k+1) \leq \max_i x_i(k+1) \leq \max_i x_i(k) \leq \max_i x_i(0),$$

where i takes values in $\{1, \ldots, n\}$.

E1.2 **Conservation of mass for compartmental systems.** Given n compartments and flows among them, consider the compartmental system (1.3)

$$\dot{q}_i = \sum\nolimits_{j=1, j \neq i}^{n} \left(F_{j \to i} - F_{i \to j} \right) - F_{i \to 0} + u_i,$$

and its linear version in equation (1.5): $\dot{q} = Cq + u$. Do the following tasks:

   (i) show that, if there are no inflows, i.e., if $u_i = 0$ for all i, then the total mass in the compartmental system does not increase with time,

   (ii) write a formula for the diagonal and off-diagonal entries of the compartmental matrix C as a function of the flow rate constants, and

   (iii) show that the column sums of C are non-positive.

E1.3 **Simulating the averaging dynamics.** Simulate in your favorite programming language and software package the linear averaging algorithm in equation (1.1). Set $n = 5$, select the initial state equal to $(1, -1, 1, -1, 1)$, and use the following undirected unweighted graphs (depicted in figure):

   (i) the complete graph,

   (ii) the cycle graph, and

   (iii) the star graph with node 1 as center.

Which value do all nodes converge to? Is it equal to the average of the initial values? Turn in your code, a few printouts (as few as possible), and your written responses.
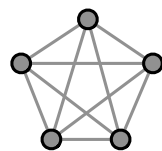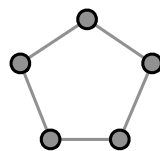
Complete graph   Cycle graph   Star graph
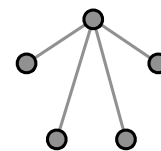
E1.4 **Computing the bugs' dynamics.** Consider the cyclic pursuit and balancing dynamics described in Section 1.4. Verify

   (i) the cyclic pursuit closed-loop equation (1.6),

   (ii) the cyclic balancing closed-loop equation (1.7), and

   (iii) the counterclockwise order of the bugs is never violated.

   Hint: *Recall the distributive property of modular addition:* $\mathrm{mod}(a \pm b, n) = \mathrm{mod}(\mathrm{mod}(a, n) \pm \mathrm{mod}(b, n), n)$.

E1.5 **Continuous-time cyclic pursuit on the plane.** Consider four mobile robots on a plane with positions $p_i \in \mathbb{C}$, $i \in \{1, \ldots, 4\}$, and moving according to $\dot{p}_i = u_i$, where $u_i \in \mathbb{C}$ are the velocity commands. The task of the robots is rendezvous at a common point (while using only onboard sensors). A simple strategy to achieve rendezvous is *cyclic pursuit*: each robot i picks another robot, say $i + 1$, and pursues it. (Here $4 + 1 = 1$.) In other words, we set $u_i = p_{i+1} - p_i$ and obtain the closed-loop system (see also corresponding simulation below):

$$\begin{bmatrix} \dot{p}_1 \\ \dot{p}_2 \\ \dot{p}_3 \\ \dot{p}_4 \end{bmatrix} = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \\ 1 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \end{bmatrix}.$$
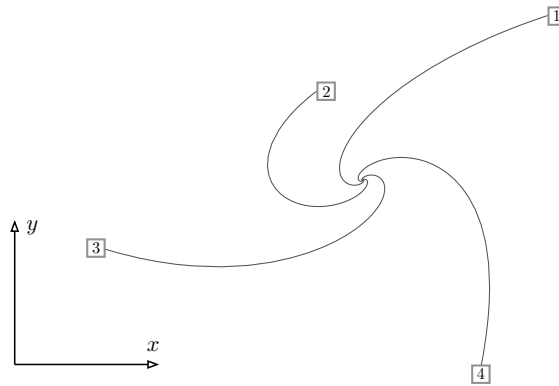
Prove that:

(i) the average robot position $\mathrm{average}(p(t)) = \sum_{i=1}^{4} p_i(t)/4$ remains constant for all $t \geq 0$;

(ii) the robots asymptotically rendezvous at the initial average robot position mass, that is,

$$\lim_{t \to \infty} p_i(t) = \mathrm{average}(p(0)) \quad \text{for } i \in \{1, \ldots, 4\};$$

(iii) if the robots are initially arranged in a square formation, then they remain in a square formation.

Hint: *Given a matrix A with semisimple eigenvalues, the solution to $\dot{x} = Ax$ is given by the modal expansion $x(t) = \sum_{i=1}^{n} e^{\lambda_i t} v_i w_i^{\mathsf{T}} x(0)$, where v_i and w_i are the right and left eigenvectors associated to the eigenvalue λ_i and normalized to $w_i^{\mathsf{T}} v_i = 1$. The modal decomposition will be reviewed in Sections 2.1 and 10.1.*

Elements of Matrix Theory

In this chapter we review basic concepts from matrix theory with a special emphasis on the so-called Perron-Frobenius theory. These concepts will be useful when analyzing the convergence of the linear dynamical systems discussed in Chapter 1, graphs and averaging algorithms defined over graphs.

Notation

It is useful to start with some basic notations from matrix theory and linear algebra. We let $f : X \to Y$ denote a function from set X to set Y. We let $\mathbb{R}$, $\mathbb{N}$ and $\mathbb{Z}$ denote respectively the set of real, natural and integer numbers; also $\mathbb{R}_{\geq 0}$ and $\mathbb{Z}_{\geq 0}$ are the set of non-negative real numbers and non-negative integer numbers. For real numbers $a < b$, we let

$$[a, b] = \{x \in \mathbb{R} \mid a \leq x \leq b\}, \qquad]a, b] = \{x \in \mathbb{R} \mid a < x \leq b\},$$
$$[a, b[= \{x \in \mathbb{R} \mid a \leq x < b\}, \qquad]a, b[= \{x \in \mathbb{R} \mid a < x < b\}.$$

Given a complex number $z \in \mathbb{C}$, its norm (sometimes referred to as complex modulus) is denoted by $|z|$, its real part by $\Re(z)$ and its imaginary part by $\Im(z)$. We let i denote the imaginary unit $\sqrt{-1}$.

We let $\mathbb{1}_n \in \mathbb{R}^n$ (respectively $\mathbb{0}_n \in \mathbb{R}^n$) be the column vector with all entries equal to $+1$ (respectively 0). Let $\mathbb{e}_1, \ldots, \mathbb{e}_n$ be the standard basis vectors of $\mathbb{R}^n$, that is, $\mathbb{e}_i$ has all entries equal to zero except for the ith entry equal to 1. The 1-norm, 2-norm, and ∞-norm of a vector $x \in \mathbb{R}^n$ are defined by, respectively,

$$\|x\|_1 = |x_1| + \cdots + |x_n|, \qquad \|x\|_2 = \sqrt{x_1^2 + \cdots + x_n^2}, \qquad \|x\|_\infty = \max\{|x_1|, \ldots, |x_n|\}.$$

We let I_n denote the n-dimensional identity matrix and $A \in \mathbb{R}^{n \times n}$ denote a square $n \times n$ matrix with real entries $\{a_{ij}\}$, $i, j \in \{1, \ldots, n\}$. The matrix A is *symmetric* if $A^\mathsf{T} = A$.

For a matrix A, $\lambda \in \mathbb{C}$ is an *eigenvalue* and $v \in \mathbb{C}^n$ is a *right eigenvector*, or simply an *eigenvector*, if they together satisfy the eigenvalue equation $Av = \lambda v$. Sometimes it will be convenient to refer to (λ, v) as an *eigenpair*. A *left eigenvector* of the eigenvalue λ is a vector $w \in \mathbb{C}^n$ satisfying $w^\mathsf{T} A = \lambda w^\mathsf{T}$.

A symmetric matrix is *positive definite* (resp. positive semidefinite) if all its eigenvalues are positive (resp. non-negative). The *kernel* of A is the subspace $\mathrm{kernel}(A) = \{x \in \mathbb{R}^n \mid Ax = \mathbb{0}_n\}$, the *image* of A is $\mathrm{image}(A) = \{y \in \mathbb{R}^n \mid Ax = y, \text{ for some } x \in \mathbb{R}^n\}$, and the *rank* of A is the dimension of its image. Given vectors $v_1, \ldots, v_j \in \mathbb{R}^n$, their span is $\mathrm{span}(v_1, \ldots, v_j) = \{a_1 v_1 + \cdots + a_j v_j \mid a_1, \ldots, a_j \in \mathbb{R}\} \subset \mathbb{R}^n$.

2.1 Linear systems and the Jordan normal form

In this section we introduce a prototypical model for dynamical systems and study its stabilities properties via the so-called Jordan normal form, that is a key tool from matrix theory. We will later apply these results to the averaging model (1.2).

2.1.1 Discrete-time linear systems

We start with a basic definition.

Definition 2.1 (Discrete-time linear system). *A square matrix A defines a* discrete-time linear system *by*

$$x(k + 1) = Ax(k), \quad x(0) = x_0, \tag{2.1}$$

or, equivalently by $x(k) = A^k x_0$, where the sequence $\{x(k)\}_{k \in \mathbb{Z}_{\geq 0}}$ is called the solution, trajectory *or* evolution *of the system.*

Sometimes it is convenient to adopt the shorthand $x^+ = f(x)$ to denote the system $x(k+1) = f(x(k))$.

We are interested in understanding when a solution from an arbitrary initial condition has an asymptotic limit as time diverges and to what value the solution converges. We formally define this property as follows.

Definition 2.2 (Semi-convergent and convergent matrices). *A matrix $A \in \mathbb{R}^{n \times n}$ is*

   (i) semi-convergent *if $\lim_{k \to +\infty} A^k$ exists, and*

   (ii) convergent *if it is semi-convergent and $\lim_{k \to +\infty} A^k = \mathbb{0}_{n \times n}$.*

It is immediate to see that, if A is semi-convergent with limiting matrix $A_\infty = \lim_{k \to +\infty} A^k$, then

$$\lim_{k \to +\infty} x(k) = A_\infty x_0.$$

In what follows we characterize the sets of semi-convergent and convergent matrices.

Remark 2.3 (Modal decomposition for symmetric matrices). *Before treating the general analysis method, we present the self-contained and instructive case of symmetric matrices. Recall that a symmetric matrix A has real eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ and corresponding orthonormal (i.e., orthogonal and unit-length) eigenvectors $v_1, \ldots, v_n$. Because the eigenvectors are an orthonormal basis for $\mathbb{R}^n$, we can write the* modal decomposition

$$x(k) = y_1(k)v_1 + \cdots + y_n(k)v_n,$$

where the ith normal mode *is defined by $y_i(k) = v_i^\mathsf{T} x(k)$. We then left-multiply the two equalities (2.1) by v_i^T and exploit $Av_i = \lambda_i v_i$ to obtain*

$$y_i(k + 1) = \lambda_i y_i(k), \quad y_i(0) = v_i^\mathsf{T} x_0, \quad \implies \quad y_i(k) = \lambda_i^k (v_i^\mathsf{T} x_0).$$

In short, the evolution of the linear system (2.1) is

$$x(k) = \lambda_1^k (v_1^\mathsf{T} x_0) v_1 + \cdots + \lambda_n^k (v_n^\mathsf{T} x_0) v_n.$$

Therefore, each evolution starting from an arbitrary initial condition satisfies

(i) $\lim_{k\to\infty} x(k) = \mathbb{0}_n$ *if and only if* $|\lambda_i| < 1$ *for all* $i \in \{1, \ldots, n\}$, *and*

(ii) $\lim_{k\to\infty} x(k) = (v_1^\mathsf{T} x_0)v_1 + \cdots + (v_m^\mathsf{T} x_0)v_m$ *if and only if* $\lambda_1 = \cdots = \lambda_m = 1$ *and* $|\lambda_i| < 1$ *for all* $i \in \{m+1, \ldots, n\}$.

●

2.1.2 The Jordan normal form

In this section we review a very useful canonical decomposition of a square matrix. Recall that two $n \times n$ matrices A and B are *similar* if $B = TAT^{-1}$ for some invertible matrix T. Also recall that a similarity transform does not change the eigenvalues of a matrix.

Theorem 2.4 (Jordan normal form). *Each matrix* $A \in \mathbb{C}^{n\times n}$ *is similar to a block diagonal matrix* $J \in \mathbb{C}^{n\times n}$, *called the* Jordan normal form *of A, given by*

$$
J = \begin{bmatrix} J_1 & 0 & \cdots & 0 \\ 0 & J_2 & \ddots & 0 \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & J_m \end{bmatrix} \in \mathbb{C}^{n\times n},
$$

where each block J_i, *called a* Jordan block, *is a square matrix of size* j_i *and of the form*

$$
J_i = \begin{bmatrix} \lambda_i & 1 & \cdots & 0 \\ 0 & \lambda_i & \ddots & 0 \\ \vdots & \ddots & \ddots & 1 \\ 0 & \cdots & 0 & \lambda_i \end{bmatrix} \in \mathbb{C}^{j_i \times j_i}. \tag{2.2}
$$

Clearly, $m \leq n$ *and* $j_1 + \cdots + j_m = n$.

We refer to (Horn and Johnson, 1985) for a standard proof of this theorem. Note that the matrix J is unique, modulo a re-ordering of the Jordan blocks.

Regarding the eigenvalues of A, we note the following. The eigenvalues of J, and therefore also of A, are the (not necessarily distinct) numbers $\lambda_1, \ldots, \lambda_m$. Given an eigenvalue λ,

(i) the *algebraic multiplicity* of λ is the sum of the sizes of all Jordan blocks with eigenvalue λ (or, equivalently, the multiplicity of λ as a root of the characteristic polynomial of A), and

(ii) the *geometric multiplicity* of λ is the number of Jordan blocks with eigenvalue λ (or, equivalently, the number of linearly-independent eigenvectors associated to λ).

An eigenvalue is

(i) *simple* if it has algebraic and geometric multiplicity equal precisely to 1, that is, a single Jordan block of size 1, and

(ii) *semisimple* if all its Jordan blocks have size 1, so that its algebraic and geometric multiplicity are equal.

Regarding the eigenvectors of A, Theorem 2.4 implies there exists an invertible matrix T such that

$$A = TJT^{-1} \tag{2.3}$$
$$\iff AT = TJ \tag{2.4}$$
$$\iff T^{-1}A = JT^{-1}. \tag{2.5}$$

Let $t_1, \ldots, t_n$ and $r_1, \ldots, r_n$ denote the columns and rows of T and T^{-1} respectively. If all eigenvalues of A are semisimple, then the equations (2.4) and (2.5) imply, for all $i \in \{1, \ldots, n\}$,

$$At_i = \lambda_i t_i \quad \text{and} \quad r_i A = \lambda_i r_i.$$

In other words, the ith column of T is the *right eigenvector* (or simply *eigenvector*) of A corresponding to the eigenvalue λ_i, and the ith row of T^{-1} is the corresponding *left eigenvector* of A.

If an eigenvalue is not semisimple, then it has larger algebraic than geometric multiplicity. For such eigenvalues, the columns of the matrix T are the right eigenvectors and the *generalized right eigenvectors* of A, whereas the rows of T^{-1} are the left eigenvectors and the *generalized left eigenvector* of A. For more details about generalized eigenvectors, we refer to reader to (Horn and Johnson, 1985).

Example 2.5 (Revisiting the wireless sensor network example). As a numerical example, let us reconsider the wireless sensor network discussed in Section 1.2 and the 4-dimensional row-stochastic matrix A_{wsn}, which we report here for convenience:

$$A_{\text{wsn}} = \begin{bmatrix} 1/2 & 1/2 & 0 & 0 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 0 & 1/3 & 1/3 & 1/3 \\ 0 & 1/3 & 1/3 & 1/3 \end{bmatrix}.$$

With the aid of a symbolic mathematics program, we compute $A_{\text{wsn}} = TJT^{-1}$ where

$$J = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{24}(5-\sqrt{73}) & 0 \\ 0 & 0 & 0 & \frac{1}{24}(5+\sqrt{73}) \end{bmatrix}, \quad T = \begin{bmatrix} 1 & 0 & -2+2\sqrt{73} & -2-2\sqrt{73} \\ 1 & 0 & -11-\sqrt{73} & -11+\sqrt{73} \\ 1 & -1 & 8 & 8 \\ 1 & 1 & 8 & 8 \end{bmatrix}, \quad \text{and}$$

$$T^{-1} = \begin{bmatrix} \frac{1}{6} & \frac{1}{3} & \frac{1}{4} & \frac{1}{4} \\ 0 & 0 & -\frac{1}{2} & \frac{1}{2} \\ -\frac{1}{96}+\frac{19}{96\sqrt{73}} & -\frac{1}{48}-\frac{5}{48\sqrt{73}} & \frac{1}{64}-\frac{3}{64\sqrt{73}} & \frac{1}{64}-\frac{3}{64\sqrt{73}} \\ -\frac{1}{96}-\frac{19}{96\sqrt{73}} & -\frac{1}{48}+\frac{5}{48\sqrt{73}} & \frac{1}{64}+\frac{3}{64\sqrt{73}} & \frac{1}{64}+\frac{3}{64\sqrt{73}} \end{bmatrix}.$$

Therefore, the eigenvalues of A are $1, 0, \frac{1}{24}(5-\sqrt{73}) \approx -0.14$, and $\frac{1}{24}(5+\sqrt{73}) \approx 0.56$. Corresponding to the eigenvalue 1, the right and left eigenvector equations are:

$$A_{\text{wsn}} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1/6 \\ 1/3 \\ 1/4 \\ 1/4 \end{bmatrix}^{\mathsf{T}} A_{\text{wsn}} = \begin{bmatrix} 1/6 \\ 1/3 \\ 1/4 \\ 1/4 \end{bmatrix}^{\mathsf{T}}. \qquad \bullet$$

2.1.3 Semi-convergence and convergence for discrete-time linear systems

We can now use the Jordan normal form to study the powers of the matrix A. We start by computing

$$A^k = \underbrace{TJT^{-1} \cdot TJT^{-1} \cdot \cdots \cdot TJT^{-1}}_{k \text{ times}} = TJ^kT^{-1} = T \begin{bmatrix} J_1^k & 0 & \cdots & 0 \\ 0 & J_2^k & \ddots & 0 \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & J_m^k \end{bmatrix} T^{-1},$$

so that, for a square matrix A with Jordan blocks J_i, $i \in \{1, \ldots, m\}$, the following statements are equivalent:

(i) A is semi-convergent (resp. convergent),

(ii) J is semi-convergent (resp. convergent), and

(iii) each block J_i is semi-convergent (resp. convergent).

Next, we compute the kth power of the generic Jordan block J_i with eigenvalue λ_i as a function of block size $1, 2, 3, \ldots, j_i$; they are, respectively,

$$\begin{bmatrix} \lambda_i^k \end{bmatrix}, \begin{bmatrix} \lambda_i^k & k\lambda_i^{k-1} \\ 0 & \lambda_i^k \end{bmatrix}, \begin{bmatrix} \lambda_i^k & k\lambda_i^{k-1} & \binom{k}{2}\lambda_i^{k-2} \\ 0 & \lambda_i^k & k\lambda_i^{k-1} \\ 0 & 0 & \lambda_i^k \end{bmatrix}, \ldots, \begin{bmatrix} \lambda_i^k & \binom{k}{1}\lambda_i^{k-1} & \binom{k}{2}\lambda_i^{k-2} & \cdots & \binom{k}{j_i-1}\lambda_i^{k-j_i+1} \\ 0 & \lambda_i^k & \binom{k}{1}\lambda_i^{k-1} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \binom{k}{2}\lambda_i^{k-2} \\ 0 & \cdots & 0 & \lambda_i^k & \binom{k}{1}\lambda_i^{k-1} \\ 0 & \cdots & \cdots & 0 & \lambda_i^k \end{bmatrix},$$

$$(2.6)$$

where the binomial coefficient $\binom{k}{m} = k!/(m!(k-m)!)$ satisfies $\binom{k}{m} \leq k^m/m!$. Note that, independently of the size of J_i, each entry of the kth power of J_i is upper bounded by a constant times $k^h\lambda_i^k$ for some non-negative integer h.

To study the limit as $k \to \infty$ of the generic block J_i^k, we study the limit as $k \to \infty$ of each term of the form $k^h\lambda_i^k$. Because exponentially-decaying factors dominate polynomially-growing terms, we know

$$\lim_{k \to \infty} k^h\lambda^k = \begin{cases} 0, & \text{if } |\lambda| < 1, \\ 1, & \text{if } \lambda = 1 \text{ and } h = 0, \\ \text{non-existent or unbounded}, & \text{if } (|\lambda| = 1 \text{ with } \lambda \neq 1) \text{ or } (|\lambda| > 1) \text{ or } (\lambda = 1 \text{ and } h = 1, 2, \ldots). \end{cases}$$

$$(2.7)$$

In summary, for each block J_i with eigenvalues λ_i, we can infer that:

(i) a block J_i of size 1 is convergent if and only if $|\lambda_i| < 1$,

(ii) a block J_i of size 1 is semi-convergent if and only if $\lambda_i \leq 1$, and

(iii) a block J_i of size larger than 1 is semi-convergent and convergent if and only if $|\lambda_i| < 1$.
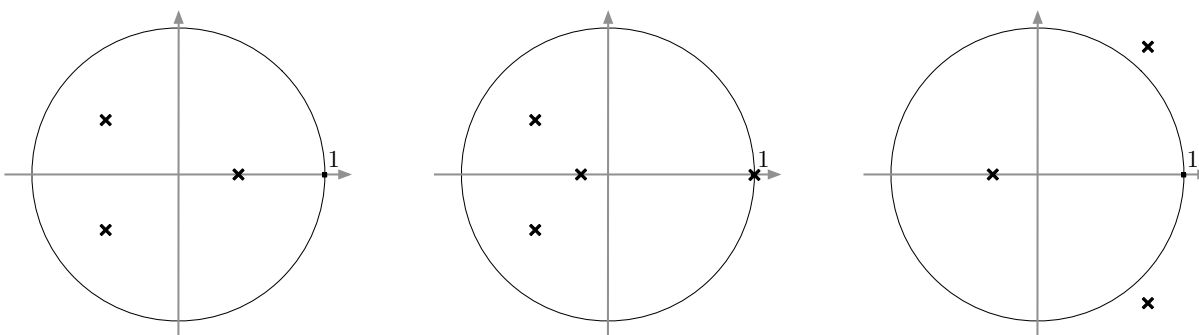
(a) The spectrum of a convergent matrix (b) The spectrum of a semi-convergent matrix, (c) The spectrum of a matrix that is not
                                        provided the eigenvalue 1 is semisimple. semi-convergent.

Figure 2.1: Eigenvalues and convergence properties of discrete-time linear systems

Based on this discussion, we are now ready to present necessary and sufficient conditions for semi-convergence and convergence of an arbitrary square matrix. We state these conditions using two two useful definitions.

Definition 2.6 (Spectrum and spectral radius of a matrix). *Given a square matrix A,*

   (i) *the spectrum of A, denoted* $\mathrm{spec}(A)$, *is the set of eigenvalues of A; and*

   (ii) *the spectral radius of A is the maximum norm of the eigenvalues of A, that is,*

$$\rho(A) = \max\{|\lambda| \mid \lambda \in \mathrm{spec}(A)\},$$

   or, equivalently, the radius of the smallest disk in $\mathbb{C}$ centered at the origin and containing the spectrum of A.

Theorem 2.7 (Convergence and spectral radius). *For a square matrix A, the following statements hold:*

   (i) *A is convergent if and only if $\rho(A) < 1$,*

   (ii) *A is semi-convergent if and only if $\rho(A) \leq 1$, no eigenvalue has unit norm other than possibly the number 1, and if 1 is an eigenvalue, then it is semisimple.*

2.2 Row-stochastic matrices and their spectral radius

Motivated by the averaging model introduced in Chapter 1, we are now interested in discrete-time linear systems defined by matrices with special properties. Specifically, we are interested in matrices with non-negative entries and whose row-sums are all equal to 1.

The square matrix $A \in \mathbb{R}^{n \times n}$ is

   (i) *non-negative* (respectively *positive*) if $a_{ij} \geq 0$ (respectively $a_{ij} > 0$) for all i and j in $\{1, \ldots, n\}$;

   (ii) *row-stochastic* if non-negative and $A\mathbb{1}_n = \mathbb{1}_n$;

(iii) *column-stochastic* if non-negative and $A^\mathsf{T}\mathbb{1}_n = \mathbb{1}_n$; and

(iv) *doubly-stochastic* if it is row- and column-stochastic.

In the following, we write $A > 0$ and $v > 0$ (respectively $A \geq 0$ and $v \geq 0$) for a positive (respectively non-negative) matrix A and vector v.

Given a finite number of points $p_1, p_2, \ldots, p_n$ in $\mathbb{R}^n$, a *convex combination of* $p_1, p_2, \ldots, p_n$ is a point of the form

$$\eta_1 p_1 + \eta_2 p_2 + \cdots + \eta_n p_n,$$

where the real numbers $\eta_1, \ldots, \eta_n$ satisfy $\eta_1 + \cdots + \eta_n = 1$ and $\eta_i \geq 0$ for all $i \in \{1, \ldots, n\}$. (For example, on the plane $\mathbb{R}^2$, the set of convex combinations of two distinct points is the segment connecting them and the set of convex combinations of three distinct points is the triangle (including its interior) defined by them.) The numbers $\eta_1, \ldots, \eta_n$ are called *convex combination coefficients* and each row of a row-stochastic matrix consists of convex combination coefficients.

2.2.1 The spectral radius for row-stochastic matrices

To characterize the spectral radius of a row-stochastic matrix, we introduce a useful general method to localize the spectrum of a matrix.

Theorem 2.8 (Geršgorin Disks Theorem). *For any square matrix $A \in \mathbb{R}^{n \times n}$,*

$$\mathrm{spec}(A) \subset \bigcup_{i \in \{1,\ldots,n\}} \underbrace{\left\{ z \in \mathbb{C} \mid |z - a_{ii}| \leq \sum_{j=1, j \neq i}^{n} |a_{ij}| \right\}}_{\text{disk in the complex plane centered at } a_{ii} \text{ with radius } \sum_{j=1, j \neq i}^{n} |a_{ij}|} .$$

Proof. Consider the eigenvalue equation $Ax = \lambda x$ for the eigenpair (λ, x), where λ and $x \neq \mathbb{0}_n$ are in general complex. Choose the index $i \in \{1, \ldots, n\}$ so that $|x_i| = \max_{j \in \{1,\ldots,n\}} |x_j| > 0$. The ith component of the eigenvalue equation can be rewritten as $\lambda - a_{ii} = \sum_{j=1, j \neq i}^{n} a_{ij} x_j / x_i$. Now, take the complex magnitude of this equality and upper-bound its right-hand side:

$$|\lambda - a_{ii}| = \left| \sum_{j=1, j\neq i}^{n} a_{ij} \frac{x_j}{x_i} \right| \leq \sum_{j=1, j\neq i}^{n} |a_{ij}| \frac{|x_j|}{|x_i|} \leq \sum_{j=1, j\neq i}^{n} |a_{ij}| .$$

This inequality defines a set of the possible locations for the arbitrary eigenvalue λ of A. The statement follows by taking the union of such sets for each eigenvalue of A. ∎

Each disk in the theorem statement is referred to as a *Geršgorin disks*, or more accurately, as a *Geršgorin row disks*; an analogous disk theorem can be stated for Geršgorin column disks. Exercise E2.17 showcases an instructive application to distributed computing of numerous topics covered so far, including convergence notions and the Geršgorin Disks Theorem.

Lemma 2.9 (Spectral properties of a row-stochastic matrix). *For a row-stochastic matrix A,*

(i) 1 *is an eigenvalue, and*

(ii) $\mathrm{spec}(A)$ *is a subset of the unit disk and $\rho(A) = 1$.*

Proof. First, recall that A being row-stochastic is equivalent to two facts: $a_{ij} \geq 0$, $i, j \in \{1, \ldots, n\}$, and $A\mathbb{1}_n = \mathbb{1}_n$. The second fact implies that $\mathbb{1}_n$ is an eigenvector with eigenvalue 1. Therefore, by definition of spectral radius, $\rho(A) \geq 1$. Next, we prove that $\rho(A) \leq 1$ by invoking the Geršgorin Disks Theorem 2.8 to show that $\mathrm{spec}(A)$ is contained in the unit disk centered at the origin. The Geršgorin disks of a row-stochastic matrix as illustrated in Figure 2.2.
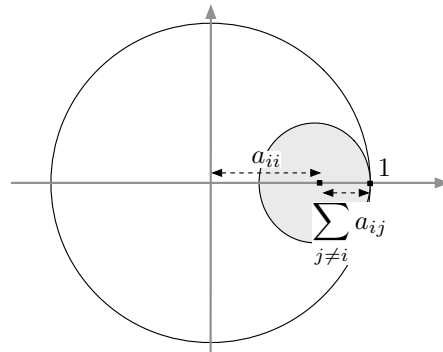
Figure 2.2: All Geršgorin disks of a row-stochastic matrix are contained in the unit disk.

Note that A being row-stochastic implies $a_{ii} \in [0, 1]$ and $a_{ii} + \sum_{j \neq i} a_{ij} = 1$. Hence, the center of the ith Geršgorin disk belongs to the positive real axis between 0 and 1, and the right-most point in the disk is at 1. ∎

Note: because 1 is an eigenvalue of each row-stochastic matrix A, clearly A is not convergent. But it is possible for A to be semi-convergent.

2.3   Perron–Frobenius theory

We have seen how row-stochastic matrices are not convergent; we now focus on characterizing those that are semi-convergent. To establish whether a row-stochastic matrix is semi-convergent, we introduce the widely-established Perron–Frobenius theory for non-negative matrices.

2.3.1   Classification of non-negative matrices

In the previous section we already defined non-negative and positive matrices. In this section we are interested in classifying non-negative matrices in terms of their zero/nonzero pattern and of the asymptotic behavior of their powers.

We start by introducing simple example non-negative matrices and related comments:

$A_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $\quad \text{spec}(A_1) = \{1, 1\}$, the zero/nonzero pattern in A_1^k is constant, and $\lim_{k \to \infty} A_1^k = I_2$,

$A_2 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, $\quad \text{spec}(A_2) = \{1, -1\}$, the zero/nonzero pattern in A_2^k is periodic, and $\lim_{k \to \infty} A_2^k$ does not exist,

$A_3 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$, $\quad \text{spec}(A_3) = \{0, 0\}$, $A_3^k = 0$ for all $k \geq 2$, and $\lim_{k \to \infty} A_3^k = 0$,

$A_4 = \dfrac{1}{2} \begin{bmatrix} 1 & 1 \\ 2 & 0 \end{bmatrix}$, $\quad \text{spec}(A_4) = \{1, -1/2\}$, $A_4^k > 0$ for all $k \geq 2$, and $\lim_{k \to \infty} A_4^k = \dfrac{1}{3} \begin{bmatrix} 2 & 1 \\ 2 & 1 \end{bmatrix}$, and

$A_5 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$, $\quad \text{spec}(A_5) = \{1, 1\}$, the zero/nonzero pattern in A_5^k is constant and $\lim_{k \to \infty} A_5^k$ is unbounded.

Based on these preliminary examples, we now introduce two sets of non-negative matrices with certain characteristic properties.

Definition 2.10 (Irreducible and primitive matrices). *For $n \geq 2$, an $n \times n$ non-negative matrix A is*

 (i) *irreducible if $\sum_{k=0}^{n-1} A^k$ is positive,*

 (ii) *primitive if there exists $k \in \mathbb{N}$ such that A^k positive.*

A matrix that is not irreducible is said to be reducible.

Note that A_1, A_3 and A_5 are reducible whereas A_2 and A_4 are irreducible. Moreover, note that A_2 is not primitive whereas A_4 is. Additionally, note that a positive matrix is clearly primitive. Finally, if there is $k \in \mathbb{N}$ such that A^k is positive, then (one can show that) all subsequent powers A^{k+1}, A^{k+2}, ... are necessarily positive as well; see Exercise E2.5.

Note: In other words, A is irreducible if, for any $(i, j) \in \{1, \ldots, n\}^2$ there is a $k = k(i, j) \leq (n - 1)$ such that $(A^k)_{ij} > 0$. There are multiple equivalent ways to define irreducible matrices. We discuss four equivalent characterizations later in Theorem 4.3.

We now state a useful result and postpone its proof to Exercise E4.5.

Lemma 2.11 (A primitive matrix is irreducible). *If a non-negative matrix is primitive, then it is also irreducible.*

As a consequence of this lemma we can draw the set diagram in Figure 2.3 describing the set of non-negative square matrices and its subsets of irreducible, primitive and positive matrices. Note that the inclusions in the diagram are strict in the sense that:

 (i) A_3 is non-negative but not irreducible;

 (ii) A_2 is irreducible but not primitive; and

 (iii) A_4 is primitive but not positive.
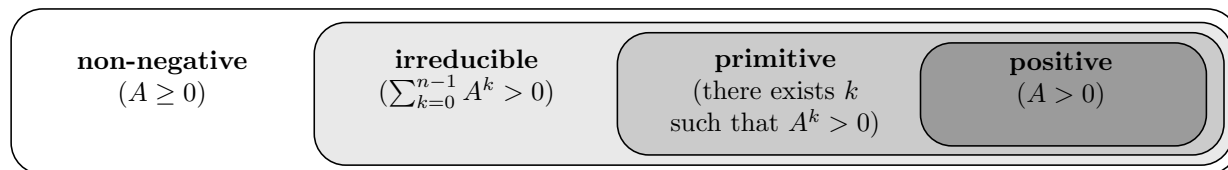
Figure 2.3: The set of non-negative square matrices and its subsets of irreducible, primitive and positive matrices.

2.3.2   Main results

We are now ready to state the main results in Perron-Frobenius theory and characterize the properties of the spectral radius of a non-negative matrix as a function of the matrix properties.

Theorem 2.12 (Perron-Frobenius Theorem). *Let $A \in \mathbb{R}^{n \times n}$, $n \geq 2$. If A is non-negative, then*

(i) *there exists a real eigenvalue $\lambda \geq |\mu| \geq 0$ for all other eigenvalues μ,*

(ii) *the right and left eigenvectors v and w of λ can be selected non-negative.*

If additionally A is irreducible, then

(iii) *the eigenvalue λ is strictly positive and simple,*

(iv) *the right and left eigenvectors v and w of λ are unique and positive, up to rescaling.*

If additionally A is primitive, then

(v) *the eigenvalue λ satisfies $\lambda > |\mu|$ for all other eigenvalues μ.*

Some remarks and some additional statements are in order. For non-negative matrices, the real non-negative eigenvalue λ is the spectral radius $\rho(A)$ of A. We refer to λ as the *dominant eigenvalue* of A; it is also referred to as the *Perron root*. The dominant eigenvalue is equivalently defined by

$$\rho(A) = \inf\{\lambda \in \mathbb{R} \mid Au \leq \lambda u \text{ for all } u > 0\}.$$

For irreducible matrices, the right and left eigenvectors v and w (unique up to rescaling and selected non-negative) of the dominant eigenvalue λ are called the *right and left dominant eigenvector*, respectively. One can show that, up to rescaling, the right dominant eigenvector is the only positive right eigenvector of a primitive matrix A (a similar statement holds for the left dominant eigenvector); see also Exercise E2.4. We refer to Theorem 4.9 and Exercise E4.9 in Section 4.5 for some useful bounds on the dominant eigenvalue and to Theorem 5.1 in Section 5.1 for a version of the Perron–Frobenius Theorem for reducible matrices.

Remark 2.13 (Examples and counterexamples). *The characterizations in the theorem are sharp in the following sense:*

(i) *the matrix $A_3 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ is non-negative and reducible, and, indeed, its dominant eigenvalue is 0;*

(ii) the matrix $A_2 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ is irreducible but not primitive and, indeed, its dominant eigenvalues $+1$ is not stricly larger, in magnitude, than the other eigenvalues -1.

2.3.3 Applications to dynamical systems

Given a primitive non-negative matrix A, the Perron–Frobenius Theorem for primitive matrices has immediate consequences for the behavior of A^k as $k \to \infty$. We start with a semiconvergence result that applies to primitive matrices. We postponed the proof to Section 2.3.4.

Theorem 2.14 (Powers of nonnegative matrices with a simple and strictly dominant eigenvalue).
Consider a nonnegative matrix A whose dominant eigenvalue λ is simple and strictly larger, in magnitude, than all other eigenvalues (e.g., assume A is primitive). Let v and w denote right and left dominant eigenvectors normalized so that $v \geq 0$ and $v^\mathsf{T} w = 1$. Then A/λ is semi-convergent and

$$\lim_{k \to \infty} \frac{A^k}{\lambda^k} = vw^\mathsf{T}.$$

The matrix vw^T is a rank-one projection matrix with numerous properties, which we discuss in Exercise E2.13.

We now apply this result to primitive row-stochastic matrices and to the averaging model $x(k+1) = Ax(k)$. For a row-stochastic A, the right eigenvector of the eigenvalue 1 is selected as $\mathbb{1}_n$.

Corollary 2.15 (Consensus for primitive row-stochastic). *For a primitive row-stochastic matrix A,*

(i) *the simple eigenvalue $\rho(A) = 1$ is strictly larger than the magnitude of all other eigenvalues, hence A is semi-convergent;*

(ii) $\lim_{k \to \infty} A^k = \mathbb{1}_n w^\mathsf{T}$, *where w is the left dominant eigenvector of A with eigenvalue 1 satisfying $w_1 + \cdots + w_n = 1$;*

(iii) *the solution to the averaging model $x(k+1) = Ax(k)$ satisfies*

$$\lim_{k \to \infty} x(k) = \left(w^\mathsf{T} x(0) \right) \mathbb{1}_n.$$

In this case we say that the dynamical system achieves consensus,

(iv) *if additionally A is doubly-stochastic, then $w = \frac{1}{n}\mathbb{1}_n$ (because $A^\mathsf{T}\mathbb{1}_n = \mathbb{1}_n$ and $\frac{1}{n}\mathbb{1}_n^\mathsf{T}\mathbb{1}_n = 1$) so that*

$$\lim_{k \to \infty} x(k) = \frac{\mathbb{1}_n^\mathsf{T} x(0)}{n} \mathbb{1}_n = \text{average}\left(x(0) \right) \mathbb{1}_n.$$

In this case we say that the dynamical system achieves average consensus.

Note: $\mathbb{1}_n w^\mathsf{T} = \begin{bmatrix} w^\mathsf{T} \\ \vdots \\ w^\mathsf{T} \end{bmatrix} = \begin{bmatrix} w_1 & w_2 & \cdots & w_n \\ \vdots & \vdots & \vdots & \vdots \\ w_1 & w_2 & \cdots & w_n \end{bmatrix}$, and $(\mathbb{1}_n w^\mathsf{T}) x(0) = (w^\mathsf{T} x(0)) \mathbb{1}_n = \begin{bmatrix} w^\mathsf{T} x(0) \\ \vdots \\ w^\mathsf{T} x(0) \end{bmatrix}$.

Note: the limiting vector is therefore a weighted average of the initial conditions. The relative weights of the initial conditions are the convex combination coefficients $w_1, \ldots, w_n$. In a social influence network, the coefficient w_i is regarded as the "social influence" of agent i.

Example 2.16 (Revisiting the wireless sensor network example). Finally, as a numerical example, let us reconsider the wireless sensor network discussed in Section 1.2 and the 4-dimensional row-stochastic matrix A_wsn. First, note that A_wsn is primitive because A_wsn^2 is positive:

$$A_\text{wsn} = \begin{bmatrix} 1/2 & 1/2 & 0 & 0 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 0 & 1/3 & 1/3 & 1/3 \\ 0 & 1/3 & 1/3 & 1/3 \end{bmatrix} \implies A_\text{wsn}^2 = \begin{bmatrix} 3/8 & 3/8 & 1/8 & 1/8 \\ 3/16 & 17/48 & 11/48 & 11/48 \\ 1/12 & 11/36 & 11/36 & 11/36 \\ 1/12 & 11/36 & 11/36 & 11/36 \end{bmatrix}.$$

Therefore, the Perron–Frobenius Theorem 2.12 for primitive matrices applies to A_wsn. The four pairs of eigenvalues and right eigenvectors of A_wsn (as computed in Example 2.5) are:

$$\left(1, \mathbb{1}_4\right), \quad \left(\frac{1}{24}(5+\sqrt{73}), \begin{bmatrix} -2-2\sqrt{73} \\ -11+\sqrt{73} \\ 8 \\ 8 \end{bmatrix}\right), \quad \left(\frac{1}{24}(5-\sqrt{73}), \begin{bmatrix} 2(-1+\sqrt{73}) \\ -11-\sqrt{73} \\ 8 \\ 8 \end{bmatrix}\right), \quad \left(0, \begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \end{bmatrix}\right).$$

Moreover, we know that A_wsn is semi-convergent. To apply the convergence results in Corollary 2.15, we numerically compute its left dominant eigenvector, normalized to have unit sum, to be $w = [1/6, 1/3, 1/4, 1/4]^\mathsf{T}$ so that we have:

$$\lim_{k\to\infty} A_\text{wsn}^k = \mathbb{1}_4 w^\mathsf{T} = \begin{bmatrix} 1/6 & 1/3 & 1/4 & 1/4 \\ 1/6 & 1/3 & 1/4 & 1/4 \\ 1/6 & 1/3 & 1/4 & 1/4 \\ 1/6 & 1/3 & 1/4 & 1/4 \end{bmatrix}.$$

Therefore, each solution to the averaging system $x(k+1) = A_\text{wsn}x(k)$ converges to a consensus vector $(w^\mathsf{T}x(0))\mathbb{1}_4$, that is, the value at each node of the wireless sensor network converges to $w^\mathsf{T}x(0) = (1/6)x_1(0)+(1/3)x_2(0)+(1/4)x_3(0)+(1/4)x_4(0)$. Note that A_wsn is not doubly-stochastic and, therefore, the averaging algorithm does not achieve average consensus and that node 2 has more influence than the other nodes.   •

Note: If A is reducible, then clearly it is not primitive. Yet, it is possible for an averaging algorithm described by a reducible matrix to converge to consensus. In other words, Corollary 2.15 provides only a sufficient condition for consensus. Here is a simple example of an averaging algorithm described by a reducible matrix that converges to consensus:

$$x_1(k+1) = x_1(k),$$
$$x_2(k+1) = x_1(k).$$

To fully understand what all phenomena are possible and what properties of A are necessary and sufficient for convergence to consensus, we will study graph theory in the next two chapters.

2.3.4 Selected proofs

We conclude this section with the proof of some selected statements.

Proof of Theorem 2.12. We start by establishing that a primitive A matrix satisfies $\rho(A) > 0$. By contradiction, if $\mathrm{spec}(A) = \{0\}$, then the Jordan normal form J of A is nilpotent, that is, there is a $k^* \in \mathbb{N}$ so that $J^k = A^k = 0$ for all $k \geq k^*$. But this is a contradiction because A being primitive implies that there is $k^* \in \mathbb{N}$ so that $A^k > 0$ for all $k \geq k^*$.

Next, we prove that $\rho(A)$ is a real positive eigenvalue with a positive right eigenvector $v > 0$. We first focus on the case that A is a positive matrix, and later show how to generalize the proof to the case of primitive matrices. Without loss of generality, assume $\rho(A) = 1$. If (λ, x) is an eigenpair for A such that $|\lambda| = \rho(A) = 1$, then

$$|x| = |\lambda||x| = |\lambda x| = |Ax| \leq |A||x| = A|x| \quad \implies \quad |x| \leq A|x|. \tag{2.8}$$

Here, we use the notation $|x| = (|x_i|)_{i \in \{1,\dots,n\}}$, $|A| = \{|a_{ij}|\}_{i,j \in \{1,\dots,n\}}$, and vector inequalities are understood component-wise. In what follows, we show $|x| = A|x|$. With the shorthands $z = A|x|$ and $y = z - |x|$, equation (2.8) reads $y \geq 0$ and we aim to show $y = 0$. By contradiction, assume y has a non-zero component. Therefore, $Ay > 0$. Independently, we also know $z = A|x| > 0$. Thus, there must exist $\varepsilon > 0$ such that $Ay > \varepsilon z$. Eliminating the variable y in the latter equation, we obtain $A_\varepsilon z > z$, where we define $A_\varepsilon = A/(1 + \varepsilon)$. The inequality $A_\varepsilon z > z$ implies $A_\varepsilon^k z > z$ for all $k > 0$. Now, observe that $\rho(A_\varepsilon) < 1$ so that $\lim_{k \to \infty} A_\varepsilon^k = \mathbb{0}_{n \times n}$ and therefore $0 > z$. Since we also knew $z > 0$, we now have a contradiction. Therefore, we know $y = 0$.

So far, we have established that $|x| = A|x|$, so that $(1, |x|)$ is an eigenpair for A. Also note that $A > 0$ and $x \neq 0$ together imply $A|x| > 0$. Therefore we have established that 1 is an eigenvalue of A with eigenvector $|x| > 0$. Next, observe that the above reasoning is correct also for primitive matrices if one replaces the first equality (2.8) by $|x| = |\lambda^k||x|$ and carries the exponent k throughout the proof.

In summary, we have established that there exists a real eigenvalue $\lambda > 0$ such that $\lambda \geq |\mu|$ for all other eigenvalues μ, and that each right (and therefore also left) eigenvector of λ can be selected positive up to rescaling. It remains to prove that λ is simple and is strictly greater than the magnitude of all other eigenvalues. For the proof of these two points, we refer to (Meyer, 2001, Chapter 8). ∎

Proof of Theorem 2.14. Because λ is simple, we write the Jordan normal form of A as

$$A = T \begin{bmatrix} \lambda & \mathbb{0}_{1 \times (n-1)} \\ \mathbb{0}_{(n-1) \times 1} & B \end{bmatrix} T^{-1}, \tag{2.9}$$

where the block-diagonal matrix $B \in \mathbb{R}^{(n-1) \times (n-1)}$ contains the Jordan blocks of all eigenvalues of A except for λ. Because λ is strictly dominant, we know that $\rho(B/\lambda) < 1$, which in turn implies

$$\lim_{k \to +\infty} B^k/\lambda^k = \mathbb{0}_{(n-1) \times (n-1)}.$$

Recall $A^k = T \begin{bmatrix} \lambda & 0 \\ 0 & B \end{bmatrix}^k T^{-1}$ so that

$$\lim_{k \to +\infty} \left(\frac{A}{\lambda}\right)^k = T \left(\lim_{k \to +\infty} \begin{bmatrix} 1^k & 0 \\ 0 & (B/\lambda)^k \end{bmatrix} \right) T^{-1} = T \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} T^{-1}.$$

Next, we let $v_1, \ldots, v_n$ (respectively, $w_1, \ldots, w_n$) denote the columns of T (respectively the rows of T^{-1}), that is, $T = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}$, and $(T^{-1})^\mathsf{T} = \begin{bmatrix} w_1 & \cdots & w_n \end{bmatrix}$. Equation (2.9) is equivalently written as

$$A \underbrace{\begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}}_{=T} = \underbrace{\begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}}_{=T} \begin{bmatrix} \lambda & 0 \\ 0 & B \end{bmatrix}.$$

The first column of the above matrix equation is $Av_1 = \lambda v_1$, that is, v_1 is the right dominant eigenvector of A, up to rescaling. Recall that λ is simple and right eigenvalue is unique up to rescaling. By analogous arguments, we find that w_1 is the left dominant eigenvector of A, up to rescaling. With this notation, some bookkeeping leads to:

$$\lim_{k \to +\infty} \left(\frac{A}{\lambda} \right)^k = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix} \begin{bmatrix} 1 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 0 \end{bmatrix} \begin{bmatrix} w_1^\mathsf{T} \\ w_2^\mathsf{T} \\ \vdots \\ w_n^\mathsf{T} \end{bmatrix} = v_1 w_1^\mathsf{T}.$$

Finally, the $(1,1)$ entry of the matrix equality $TT^{-1} = I_n$ gives precisely the normalization $v_1^\mathsf{T} w_1 = 1$. In summary, v_1 and w_1 are the right and left dominant eigenvectors, up to rescaling, and they are known to satisfy $v_1^\mathsf{T} w_1 = 1$. Hence, $vw^\mathsf{T} = v_1 w_1^\mathsf{T}$. This concludes the proof of Theorem 2.14.   ∎

2.4   Historical notes and further reading

For comprehensive treatments on matrix theory we refer to the classic texts by Gantmacher (1959), Horn and Johnson (1985), and Meyer (2001).

   Regarding the main Perron–Frobenius Theorem 2.12, historically, Perron (1907) established the original result fo the case of positive matrices. Frobenius (1912) provided the substantial extension to the settings of primitive and irreducible matrices. More historical information is given in (Meyer, 2001, Chapter 8).

   Theorem 2.14 is generalized as follows: an irreducible row-stochastic matrix A with left-dominant eigenvector w satisfies $\lim_{k \to \infty} \frac{1}{k} (I_n + A + \cdots + A^{k-1}) = \mathbb{1}_n w^\mathsf{T}$. We refer to (Meyer, 2001, Section 8.4) for more details on this result and to (Breiman, 1992, Chapter 6) for the more general Ergodic Theorem.

2.5 Exercises

E2.1 **Simple properties of stochastic matrices.** Let $A_1, A_2, \ldots, A_k$ be $n \times n$ matrices, let $A_1 A_2 \cdots A_k$ be their product and let $\eta_1 A_1 + \cdots + \eta_k A_k$ be their convex combination with arbitrary convex combination coefficients. Show that

(i) if $A_1, A_2, \ldots, A_k$ are non-negative, then their product and all their convex combinations are non-negative,

(ii) if $A_1, A_2, \ldots, A_k$ are row-stochastic, then their product and all their convex combinations are row-stochastic, and

(iii) if $A_1, A_2, \ldots, A_k$ are doubly-stochastic, then their product and all their convex combinations are doubly-stochastic.

E2.2 **Semi-convergence and Jordan block decomposition.** Consider a matrix $A \in \mathbb{C}^{n \times n}$, $n \geq 2$, with $\rho(A) = 1$. Show that the following statements are equivalent:

(i) A is semi-convergent,

(ii) either $A = I_n$ or there exists a nonsingular matrix $T \in \mathbb{C}^{n \times n}$ and a number $m \in \{1, \ldots, n-1\}$ such that

$$A = T \begin{bmatrix} I_m & \mathbb{0}_{m \times (n-m)} \\ \mathbb{0}_{(n-m) \times m} & B \end{bmatrix} T^{-1},$$

where $B \in \mathbb{C}^{(n-m) \times (n-m)}$ is convergent, that is, $\rho(B) < 1$.

Note: If A is real, then it is possible to find real-valued matrices T and B in statement (ii) by using the notion of real Jordan normal form (Hogben, 2013).

E2.3 **Row-stochastic matrices after pairwise-difference similarity transform.** For $n \geq 2$, let $A \in \mathbb{R}^{n \times n}$ be row stochastic. Define $T \in \mathbb{R}^{n \times n}$ by

$$T = \begin{bmatrix} -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \\ 1/n & 1/n & \ldots & 1/n \end{bmatrix}.$$

Perform the following tasks:

(i) for $x = [x_1, \ldots, x_n]^{\mathsf{T}}$, write Tx in components and show T is invertible,

(ii) show $TAT^{-1} = \begin{bmatrix} A_{\text{stable}} & \mathbb{0}_{n-1} \\ c^{\mathsf{T}} & 1 \end{bmatrix}$ for some $A_{\text{stable}} \in \mathbb{R}^{(n-1) \times (n-1)}$ and $c \in \mathbb{R}^{n-1}$,

(iii) if A is doubly-stochastic, then $c = 0$,

(iv) show that A primitive implies $\rho(A_{\text{stable}}) < 1$, and

(v) compute TAT^{-1} for $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$.

E2.4 **Uniqueness of the non-negative eigenvector in irreducible non-negative matrices.** Given a square matrix $A \in \mathbb{R}^{n \times n}$, show that:

(i) if v_1 is a right eigenvector of A corresponding to the eigenvalue λ_1, w_2 is a left eigenvector of A relative to λ_2, and $\lambda_1 \neq \lambda_2$, then $v_1 \perp w_2$; and

(ii) if A is non-negative and irreducible and $u \in \mathbb{R}^n_{\geq 0}$ is a right non-negative eigenvector of A, then u is an eigenvector corresponding to the eigenvalue $\rho(A)$.

E2.5   **Powers of primitive matrices.** Let $A \in \mathbb{R}^{n \times n}$ be non-negative. Show that $A^k > 0$, for some $k \in \mathbb{N}$, implies $A^m > 0$ for all $m \geq k$.

E2.6   **Sufficient condition for primitivity.** Consider a non-negative matrix $A \in \mathbb{R}^{n \times n}$. If there exists $r \in \{1, \ldots, n\}$ such that $A_{rj} > 0$ and $A_{ir} > 0$ for all $i, j \in \{1, \ldots, n\}$, that is, if A has the sparsity pattern

$$
A = \begin{bmatrix}
 & & & & \star & & \\
 & & & & \star & & \\
 & & & & \star & & \\
 & & & & \star & & \\
\star & \star & \star & \star & \star & \star & \star \\
 & & & & \star & & \\
 & & & & \star & &
\end{bmatrix},
$$

then the matrix A is primitive. (Here the symbol $\star$ denotes a strictly positive entry. The absence of a symbol denotes a positive or zero entry.)

E2.7   **Reducibility fallacy.** Consider the following statement:

> *Any non-negative square matrix $A \in \mathbb{R}^{n \times n}$ with a zero entry is reducible, because the zero entry can be moved in position $A_{n,1}$ via a permutation similarity transformation.*

Is the statement true? If yes, explain why; if not, provide a counterexample.

E2.8   **Symmetric doubly-stochastic matrix.** Let $A \in \mathbb{R}^{n \times n}$ be doubly-stochastic. Show that:

   (i) the matrix $A^\mathsf{T} A$ is doubly-stochastic and symmetric,

   (ii) $\operatorname{spec}(A^\mathsf{T} A) \subset [0, 1]$,

   (iii) the eigenvalue 1 of $A^\mathsf{T} A$ is not necessarily simple even if A is irreducible.

E2.9   **On some non-negative matrices.** How many 2×2 matrices exist that are simultaneously doubly stochastic, irreducible and not primitive? Justify your claim.

E2.10 **Discrete-time affine systems.** Given $A \in \mathbb{R}^{n \times n}$ and $b \in \mathbb{R}^n$, consider the discrete-time affine system

$$
x(k + 1) = Ax(k) + b.
$$

Assume A is convergent and show that

   (i) the matrix $(I_n - A)$ is invertible,

   (ii) the only equilibrium point of the system is $(I_n - A)^{-1}b$, and

   (iii) $\lim_{k \to \infty} x(k) = (I_n - A)^{-1}b$ for all initial conditions $x(0) \in \mathbb{R}^n$.
   Hint: *Define a new sequence $y(k)$, $k \in \mathbb{Z}_{\geq 0}$, by $y(k) = x(k) - x^*$ for an appropriate x^*.*

E2.11 **An affine averaging system.** Given a primitive doubly-stochastic matrix A and a vector b satisfying $\mathbb{1}_n^\mathsf{T} b = 0$, consider the dynamical system
$$
x(k + 1) = Ax(k) + b.
$$

Show that

   (i) the quantity $k \mapsto \mathbb{1}_n^\mathsf{T} x(k)$ is constant,

   (ii) for each $\alpha \in \mathbb{R}$, there exists a unique equilibrium point x_α^* satisfying $\mathbb{1}_n^\mathsf{T} x_\alpha^* = \alpha$ and satisfying generically $x_\alpha^* \notin \operatorname{span}\{\mathbb{1}_n\}$, and

   (iii) all solutions with initial condition $x(0)$ satisfying $\mathbb{1}_n^\mathsf{T} x(0) = \alpha$ converge to x_α^*.

Hint: First, use Exercise E2.2 and study the properties of the similarity transformation matrix T and its inverse T^{-1}. Second, define $y(k) = T^{-1}x(k)$, show the evolution of $y_1(k)$ is decoupled from that of the other entries and apply E2.10.

E2.12 **The Neumann series.** For $A \in \mathbb{C}^{n \times n}$, show that the following statements are equivalent:

(i) $\rho(A) < 1$,

(ii) $\lim_{k \to \infty} A^k = \mathbb{0}_{n \times n}$, and

(iii) the *Neumann series* $\sum_{k=0}^{\infty} A^k$ converges.

Additionally show that, if any and hence all of these conditions hold, then

(iv) the matrix $(I_n - A)$ is invertible, and

(v) $\displaystyle\sum_{k=0}^{\infty} A^k = (I_n - A)^{-1}$.

Hint: This statement, written in the style of (Meyer, 2001, Section 7.10), is an extension of Theorem 2.7 and a generalization of the classic geometric series $\frac{1}{1-x} = \sum_{k=0}^{\infty} x^k$, convergent for all $|x| < 1$. For the proof, the hint is to use the Jordan normal form.

E2.13 **The rank-one projection matrix defined by a primitive matrix.** This exercise requires the following notions from linear algebra: a square matrix B is a *projection matrix* if $B^2 = B$, a vector space V is the *direct sum* of two subspaces U and W, written $V = U \oplus W$, if each $v \in V$ defines unique $u \in U$ and $w \in W$ such that $v = u + w$, and a subspace U is *invariant under a linear map* B if $u \in U$ implies $Bu \in U$.

Let A be an n-dimensional primitive matrix with dominant eigenvalue λ, right dominant eigenvector $v > 0$ and left dominant eigenvector w with the normalization $v^{\mathsf{T}}w = 1$. Define the rank-one matrix $J_A := vw^{\mathsf{T}}$. Show that:

(i) $J_A = J_A^2$ is a projection matrix whose image is $\mathrm{span}\{v\}$,

(ii) $I_n - J_A = (I_n - J_A)^2$ is a projection matrix whose image is $\mathrm{kernel}(J_A) = \{q \in \mathbb{R}^n \mid w^{\mathsf{T}}q = 0\} = \mathrm{span}\{w\}^{\perp}$,

(iii) $AJ_A = J_A A = \lambda J_A$,

(iv) $\mathbb{R}^n = \mathrm{span}\{v\} \oplus \mathrm{span}\{w\}^{\perp}$ and both subspaces $\mathrm{span}\{v\}$ and $\mathrm{span}\{w\}^{\perp}$ are invariant under A,

(v) if A is symmetric, then J_A is an orthogonal projection,

(vi) the restriction of A to $\mathrm{span}\{w\}^{\perp}$ has all its eigenvalues strictly less than λ in magnitude, and the restriction of A to the $\mathrm{span}\{v\}$ is multiplication by λ, and

(vii) $\rho(A - \lambda J_A) < \lambda$.

E2.14 **Permutation and orthogonal matrices.** A set G with a binary operation mapping two elements of G into another element of G, denoted by $(a, b) \mapsto a \star b$, is a *group* if:

(G1) $a \star (b \star c) = (a \star b) \star c$ for all $a, b, c \in G$ (associativity property);

(G2) there exists $e \in G$ such that $a \star e = e \star a = a$ for all $a \in G$ (existence of an identity element); and

(G3) there exists $a^{-1} \in G$ such that $a \star a^{-1} = a^{-1} \star a = e$ for all $a \in G$ (existence of inverse elements).

Recall that: a permutation matrix is an $n \times n$ binary (i.e., entries equal to 0 and 1) matrix with precisely one entry equal to 1 in every row and every column; a permutation matrix acts on a vector by permuting its entries. Also recall that an orthogonal matrix R is an $n \times n$ matrix whose columns and rows are orthonormal vectors, i.e., $RR^{\mathsf{T}} = I_n$; an orthogonal matrix acts on a vector like a rotation and/or reflection. Prove that

(i) the set of $n \times n$ permutation matrices with the operation of matrix multiplication is a group;

(ii) the set of $n \times n$ orthogonal matrices with the operation of matrix multiplication is a group;

(iii) each permutation matrix is orthogonal.

E2.15 **On doubly-stochastic and permutation matrices.** The following result is known as the Birkhoff – Von Neumann Theorem. For a matrix $A \in \mathbb{R}^{n \times n}$, the following statements are equivalent:

  (i) A is doubly-stochastic; and

  (ii) A is a convex combination of permutation matrices.

Do the following:

  - show that the set of doubly-stochastic matrices is convex (i.e., given any two doubly-stochastic matrices A_1 and A_2, any matrix of the form $\lambda A_1 + (1 - \lambda)A_2$, for $\lambda \in [0, 1]$, is again doubly-stochastic);

  - show that (ii) $\implies$ (i);

  - find in the literature a proof of (i) $\implies$ (ii) and sketch it in one or two paragraphs.

E2.16 **Determinants of block matrices (Silvester, 2000).** Given square matrices $A, B, C, D \in \mathbb{R}^{n \times n}$, $n \geq 1$, useful identities are

$$\det \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{cases} \det(D)\det(A - BD^{-1}C), & \text{if } D \text{ is invertible,} & \text{(E2.1a)} \\ \det(AD - BC), & \text{if } CD = DC, & \text{(E2.1b)} \\ \det(DA - BC), & \text{if } BD = DB. & \text{(E2.1c)} \end{cases}$$

  (i) Prove equality (E2.1a).

  (ii) Prove equality (E2.1b) and (E2.1c) assuming D is invertible.

Hint: Show $\begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} I_n & \mathbb{0}_{n \times n} \\ -D^{-1}C & I_n \end{bmatrix} = \begin{bmatrix} A - BD^{-1}C & B \\ \mathbb{0}_{n \times n} & D \end{bmatrix}$. *We refer to (Silvester, 2000) for the complete proofs and for the additional identities*

$$\det \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{cases} \det(AD - CB), & \text{if } AC = CA, & \text{(E2.2a)} \\ \det(DA - CB), & \text{if } AB = BA. & \text{(E2.2b)} \end{cases}$$

E2.17 **The Jacobi relaxation in parallel computation.** Consider n distributed processors that aim to collectively solve the linear equation $Ax = b$, where $b \in \mathbb{R}^n$ and $A \in \mathbb{R}^{n \times n}$ is invertible and its diagonal elements a_{ii} are nonzero. Each processor stores a variable $x_i(k)$ as the discrete-time variable k evolves and applies the following iterative strategy termed *Jacobi relaxation*. At time step $k \in \mathbb{N}$ each processor performs the local computation

$$x_i(k+1) = \frac{1}{a_{ii}} \left(b_i - \sum_{j=1, j \neq i}^{n} a_{ij} x_j(k) \right), \quad i \in \{1, \ldots, n\}.$$

Next, each processor $i \in \{1, \ldots, n\}$ sends its value $x_i(k+1)$ to all other processors $j \in \{1, \ldots, n\}$ with $a_{ji} \neq 0$, and they iteratively repeat the previous computation. The initial values of the processors are arbitrary.

  (i) Assume the Jacobi relaxation converges, i.e., assume $\lim_{k \to \infty} x(k) = x^*$. Show that $Ax^* = b$.

  (ii) Give a necessary and sufficient condition for the Jacobi relaxation to converge.

  (iii) Use Geršgorin Disks Theorem 2.8 to show that the Jacobi relaxation converges if A is *strictly row diagonally dominant*, that is, if $|a_{ii}| > \sum_{j=1, j \neq i}^{n} |a_{ij}|$ for all $i \in \{1, \ldots, n\}$.

E2.18 **The Jacobi over-relaxation in parallel computation.** We now consider a more sophisticated version of the Jacobi relaxation presented in Exercise E2.17. Consider again n distributed processors that aim to collectively solve the linear equation $Ax = b$, where $b \in \mathbb{R}^n$ and $A \in \mathbb{R}^{n \times n}$ is invertible and its diagonal elements a_{ii} are nonzero. Each processor stores a variable $x_i(k)$ as the discrete-time variable k evolves and applies the

following iterative strategy termed *Jacobi over-relaxation*. At time step $k \in \mathbb{N}$ each processor performs the local computation

$$x_i(k+1) = (1-\omega)x_i(k) + \frac{\omega}{a_{ii}}\Big(b_i - \sum_{j=1, j\neq i}^{n} a_{ij}x_j(k)\Big), \quad i \in \{1, \ldots, n\},$$

where $\omega \in \mathbb{R}$ is an adjustable parameter. Next, each processor $i \in \{1, \ldots, n\}$ sends its value $x_i(k+1)$ to all other processors $j \neq i$ with $a_{ji} \neq 0$, and they iteratively repeat the previous computation. The initial values of the processors are arbitrary.

(i) Assume the Jacobi over-relaxation converges to $x^\star$ and show that $Ax^\star = b$ if $\omega \neq 0$.

(ii) Find the expression governing the dynamics of the error variable $e(k) := x(k) - x^\star$.

(iii) Suppose that A is *strictly row diagonally dominant*, that is $|a_{ii}| > \sum_{j\neq i} |a_{ij}|$. Use the Geršgorin Disks Theorem 2.8 to discuss the convergence properties of the algorithm for all possible values of $\omega \in \mathbb{R}$.

   Hint: Consider different thresholds for ω.

E2.19 **Simulation (cont'd).** This is a followup to Exercise E1.3. Consider the linear averaging algorithm in equation (1.1): set $n = 5$, select the initial state equal to $(1, -1, 1, -1, 1)$, and use (a) the complete graph (b) a cycle graph, and (c) a star graph with node 1 as center.

(i) To which value do all nodes converge to?

(ii) Compute the dominant left eigenvector of the averaging matrix associated to each of the three graphs and verify that the result in Corollary 2.15(iii) is correct.

E2.20 **Continuous- and discrete-time control control of mobile robots.** Consider n robots moving on the line with positions $z_1, z_2, \ldots z_n \in \mathbb{R}$. In order to gather at a common location (i.e., reach rendezvous), each robot heads for the centroid of its neighbors, that is,

$$\dot{z}_i = \frac{1}{n-1}\Big(\sum_{j=1, j\neq i}^{n} z_j\Big) - z_i.$$

(i) Will the robots asymptotically rendezvous at a common location?

(ii) Consider the Euler discretization of the above closed-loop dynamics with sampling rate $T > 0$:

$$z_i(k+1) = z_i(k) + T\Big(\frac{1}{n-1}\Big(\sum_{j=1, j\neq i}^{n} z_j(k)\Big) - z_i(k)\Big).$$

For which values of the sampling period T will the robots rendezvous?

Hint: Use the modal decomposition in Remark 2.3 (and its extension to ordinary differential equations for part (i)).

Elements of Graph Theory

Graph theory provides key concepts to model, analyze and design network systems and distributed algorithms; the language of graphs pervades modern science and technology and is therefore essential.

3.1 Graphs and digraphs

[Graphs] An *undirected graph* (in short, a *graph*) consists of a set V of elements called *vertices* and of a set E of unordered pairs of vertices, called *edges*. For $u, v \in V$ and $u \neq v$, the set $\{u, v\}$ denotes an unordered edge. We define and visualize some basic examples graphs in Figure 3.1.

[Neighbors and degrees in graphs] Two vertices u and v of a given graph are *neighbors* if $\{u, v\}$ is an undirected edge. Given a graph G, we let $\mathcal{N}_G(v)$ denote the set of neighbors of v.

The *degree* of v is the number of neighbors of v. A graph is *regular* if all the nodes have the same degree; e.g., in Figure 3.1, the cycle graph is regular with degree 2 whereas the complete bipartite graph $K(3,3)$ and the Petersen graph are regular with degree 3.

[Digraphs and self-loops] A *directed graph* (in short, a *digraph*) of order n is a pair $G = (V, E)$, where V is a set with n elements called *vertices* (or *nodes*) and E is a set of ordered pairs of vertices called *edges*. In other words, $E \subseteq V \times V$. As for graphs, V and E are the *vertex set* and *edge set*, respectively. For $u, v \in V$, the ordered pair (u, v) denotes an edge *from u to v*. A digraph is *undirected* if $(v, u) \in E$ anytime $(u, v) \in E$. In a digraph, a *self-loop* is an edge from a node to itself. Consistently with a customary convention, self-loops are not allowed in graphs. We define and visualize some basic examples digraphs in Figure 3.2.

(a) *Path graph* with 6 nodes, denoted by P_6

(b) *Cycle (ring) graph* with 6 nodes, denoted by C_6

(c) *Star graph* with 6 nodes, denoted by S_6

(d) *Complete graph* with 6 nodes, denoted by K_6

(e) *Complete bipartite graph* with $3+3$ nodes, denoted by $K_{3,3}$

(f) A Cartesian grid graph

(g) *Petersen graph*
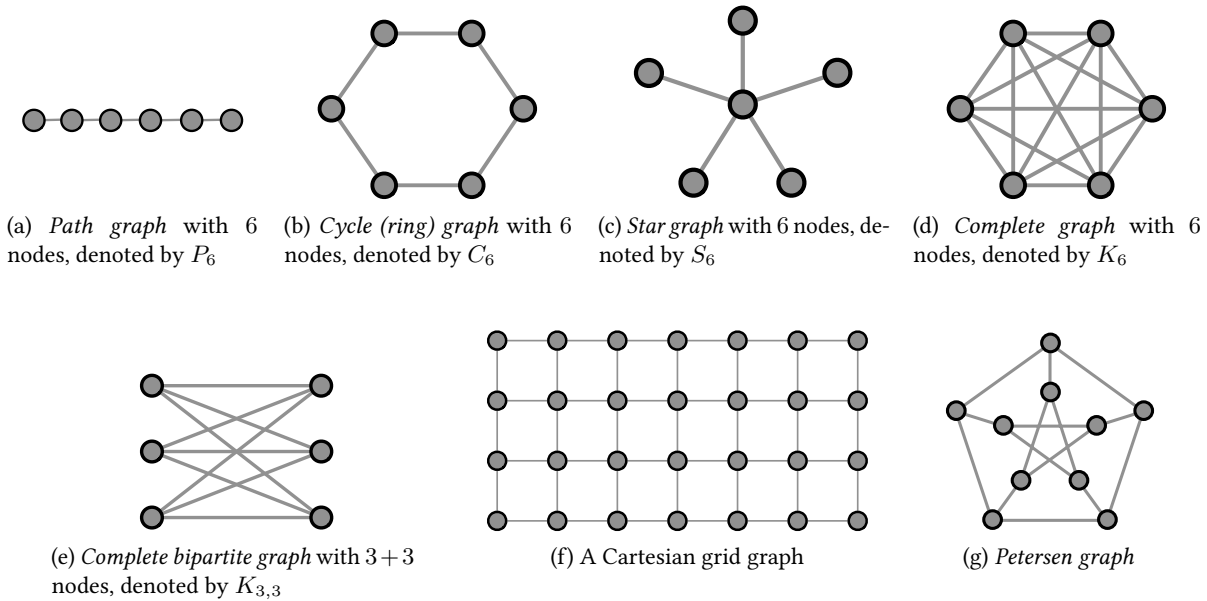
Figure 3.1: Example graphs. *Path graph:* nodes are ordered in a sequence and edges connect subsequent nodes in the sequence. *Cycle (or ring) graph:* all nodes and edges can be arranged as the vertices and edges of a regular polygon. *Star graph:* edges connect a specific node, called the *center*, to all other nodes. *Complete graph:* every pair of nodes is connected by an edge. *Complete bipartite graph:* nodes are divided into two sets and every node of the first set is connected with every node of the second set.

(a) *Cycle digraph* with 6 nodes

(b) *Complete digraph* with 6 nodes

(c) A digraph with no directed cycles
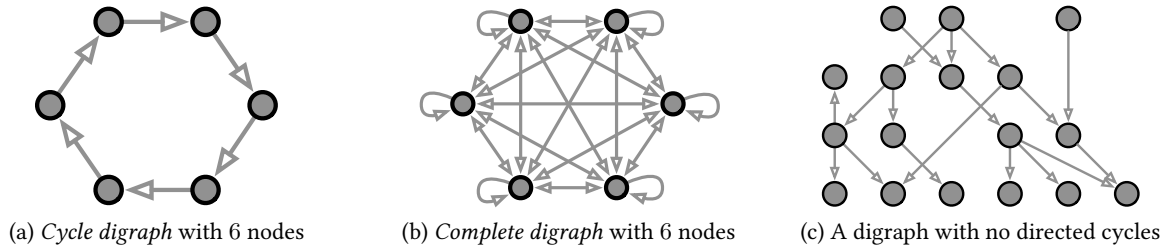
Figure 3.2: Example digraphs

[Subgraphs] A digraph (V', E') is a *subgraph* of a digraph (V, E) if $V' \subseteq V$ and $E' \subseteq E$. A digraph (V', E') is a *spanning subgraph* of (V, E) if it is a subgraph and $V' = V$. The subgraph of (V, E) *induced by* $V' \subseteq V$ is the digraph (V', E'), where E' contains all edges in E between two vertices in V'.

[In- and out-neighbors] In a digraph G with an edge $(u, v) \in E$, u is called an *in-neighbor* of v, and v is called an *out-neighbor* of u. We let $\mathcal{N}^{\text{in}}(v)$ (resp., $\mathcal{N}^{\text{out}}(v)$) denote the set of in-neighbors, (resp. the set of out-neighbors) of v. Given a digraph $G = (V, E)$, an *in-neighbor* of a nonempty set of nodes U is a node $v \in V \setminus U$ for which there exists an edge $(v, u) \in E$ for some $u \in U$.

[In- and out-degree] The *in-degree* $d_{\mathrm{in}}(v)$ and *out-degree* $d_{\mathrm{out}}(v)$ of v are the number of in-neighbors and out-neighbors of v, respectively. Note that a self-loop at a node v makes v both an in-neighbor as well as an out-neighbor of itself. A digraph is *topologically balanced* if each vertex has the same in- and out-degrees (even if distinct vertices have distinct degrees).

3.2   Paths and connectivity in undirected graphs

[Paths] A *path* in a graph is an ordered sequence of vertices such that any pair of consecutive vertices in the sequence is an edge of the graph. A path is *simple* if no vertex appears more than once in it, except possibly for the case in which the initial vertex is the same as the final vertex. (Note: some authors adopt the term "walk" to refer to what we call here path.)

[Connectivity and connected components] A graph is *connected* if there exists a path between any two vertices. If a graph is not connected, then it is composed of multiple *connected components*, that is, multiple connected subgraphs.

[Cycles] A *cycle* is a simple path that starts and ends at the same vertex and has at least three distinct vertices. A graph is *acyclic* if it contains no cycles. A connected acyclic graph is a *tree*.
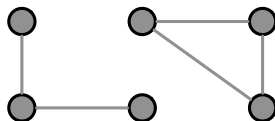
Figure 3.3: This graph has two connected components. The leftmost connected component is a tree, while the rightmost connected component is a cycle.

3.3   Paths and connectivity in digraphs

[Directed paths] A *directed path* in a digraph is an ordered sequence of vertices such that any pair of consecutive vertices in the sequence is a directed edge of the digraph. A directed path is *simple* if no vertex appears more than once in it, except possibly for the initial and final vertex.

[Cycles in digraphs] A *cycle* in a digraph is a simple directed path that starts and ends at the same vertex. It is customary to accept, as feasible cycles in digraphs, also cycles of length 1 (that is, a self-loop) and cycles of length 2 (that is, composed of just 2 nodes). The set of cycles of a directed graph is finite. A digraph is *acyclic* if it contains no cycles.

[Sources and sinks] In a digraph, every vertex with in-degree 0 is called a *source*, and every vertex with out-degree 0 is called a *sink*. Every acyclic digraph has at least one source and at least one sink; see Figure 3.4 and Exercise E3.1.

(a) An acyclic digraph with one sink and two sources   (b) A directed cycle
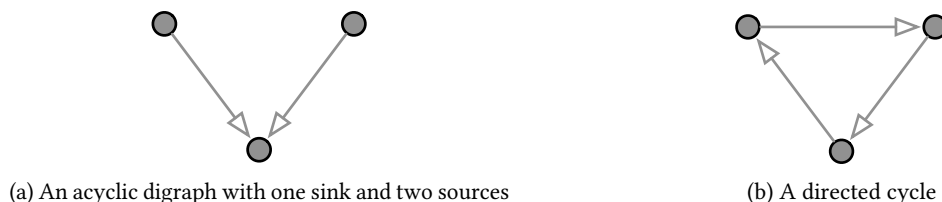
Figure 3.4: Examples of sources and sinks

[Directed trees] A *directed tree* (sometimes called a *rooted tree*) is an acyclic digraph with the following property: there exists a vertex, called the *root*, such that any other vertex of the digraph can be reached by one and only one directed path starting at the root. A *directed spanning tree* of a digraph is a spanning subgraph that is a directed tree.

3.3.1 Connectivity properties of digraphs

Next, we present four useful connectivity notions for a digraph G:

(i) G is *strongly connected* if there exists a directed path from any node to any other node;

(ii) G is *weakly connected* if the undirected version of the digraph is connected;

(iii) G possesses a *globally reachable node* if one of its nodes can be reached from any other node by traversing a directed path; and

(iv) G possesses a *directed spanning tree* if one of its nodes is the root of directed paths to every other node.

These notions are illustrated in Figure 3.5.

(a) A strongly connected digraph   (b) A weakly connected digraph with a globally reachable node
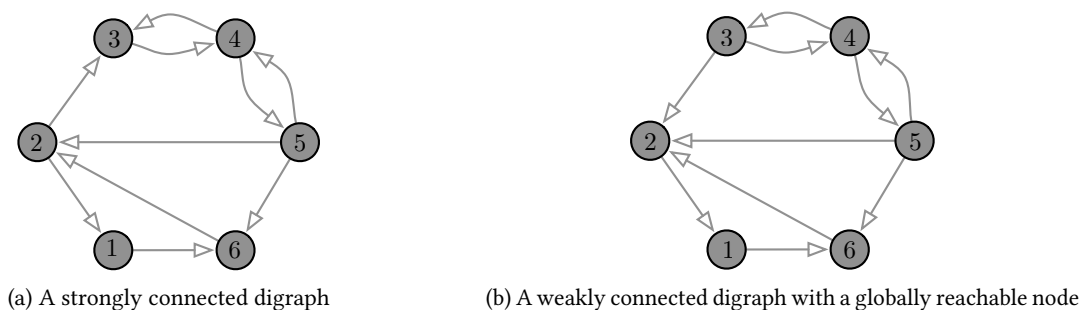
Figure 3.5: Connectivity examples for digraphs

For a digraph $G = (V, E)$, the *reverse digraph* $G(\text{rev})$ has vertex set V and edge set $E(\text{rev})$ composed of all edges in E with reversed direction. Clearly, a digraph contains a directed spanning tree if and only if the reverse digraph contains a globally reachable node.

3.3.2 Periodicity of strongly-connected digraphs

[Periodic and aperiodic digraphs] A strongly-connected directed graph is *periodic* if there exists a $k > 1$, called the *period*, that divides the length of every cycle of the graph. In other words, a digraph is periodic if the greatest common divisor of the lengths of all its cycles is larger than one. A digraph is *aperiodic* if it is not periodic.

(a) A periodic digraph with period 2 (b) An aperiodic digraph with cycles of length 1 and 2. (c) An aperiodic digraph with cycles of length 2 and 3.
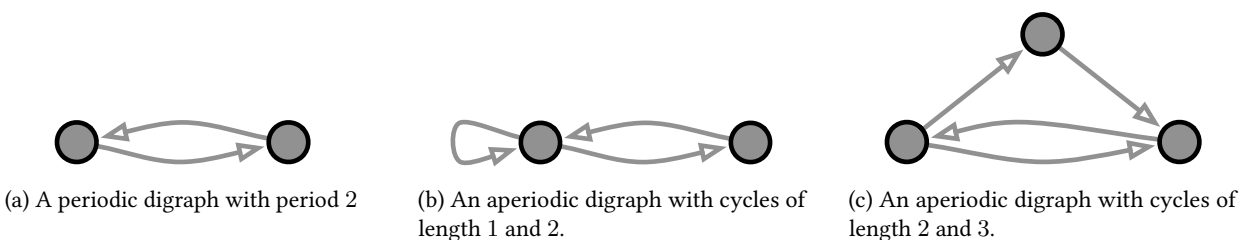
Figure 3.6: Example periodic and aperiodic digraphs.

Note: the definition of periodic digraph is well-posed because a digraph has only a finite number of cycles (because of the assumptions that nodes are not repeated in simple paths). The notions of periodicity and aperiodicity only apply to digraphs and not to undirected graphs (where the notion of a cycle is defined differently). Any strongly-connected digraph with a self-loop is aperiodic.

3.3.3 Condensation digraphs

[Strongly connected components] A subgraph H is a *strongly connected component* of G if H is strongly connected and any other subgraph of G strictly containing H is not strongly connected.

[Condensation digraph] The *condensation digraph* of a digraph G, denoted by $C(G)$, is defined as follows: the nodes of $C(G)$ are the strongly connected components of G, and there exists a directed edge in $C(G)$ from node H_1 to node H_2 if and only if there exists a directed edge in G from a node of H_1 to a node of H_2. The condensation digraph has no self-loops.

(a) An example digraph G (b) The strongly connected components of G (c) The condensation $C(G)$
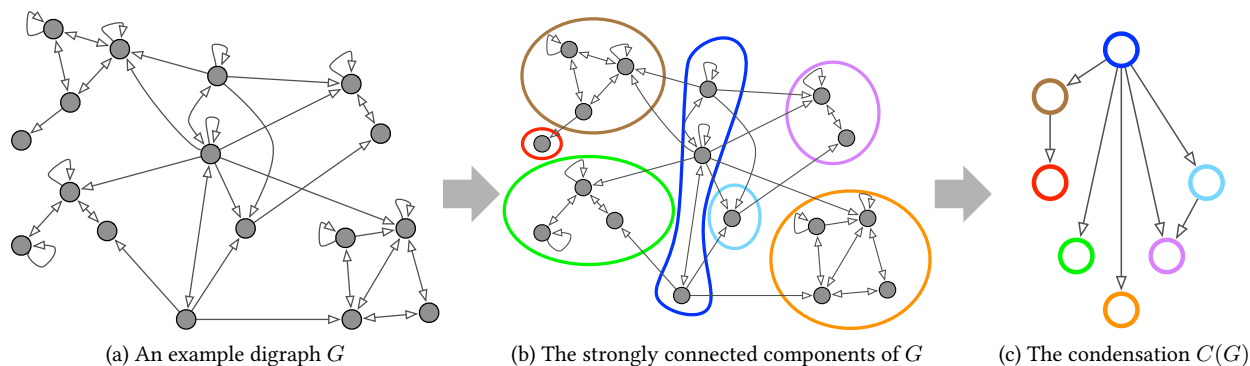
Figure 3.7: An example digraph, its strongly connected components and its condensation digraph.

Lemma 3.1 (Properties of the condensation digraph). *For a digraph G and its condensation digraph $C(G)$,*

(i) *$C(G)$ is acyclic,*

(ii) *G is weakly connected if and only if $C(G)$ is weakly connected, and*

(iii) *the following statement are equivalent:*

    a) *G contains a globally reachable node,*

    b) *$C(G)$ contains a globally reachable node, and*

    c) *$C(G)$ contains a unique sink.*

Proof. We prove statement (i) by contradiction. If there exists a cycle $(H_1, H_2, \ldots, H_m, H_1)$ in $C(G)$, then the set of vertices $H_1, \ldots, H_m$ are strongly connected in $C(G)$. But this implies that also the subgraph of G containing all node of $H_1, \ldots, H_m$ is strongly connected in G. But this is a contradiction with the fact that any subgraph of G strictly containing any of the $H_1, \ldots, H_m$ must be not strongly connected. Statement (ii) is intuitive and simple to prove; we leave this task to the reader.

Regarding statement (iii), we start by proving that (iii)a $\implies$ (iii)b. Let v be a a globally reachable node in G and let H denote the node in $C(G)$ containing v. Pick an arbitrary node $\bar{H}$ of $C(G)$ and let $\bar{v}$ be a node of G in $\bar{H}$. Since v is globally reachable, there exists a directed path from $\bar{v}$ to v in G. This directed path induces naturally a directed path in $C(G)$ from $\bar{H}$ to H. This shows that H is a globally reachable node in $C(G)$.

Regarding (iii)b $\implies$ (iii)a, let H be a globally reachable node of $C(G)$ and pick a node v in H. We claim v is globally reachable in G. Indeed, pick any node $\bar{v}$ in G belonging to a strongly connected component $\bar{U}$ of G. Because H is globally reachable in $C(G)$, there exists a directed path of the form $\bar{H} = H_0, H_1, \ldots, H_k, H_{k+1} = H$ in $C(G)$. One can now piece together a directed path in G from $\bar{v}$ to v, by walking inside each of the strongly connected components H_i and moving to the subsequent strongly connected components H_{i+1}, for $i \in \{0, \ldots, k\}$.

The final equivalence between statement (iii)b and statement (iii)c is an immediate consequence of $C(G)$ being acyclic. ∎

3.4   Weighted digraphs

A *weighted digraph* is a triplet $G = (V, E, \{a_e\}_{e \in E})$, where the pair (V, E) is a digraph with nodes $V = \{v_1, \ldots, v_n\}$, and where $\{a_e\}_{e \in E}$ is a collection of strictly positive weights for the edges E.

Note: for simplicity we let $V = \{1, \ldots, n\}$. It is therefore equivalent to write $\{a_e\}_{e \in E}$ or $\{a_{ij}\}_{(i,j) \in E}$.
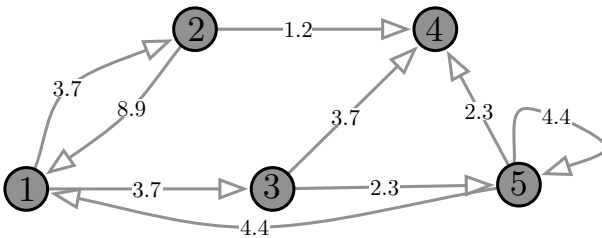
The set of weights for this weighted digraph is

$$a_{12} = 3.7, \ a_{13} = 3.7, \ a_{21} = 8.9,$$
$$a_{24} = 1.2, \ a_{34} = 3.7, \ a_{35} = 2.3,$$
$$a_{51} = 4.4, \ a_{54} = 2.3, \ a_{55} = 4.4.$$

A digraph $G = (V = \{v_1, \ldots, v_n\}, E)$ can be regarded as a weighted digraph by defining its set of weights to be all equal to 1, that is, setting $a_e = 1$ for all $e \in E$. A weighted digraph is *undirected* if $a_{ij} = a_{ji}$ for all $i, j \in \{1, \ldots, n\}$.

The notions of connectivity and definitions of in- and out-neighbors, introduced for digraphs, remain equally valid for weighted digraphs. The notions of in- and out-degree are generalized to weighted digraphs as follows. In a weighted digraph with $V = \{v_1, \ldots, v_n\}$, the *weighted out-degree* and the *weighted in-degree* of vertex v_i are defined by, respectively,

$$d_{\text{out}}(v_i) = \sum_{j=1}^{n} a_{ij}, \qquad \text{(i.e., } d_{\text{out}}(v_i) \text{ is the sum of the weights of all the out-edges of } v_i\text{)} ,$$

$$d_{\text{in}}(v_i) = \sum_{j=1}^{n} a_{ji}, \qquad \text{(i.e., } d_{\text{in}}(v_i) \text{ is the sum of the weights of all the in-edges of } v_i\text{)} .$$

The weighted digraph G is *weight-balanced* if $d_{\text{out}}(v_i) = d_{\text{in}}(v_i)$ for all $v_i \in V$.

3.5   Appendix: Database collections and software libraries

Useful collections of example networks are freely available online; here are some examples:

 (i) The Koblenz Network Collection, available at http://konect.uni-koblenz.de and described in (Kunegis, 2013), contains model graphs in easily accessible Matlab format (as well as a Matlab toolbox for network analysis and a compact overview the various computed statistics and plots for the networks in the collection).

 (ii) A broad range of example networks is available online at the Stanford Large Network Dataset Collection, see http://snap.stanford.edu/data.

 (iii) The The SuiteSparse Matrix Collection (formerly known as the University of Florida Sparse Matrix Collection), available at http://suitesparse.com and described in (Davis and Hu, 2011), contains a large and growing set of sparse matrices and complex graphs arising in a broad range of applications; e.g., see Figure 3.8.

(iv) The UCI Network Data Repository, available at `http://networkdata.ics.uci.edu`, is an effort to facilitate the scientific study of networks; see also (DuBois, 2008).

(a) IEEE 118 bus system

(b) Klavzar bibliography

(c) Pajek network GD99c
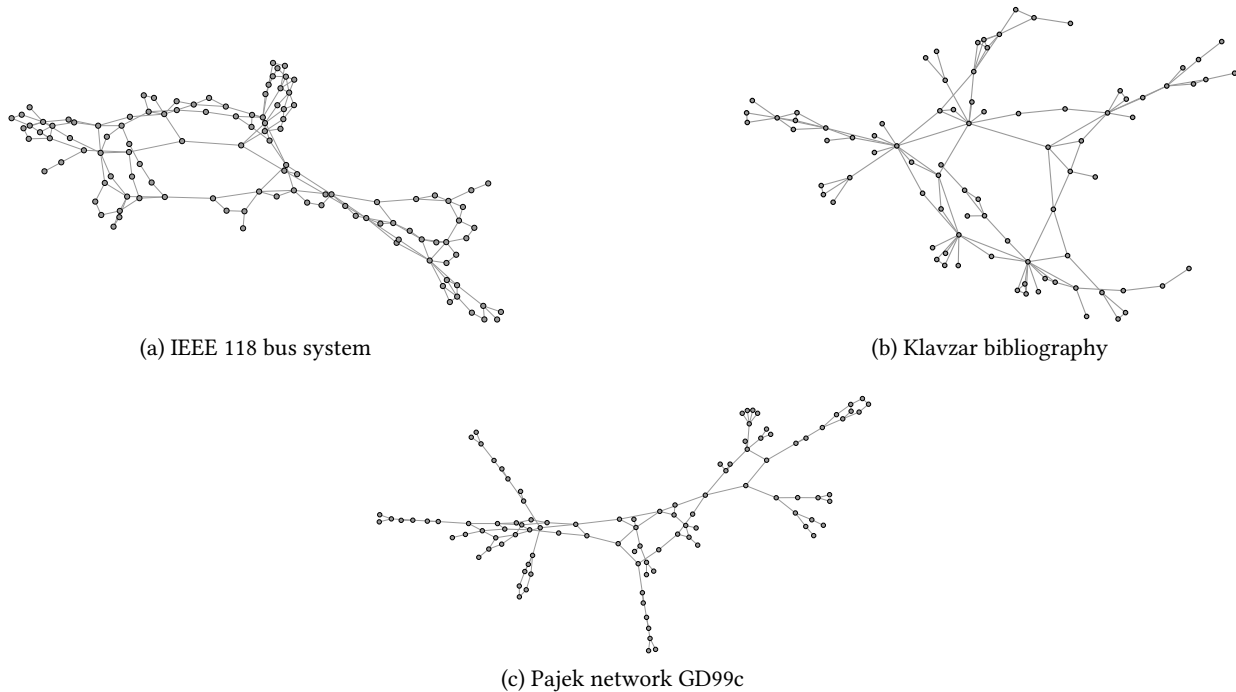
Figure 3.8: Example networks from distinct domains: Figure 3.8a shows the standard IEEE 118 power grid testbed (118 nodes); Figure 3.8b shows the Klavzar bibliography network (86 nodes); Figure 3.8c shows the GD99c Pajek network (105 nodes). Networks parameters are available at `http://suitesparse.com`.

Useful software libraries for network analysis and visualization are freely available online; here are some examples:

(i) NetworkX, available at `http://networkx.github.io`, is a Python library for network analysis. For example, one feature is the ability to compute condensation digraphs.

(ii) Gephi, available at `https://gephi.org`, is an interactive visualization and exploration platform for all kinds of networks and complex systems, dynamic and hierarchical graphs.

(iii) Cytoscape, available at `http://www.cytoscape.org`, is an open-source software platform for visualizing complex networks and integrating them with attribute data.

(iv) Mathematica provides functionality for modeling, analyzing, synthesizing, and visualizing graphs and networks – beside the ability to simulate dynamical systems; see description at `http://www.wolfram.com/language/elementary-introduction/21-graphs-and-networks.html`.

(v) Graphviz, available at `http://www.graphviz.org/`, is an open source graph visualization software which is also compatible with Matlab:
`http://www.mathworks.com/matlabcentral/fileexchange/4518-matlab-graphviz-interface`.

3.6 Historical notes and further reading

Paraphrasing from Chapter 1 "Discovery!" in the classic work by Harary (1969),

> (Euler, 1741) became the father of graph theory as well as topology when he settled a famous unsolved problem of his day called the Königsberg Bridge Problem.

> Subsequent rediscoveries of graph theory by Kirchhoff (1847) and Cayley (1857) also had their roots in the physical world. Kirchhoff's investigations of electric networks led to his development of the basic concepts and theorems concerning trees in graphs, while Cayley considered trees arising from the enumeration of organic chemical isomers.

For modern comprehensive treatments we refer the reader to standard books in graph theory such as (Diestel, 2000; Bollobás, 1998).

A classic reference in graph drawing is (Fruchterman et al., 1991), the layout of the three graphs in Figure 3.8 is obtained via the algorithm proposed by Hu (2005).

3.7   Exercises

E3.1   **Acyclic digraphs.** Let G be an acyclic digraph with nodes $\{1, \ldots, n\}$. A *topological sort* of G is a re-numbering of the vertices of G with the property that, if (u, v) is an edge of G, then $u > v$.

    (i) Show that G contains at least one sink, i.e., a vertex without out-neighbors and at least one source, i.e., a vertex without in-neighbors.

    (ii) Provide an algorithm to perform a topological sort of G. Is the topological sort unique?

        Hint: Use high-level pseudo-code instructions such as "select a node satisfying property A" or "remove all edges satisfying property B."

    (iii) Show that, after topologically sorting the vertices of G, the adjacency matrix of G is lower-triangular, i.e., all its entries above the main diagonal are equal to zero.

E3.2   **Condensation digraphs.** Draw the condensation for each of the following digraphs.
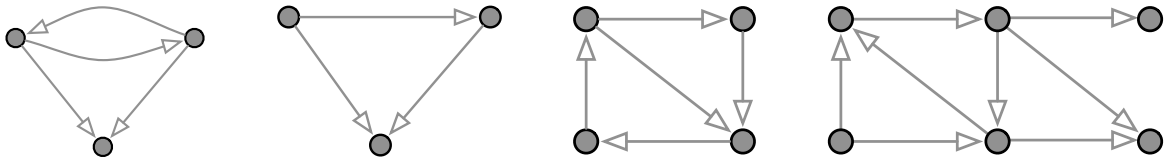
E3.3   **Directed spanning trees in the condensation digraph.** For a digraph G and its condensation digraph $C(G)$, show that the following statements are equivalent:

    (i) G contains a directed spanning tree, and

    (ii) $C(G)$ contains a directed spanning tree.

E3.4   **Properties of trees.** Consider an undirected graph G with n nodes and m edges (and without self-loops). Show that the following statements are equivalent:

    (i) G is a tree;

    (ii) G is connected and $m = n - 1$; and

    (iii) G is acyclic and $m = n - 1$.

E3.5   **Connectivity in topologically balanced digraphs.** Prove the following statement: If a digraph G is topologically balanced and contains either a globally reachable vertex or a directed spanning tree, then G is strongly connected.

E3.6   **Globally reachable nodes and disjoint closed subsets (Lin et al., 2005; Moreau, 2005).** Consider a digraph $G = (V, E)$ with at least two nodes. Prove that the following statements are equivalent:

    (i) G has a globally reachable node, and

    (ii) for every pair S_1, S_2 of non-empty disjoint subsets of V, there exists a node that is an out-neighbor of S_1 or S_2.

E3.7   **Swiss railroads.** Consider the fictitious railroad map of Switzerland given in Figure E3.1.

    (i) Can a passenger go from any station to any other?

    (ii) Is the graph acyclic? Is it aperiodic? If not, what is its period?
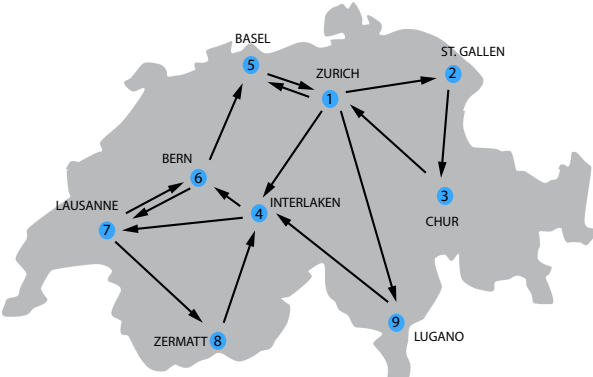
Figure E3.1: Fictitious railroad map connections in Switzerland

The Adjacency Matrix

In this chapter we present results on the adjacency matrices as part of the broader field of algebraic graph theory. The key results in this area relate, through necessary and sufficient conditions, matrix properties with graphical properties. For example, we will show how a matrix is primitive if and only if its associated digraph is strongly connected and aperiodic.

4.1 The adjacency matrix

Given a weighted digraph $G = (V, E, \{a_e\}_{e \in E})$, with $V = \{1, \ldots, n\}$, the *weighted adjacency matrix* of G is the $n \times n$ non-negative matrix A defined as follows: for each edge $(i, j) \in E$, the entry (i, j) of A is equal to the weight $a_{(i,j)}$ of the edge (i, j), and all other entries of A are equal to zero. In other words, $a_{ij} > 0$ if and only if (i, j) is an edge of G, and $a_{ij} = 0$ otherwise. Figure 4.1 shows a example of a weighted digraph.

$$A = \begin{bmatrix} 0 & 3.7 & 3.7 & 0 & 0 \\ 8.9 & 0 & 0 & 1.2 & 0 \\ 0 & 0 & 0 & 3.7 & 2.3 \\ 0 & 0 & 0 & 0 & 0 \\ 4.4 & 0 & 0 & 2.3 & 4.4 \end{bmatrix}.$$
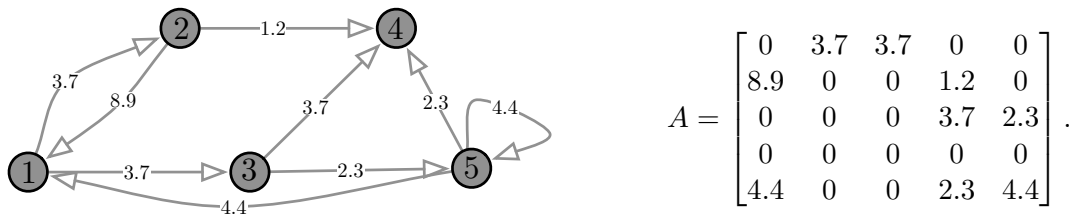
Figure 4.1: A weighted digraph and its adjacency matrix

The *binary adjacency matrix* $A \in \{0, 1\}^{n \times n}$ of a digraph $G = (V = \{1, \ldots, n\}, E)$ or of a weighted digraph is defined by

$$a_{ij} = \begin{cases} 1, & \text{if } (i, j) \in E, \\ 0, & \text{otherwise.} \end{cases} \tag{4.1}$$

Here, a binary matrix is any matrix with entries taking values in $0, 1$.

Finally, the *weighted out-degree matrix* D_{out} and the *weighted in-degree matrix* D_{in} of a weighted digraph are the diagonal matrices defined by

$$
D_{\text{out}} = \operatorname{diag}(A\mathbb{1}_n) = \begin{bmatrix} d_{\text{out}}(1) & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & d_{\text{out}}(n) \end{bmatrix}, \quad \text{and} \quad D_{\text{in}} = \operatorname{diag}(A^{\mathsf{T}}\mathbb{1}_n),
$$

where $\operatorname{diag}(z_1, \ldots, z_n)$ is the diagonal matrix with diagonal entries equal to $z_1, \ldots, z_n$.

We conclude this section with some basic examples.

Example 4.1 (Basic graphs and their adjacency matrices). Recall the definitions of path, cycle, star, complete and complete bipartite graph from Figure 3.1. Figure 4.2 illustrates their adjacency matrices.

$$P_6 \qquad\qquad C_6 \qquad\qquad S_6 \qquad\qquad K_6 \qquad\qquad K_{3,3}$$
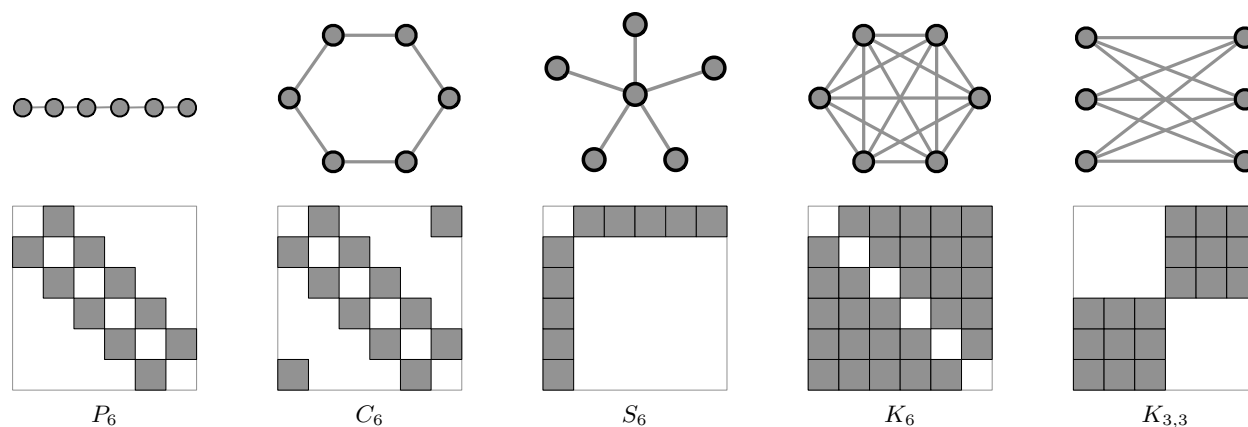
Figure 4.2: Path, cycle, star, complete and complete bipartite graph (from Figure 3.1) and their binary adjacency matrices

Note that the adjacency matrices of path and cycle graphs have a particular structure. An $n \times n$ matrix T is *Toeplitz* (also called *diagonal-constant*) if there exist scalar numbers $a_{-(n-1)}, \ldots, a_{-1}, a_0, a_1, \ldots, a_{(n-1)}$ such that

$$
T = \begin{bmatrix} a_0 & a_1 & \ldots & \ldots & a_{n-1} \\ a_{-1} & a_0 & a_1 & \ldots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ldots & a_{-1} & a_0 & a_1 \\ a_{-(n-1)} & \ldots & \ldots & a_{-1} & a_0 \end{bmatrix}.
$$

Two special cases are of interest, namely, those of tridiagonal Toeplitz and circulant matrices. For these two cases it is possible to compute eigenvalues and eigenvectors; we refer to Exercises E4.16 and E4.17 for more information. For here instead, we conclude with a table containing the *adjacency spectrum* of the basic graphs, i.e., the spectrum of their binary adjacency matrices.

Graph	Adjacency Matrix	Adjacency Spectrum
path graph P_n	Toeplitz tridiagonal	$\{2\cos(\pi i/(n+1)) \mid i \in \{1,\ldots,n\}\}$
cycle graph C_n	circulant	$\{2\cos(2\pi i/n)) \mid i \in \{1,\ldots,n\}\}$
star graph S_n	$\mathbb{e}_1 \mathbb{e}_{-1} + \mathbb{e}_{-1}\mathbb{e}_1$, where $\mathbb{e}_{-i} = \mathbb{1}_n - \mathbb{e}_i$	$\{\sqrt{n-1}, 0, \ldots, 0, -\sqrt{n-1}\}$
path graph K_n	$\mathbb{1}_n \mathbb{1}_n^{\mathsf{T}} - I_n$	$\{(n-1), -1, \ldots, -1\}$
complete bipartite $K_{n,m}$	$\begin{bmatrix} \mathbb{0}_{n\times n} & \mathbb{1}_{n\times m} \\ \mathbb{1}_{m\times n} & \mathbb{0}_{m\times m} \end{bmatrix}$	$\{\sqrt{nm}, 0, \ldots, 0, -\sqrt{nm}\}$

Table 4.1: Adjacency spectrum for basic graphs

We ask the reader to prove the statements in the table in Exercise .   ●

4.2 Algebraic graph theory: basic and prototypical results

In this section we review some basic and prototypical results that involve correspondences between graphs and adjacency matrices.

In what follows we let G denote a weighted digraph and A its weighted adjacency matrix or, equivalently, we let A be a non-negative matrix and G be its *associated weighted digraph* (i.e., the digraph with nodes $\{1,\ldots,n\}$ and with weighted adjacency matrix A). We start with some straightforward statements:

(i) G is undirected if and only if A is symmetric and its diagonal entries are equal to 0;

(ii) G is weight-balanced if and only if $A\mathbb{1}_n = A^{\mathsf{T}}\mathbb{1}_n$, i.e., $D_{\text{out}} = D_{\text{in}}$;

(iii) in a digraph G without self-loops, the node i is a sink in G if and only if ith row-sum of A is zero;

(iv) in a digraph G without self-loops, the node i is a source in G if and only if ith column-sum of A is zero;

(v) A is row-stochastic if and only if each node of G has weighted out-degree equal to 1 (so that $D_{\text{out}} = I_n$); and

(vi) A is doubly-stochastic if and only if each node of G has weighted out-degree and weighted in-degree equal to 1 (so that $D_{\text{out}} = D_{\text{in}} = I_n$ and, in particular, G is weight-balanced).

Next, we relate the powers of the adjacency matrix with the existence of directed paths in the digraph. We start with some simple observation. First, pick two nodes i and j and note that there exists a directed path from i to j of length 1 (i.e., an edge) if and only if $(A)_{ij} > 0$. Next, consider the formula for the matrix power:

$$(A^2)_{ij} = (i\text{th row of } A) \cdot (j\text{th column of } A) = \sum_{h=1}^{n} A_{ih} A_{hj}.$$

A directed path from i to j of length 2 exists if and only if there exists a node k such that (i,k) and (k,j) are edges of G. In turn, (i,k) and (k,j) are edges if and only if $A_{ik} > 0$ and $A_{kj} > 0$ and therefore $(A^2)_{ij} > 0$. In short, we know that a directed path from i to j of length 2 exists if and only if $(A^2)_{ij} > 0$. These observations lead to the following result, whose proof we leave as Exercise .

Lemma 4.2 (Directed paths and powers of the adjacency matrix). *Let G be a weighted digraph with n nodes, with weighted adjacency matrix A, with unweighted adjacency matrix $A_{0,1} \in \{0,1\}^{n \times n}$, and possibly with self-loops. For all $i, j \in \{1, \ldots, n\}$ and $k \in \mathbb{N}$*

(i) *the (i, j) entry of $A_{0,1}^k$ equals the number of directed paths of length k (including paths with self-loops) from node i to node j; and*

(ii) *the (i, j) entry of A^k is positive if and only if there exists a directed path of length k (including paths with self-loops) from node i to node j.*

4.3 Graph theoretical characterization of irreducible matrices

In this section we provide three equivalent characterizations of the notion of irreducibility an we can now characterize certain connectivity properties of digraphs based on the powers of the adjacency matrix.

Before proceeding, we introduce a few useful concepts. First, $\{\mathcal{I}, \mathcal{J}\}$ is a *partition* of the index set $\{1, \ldots, n\}$ if $\mathcal{I} \cup \mathcal{J} = \{1, \ldots, n\}, \mathcal{I} \neq \emptyset, \mathcal{J} \neq \emptyset$, and $\mathcal{I} \cap \mathcal{J} = \emptyset$. Second, a *permutation matrix* is a square binary matrix with precisely one entry equal to 1 in every row and every columns. (In other words, the columns of a permutation matrix are a reordering of the basis vectors $e_1, \ldots, e_n$; a permutation matrix acts on a vector by permuting its entries.) Finally, an $n \times n$ matrix A is *block triangular* if there exists $r \in \{1, \ldots, n-1\}$ such that

$$A = \left[\begin{array}{c|c} B & C \\ \hline \mathbb{0}_{(n-r) \times r} & D \end{array} \right],$$

where $B \in \mathbb{R}^{r \times r}, C \in \mathbb{R}^{r \times (n-r)}$ and $D \in \mathbb{R}^{(n-r) \times (n-r)}$ are arbitrary.

We are now ready to state the main result of this section.

Theorem 4.3 (Connectivity properties of the digraph and positive powers of the adjacency matrix). *Let G be a weighted digraph with $n \geq 2$ nodes and with weighted adjacency matrix A. The following statements are equivalent:*

(i) *A is irreducible, that is, $\sum_{k=0}^{n-1} A^k > 0$;*

(ii) *there exists no permutation matrix P such that $P^{\mathsf{T}} A P$ is block triangular;*

(iii) *G is strongly connected;*

(iv) *for all partitions $\{\mathcal{I}, \mathcal{J}\}$ of the index set $\{1, \ldots, n\}$, there exists $i \in \mathcal{I}$ and $j \in \mathcal{J}$ such that (i, j) is a directed edge in G.*

Note: as the theorem establishes, there are four equivalent characterizations of irreducibility. In the literature, it is common to define irreducibility through property (ii) or (iv). We next see two simple examples.
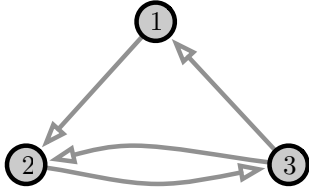
This digraph is strongly connected and, accordingly, its adjacency matrix is irreducible:

$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}.$$
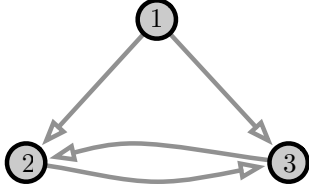
This digraph is not strongly connected (vertices 2 and 3 are globally reachable, but 1 is not) and, accordingly, its adjacency matrix is reducible:

$$\begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}.$$

Proof of Theorem 4.3. Regarding (iii) $\implies$ (iv), pick a partition $\{\mathcal{I}, \mathcal{J}\}$ of the index set $\{1, \ldots, n\}$ and two nodes $i_0 \in \mathcal{I}$ and $j_0 \in \mathcal{J}$. By assumptions there exists a directed path from i_0 to j_0. Hence there must exist an edge from a node in $\mathcal{I}$ to a node in $\mathcal{J}$.

Regarding (iv) $\implies$ (iii), pick a node $i \in \{1, \ldots, n\}$ and let $R_i \subset \{1, \ldots, n\}$ be the set of nodes reachable from i, i.e., the set of nodes that belong to directed paths originating from node i. Denote the unreachable nodes by $U_i = \{1, \ldots, n\} \setminus R_i$. Second, by contradiction, assume U_i is not empty. Then $R_i \cup U_i$ is a partition of the index set $\{1, \ldots, n\}$ and irreduciblity implies the existence of a non-zero entry a_{jh} with $j \in R_i$ and $h \in U_i$. But then the node h is reachable. Therefore, $U_i = \emptyset$, and all nodes are reachable from i.

Regarding (iii) $\implies$ (i), because G is strongly connected, there exists a directed path of length k' connecting node i to node j, for all i and j. By removing any cycle from such a path (so that no intermediate node is repeated), one can compute a path from i to j of length $k < n$. Hence, by Lemma 4.2(ii), the entry $(A^k)_{ij}$ is strictly positive and, in turn, so is the entire matrix sum $\sum_{k=0}^{n-1} A^k$.

Regarding (i) $\implies$ (iii), pick two nodes i and j. Because $\sum_{k=0}^{n-1} A^k > 0$, there must exists k such that $(A^k)_{ij} > 0$. Lemma 4.2(ii) implies the existence of a path of length k from i to j. Hence, G is strongly connected.

Regarding (ii) $\implies$ (iv), by contradiction, assume there exists a partition $(\mathcal{I}, \mathcal{J})$ of $\{1, \ldots, n\}$ such that $a_{ij} = 0$ for all $(i, j) \in \mathcal{I} \times \mathcal{J}$. Let $\pi : \{1, \ldots, n\} \to \{1, \ldots, n\}$ be the permutation that maps all entries of $\mathcal{I}$ into the first $|\mathcal{I}|$ entries of $\{1, \ldots, n\}$. Here we let $|\mathcal{I}|$ denote the number of elements of $\mathcal{I}$. Let P be the corresponding permutation matrix. We now compute PAP^T and block partition it as:

$$PAP^\mathsf{T} = \begin{bmatrix} A_{\mathcal{I}\mathcal{I}} & A_{\mathcal{I}\mathcal{J}} \\ A_{\mathcal{J}\mathcal{I}} & A_{\mathcal{J}\mathcal{J}} \end{bmatrix},$$

where $A_{\mathcal{I}\mathcal{I}} \in \mathbb{R}^{|\mathcal{I}| \times |\mathcal{I}|}$, $A_{\mathcal{I}\mathcal{J}} \in \mathbb{R}^{|\mathcal{I}| \times |\mathcal{J}|}$, $A_{\mathcal{J}\mathcal{I}} \in \mathbb{R}^{|\mathcal{J}| \times |\mathcal{I}|}$, and $A_{\mathcal{J}\mathcal{J}} \in \mathbb{R}^{|\mathcal{J}| \times |\mathcal{J}|}$. By construction, $A_{\mathcal{J}\mathcal{I}} = \mathbb{0}_{|\mathcal{J}| \times |\mathcal{I}|}$ so that PAP^T is block triangular, which is in contradiction with the assumed statement (ii).

Regarding (iv) $\implies$ (ii), by contradiction, assume there exists a permutation matrix P and a number $r < n$ such that

$$PAP^\mathsf{T} = \begin{bmatrix} B & C \\ \mathbb{0}_{(n-r) \times r} & D \end{bmatrix},$$

where the matrices $B \in \mathbb{R}^{r \times r}$, $C \in \mathbb{R}^{r \times (n-r)}$, and $D \in \mathbb{R}^{(n-r) \times (n-r)}$ are arbitrary. The permutation matrix P defines a unique permutation $\pi : \{1, \ldots, n\} \to \{1, \ldots, n\}$ with the property that the columns of P are $\mathbb{e}_{\pi(1)}, \ldots, \mathbb{e}_{\pi(n)}$. Let $\mathcal{J} = \{\pi(1), \ldots, \pi(r)\}$ and $\mathcal{I} = \{1, \ldots, n\} \setminus \mathcal{J}$. Then, by construction, for any pair $(i,j) \in \mathcal{I} \times \mathcal{J}$, we know $a_{ij} = 0$, which is in contradiction with the assumed statement (iv).   ∎

Next we present two results, whose proof are analogous to those of the previous theorem and left to the reader as an exercise.

Lemma 4.4 (Global reachability and powers of the adjacency matrix). *Let G be a weighted digraph with $n \geq 2$ nodes and weighted adjacency matrix A. For any $j \in \{1, \ldots, n\}$, the following statements are equivalent:*

  (i) *the jth node of G is globally reachable, and*

  (ii) *the jth column of $\sum_{k=0}^{n-1} A^k$ is positive.*

Next, we notice that if node j is reachable from node i via a path of length k and at least one node along that path has a self-loop, then node j is reachable from node i via paths of length k, $k+1$, $k+2$, and so on. This observation and the last lemma lead to the following corollary.

Corollary 4.5 (Connectivity properties of the digraph and positive powers of the adjacency matrix: cont'd). *Let G be a weighted digraph with n nodes, weighted adjacency matrix A and a self-loop at each node. The following statements are equivalent:*

  (i) *G is strongly connected; and*

  (ii) *A^{n-1} is positive, so that A is primitive.*

For any $j \in \{1, \ldots, n\}$, the following two statements are equivalent:

  (iii) *the jth node of G is globally reachable; and*

  (iv) *the jth column of A^{n-1} has positive entries.*

Finally, we conclude this section with a clarification.

Remark 4.6 (Similarity transformations defined by permutation matrices). *Note that $P^\mathsf{T} A P$ is the similarity transformation of A defined by P because the permutation matrix P satisfies $P^{-1} = P^\mathsf{T}$; see Exercise E2.14. Moreover, note that $P^\mathsf{T} A P$ is simply a reordering of rows and columns. For example, consider*

$$P = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \text{ with } P^\mathsf{T} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}. \text{ Note } P \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix} \text{ as well as } P^\mathsf{T} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix} \text{ and compute}$$

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \implies P^\mathsf{T} A P = \begin{bmatrix} a_{22} & a_{23} & a_{21} \\ a_{32} & a_{33} & a_{31} \\ a_{12} & a_{13} & a_{11} \end{bmatrix},$$

so that the entries of the 1st, 2nd and 3rd rows of A are mapped respectively to the 3rd, 1st and 2nd rows of $P^\mathsf{T} A P$ — and, at the same time, — the entries of the 1st, 2nd and 3rd columns of A are mapped respectively to the 3rd, 1st and 2nd columns of $P^\mathsf{T} A P$.   •

4.4 Graph theoretical characterization of primitive matrices

In this section we present the main result of this chapter, an immediate corollary and its proof; see also Figure 4.3.

Theorem 4.7 (Strongly connected and aperiodic digraph and primitive adjacency matrix). *Let G be a weighted digraph with $n \geq 2$ nodes and with weighted adjacency matrix A. The following two statements are equivalent:*

(i) *G is strongly connected and aperiodic; and*

(ii) *A is primitive, that is, there exists $k \in \mathbb{N}$ such that A^k is positive.*
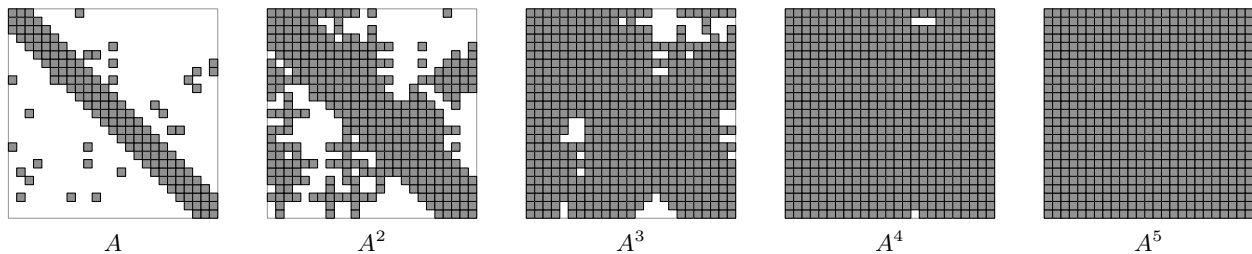
Figure 4.3: Increasing powers of a non-negative matrix $A \in \mathbb{R}^{25 \times 25}$. The digraph associated to A is strongly connected and has self-loops at each node; as predicted by Theorem 4.7, there exists $k = 5$ such that $A^k > 0$.

Before proving Theorem 4.7, we introduce a useful fact from number theory, whose proof we leave as Exercise E4.13. First, we recall a useful notion: a set of integers are *coprime* if its elements share no common positive factor except 1, that is, their greatest common divisor is 1. Loosely, the following lemma states that coprime numbers generate, via linear combinations with non-negative integer coefficients, all numbers larger than a given threshold.

Lemma 4.8 (Frobenius number). *Given a finite set $A = \{a_1, a_2, \ldots, a_n\}$ of positive coprime integers, an integer M is said to be* representable *by A if there exist non-negative integers $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ such that $M = \alpha_1 a_1 + \cdots + \alpha_N a_N$. The following statements are equivalent:*

(i) *there exists a finite largest unrepresentable integer, called the* Frobenius number *of A, and*

(ii) *the greatest common divisor of A is 1.*

Proof of Theorem 4.7. Regarding (i) $\implies$ (ii), pick any ordered pair (i, j). We claim that there exists a number $k(i, j)$ with the property that, for all $m > k(i, j)$, we have $(A^m)_{ij} > 0$, that is, there exists a directed path from i to j of length m for all $m \geq k(i, j)$. If this claim is correct, then the statement (ii) is proved with $k = \max\{k(i, j) \mid i, j \in \{1, \ldots, n\}\}$. To show this claim, let $\{c_1, \ldots, c_N\}$ be the set of the cycles of G and let $\{k_1, \ldots, k_N\}$ be their lengths. Because G is aperiodic, the lengths $\{k_1, \ldots, k_N\}$ are coprime and Lemma 4.8 implies the existence of a number $h(k_1, \ldots, k_N)$ such that any number larger than $h(k_1, \ldots, k_N)$ is a linear combination of $k_1, \ldots, k_N$ with non-negative integer as coefficients. Because G is strongly connected, there exists a path γ of arbitrary length $\Gamma(i, j)$ that starts at i, contains a vertex of each

of the cycles $c_1, \ldots, c_N$, and terminates at j. Now, we claim that $k(i, j) = \Gamma(i, j) + h(k_1, \ldots, k_N)$ has the desired property. Indeed, pick any number $m > k(i, j)$ and write it as $m = \Gamma(i, j) + \beta_1 k_1 + \cdots + \beta_N k_N$ for appropriate numbers $\beta_1, \ldots, \beta_N \in \mathbb{N}$. A directed path from i to j of length m is constructed by attaching to the path γ the following cycles: β_1 times the cycle c_1, β_2 times the cycle c_2, $\ldots$, β_N times the cycle c_N.

Regarding (ii) $\implies$ (i), from Lemma 4.2 we know that $A^k > 0$ means that there are paths of length k from every node to every other node. Hence, the digraph G is strongly connected. Next, we prove aperiodicity. Because G is strongly connected, each node of G has at least one outgoing edge, that is, for all i, there exists at least one index j such that $a_{ij} > 0$. This fact implies that the matrix $A^{k+1} = A A^k$ is positive via the following simple calculation: $(A^{k+1})_{il} = \sum_{h=1}^n a_{ih}(A^k)_{hl} \geq a_{ij}(A^k)_{jl} > 0$. In summary, if A^k is positive for some k, then A^m is positive for all subsequent $m > k$ (see also Exercise E2.5). Therefore, there are closed paths in G of any sufficiently large length. This fact implies that G is aperiodic; indeed, by contradiction, if the cycle lengths were not coprimes, then G would not possess such closed paths of arbitrary sufficiently large length.   ∎

4.5   Elements of spectral graph theory

In this section we provide some elementary results on the spectral radius of a non-negative matrix A. (We provide bounds on the eigenvalues of the Laplacian matrix in Section 6.1.2 and Exercise E6.3.) Recall that ith entry of the vector $A \mathbb{1}_n$ contains the ith row-sum of the matrix A and the out-degree of the ith node of the digraph associated to A. In other words, $d_{\text{out}}(i) = \mathbb{e}_i^\mathsf{T} A \mathbb{1}_n$.

Theorem 4.9 (Bounds on the spectral radius of a non-negative matrix). *For a non-negative $n \times n$ matrix A with associated digraph G, the following statements hold:*

  (i) $\min(A \mathbb{1}_n) \leq \rho(A) \leq \max(A \mathbb{1}_n)$*;*

 (ii) if $\min(A \mathbb{1}_n) = \max(A \mathbb{1}_n)$*, then* $\rho(A) = \max(A \mathbb{1}_n)$*; and*

(iii) if $\min(A \mathbb{1}_n) < \max(A \mathbb{1}_n)$*, then the following two statements are equivalent:*

> *a) for each node i with $\mathbb{e}_i^\mathsf{T} A \mathbb{1}_n = \max(A \mathbb{1}_n)$, there exists a directed path in G from node i to a node j with $\mathbb{e}_j^\mathsf{T} A \mathbb{1}_n < \max(A \mathbb{1}_n)$; and*
>
> *b)* $\rho(A) < \max(A \mathbb{1}_n)$*.*

An illustration of this result is given in Figure 4.4. Before providing the proof, we introduce a useful notion and establish a corollary.

Definition 4.10 (Row-substochastic matrix). *A non-negative $n \times n$ matrix A is* row-substochastic *if its row-sums are at most 1 and at least one row-sum is strictly less than 1, that is,*

$$A \mathbb{1}_n \leq \mathbb{1}_n, \text{ and there exists } i \in \{1, \ldots, n\} \text{ such that } \mathbb{e}_i^\mathsf{T} A \mathbb{1}_n < 1.$$

Note that a row-substochastic matrix with at least one row-sum equal to 1 satisfies $\min(A \mathbb{1}_n) < \max(A \mathbb{1}_n)$ and that any irreducible row-substochastic matrix satisfies condition (iii)a because the associated digraph is strongly connected. These two observations lead immediately to the following rewriting of the previous theorem.

(a) Complete bipartite graph $K(3,3)$ with binary adjacency matrix $A_{K(3,3)}$

(b) Cartesian grid graph with binary adjacency matrix A_{grid}
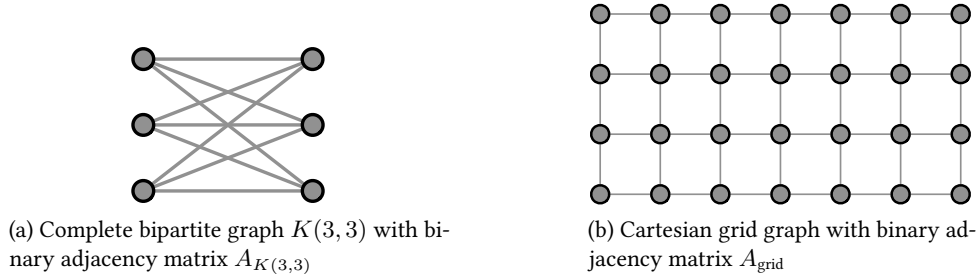
Figure 4.4: Illustration of Theorem 4.9: by counting the number of neighbors of each node (i.e., by computing the row sums of A) and observing that the grid graph is connected, we can establish that $\rho(A_{K(3,3)}) = 3$ and $2 < \rho(A_{\text{grid}}) < 4$.

Corollary 4.11 (Convergent row-substochastic matrices). *A row-substochastic matrix is convergent if and only if its associated digraph contains directed paths from each node with out-degree 1 to a node with out-degree less than 1. Specifically, an irreducible row-substochastic matrix is convergent.*

We now present the proof of the main theorem in this section.

Proof of Theorem 4.9. Regarding statement (i), the Perron–Frobenius Theorem 2.12 applied to the non-negative matrix A implies the existence of a vector $x \geq \mathbb{0}_n$, $x \neq \mathbb{0}_n$, such that

$$Ax = \rho(A)x \qquad \Longrightarrow \qquad \rho(A)x_i = \sum_{j=1}^{n} a_{ij}x_j \ \text{ for all } i \in \{1, \ldots, n\}.$$

Let $\ell \in \operatorname{argmax}_{i \in \{1, \ldots, n\}}\{x_i\}$ be the index (or one of the indices) satisfying $x_\ell = \max\{x_1, \ldots, x_n\} > 0$ and compute

$$\rho(A) = \sum_{j=1}^{n} a_{\ell j}\frac{x_j}{x_\ell} \leq \sum_{j=1}^{n} a_{\ell j} \leq \max(A\mathbb{1}_n).$$

We leave the proof of the lower bound to the reader in Exercise E4.9.

Regarding statement (ii), note that $\mathbb{1}_n$ is an eigenvector with eigenvalue $\max(A\mathbb{1}_n)$ so that we know $\rho(A) \geq \max(A\mathbb{1}_n)$. But we also know from statement (i) that $\rho(A) \leq \max(A\mathbb{1}_n)$.

Next, we establish that the condition (iii)a implies the bound (iii)b. It suffices to focus on row-substochastic matrices (if $\max(A\mathbb{1}_n) \neq 1$, we consider the row-substochastic matrix $A/\rho(A)$). We now claim that:

(1) if $\mathbb{e}_i^{\mathsf{T}} A\mathbb{1}_n < 1$, then $\mathbb{e}_i^{\mathsf{T}} A^2\mathbb{1}_n < 1$,

(2) if i has an outneighbor j (that is, $A_{ij} > 0$) with $\mathbb{e}_j^{\mathsf{T}} A\mathbb{1}_n < 1$, then $\mathbb{e}_i^{\mathsf{T}} A^2\mathbb{1}_n < 1$,

(3) there exists k such that $A^k\mathbb{1}_n < \mathbb{1}_n$, and

(4) $\rho(A) < 1$.

Regarding statement *(1)*, for a node i satisfying $\mathbb{e}_i^{\mathsf{T}} A\mathbb{1}_n < 1$, we compute

$$\mathbb{e}_i^{\mathsf{T}} A\mathbb{1}_n < 1 \quad \Longrightarrow \quad \mathbb{e}_i^{\mathsf{T}} A^2\mathbb{1}_n = \mathbb{e}_i^{\mathsf{T}} A(A\mathbb{1}_n) \leq \mathbb{e}_i^{\mathsf{T}} A\mathbb{1}_n < 1,$$

where we used the implication: if $\mathbb{0}_n \leq v \leq \mathbb{1}_n$ and $w \geq \mathbb{0}_n$, then $w^\mathsf{T} v \leq w^\mathsf{T} \mathbb{1}_n$. This proves statement *(1)*. Next, note that $0 \leq \mathbb{e}_j^\mathsf{T} A \mathbb{1}_n < 1$ and $A \mathbb{1}_n \leq \mathbb{1}_n$ together imply the useful inequality

$$A \mathbb{1}_n \leq \mathbb{1}_n - \big(1 - \mathbb{e}_j^\mathsf{T} A \mathbb{1}_n\big)\mathbb{e}_j, \qquad \text{because } 1 - \mathbb{e}_j^\mathsf{T} A \mathbb{1}_n > 0.$$

Therefore, we compute

$$
\begin{aligned}
\mathbb{e}_i^\mathsf{T} A^2 \mathbb{1}_n &= (\mathbb{e}_i^\mathsf{T} A)(A \mathbb{1}_n) \\
&\leq (\mathbb{e}_i^\mathsf{T} A)\Big(\mathbb{1}_n - \big(1 - \mathbb{e}_j^\mathsf{T} A \mathbb{1}_n\big)\mathbb{e}_j\Big) \\
&= \mathbb{e}_i^\mathsf{T} A \mathbb{1}_n - \big(1 - \mathbb{e}_j^\mathsf{T} A \mathbb{1}_n\big)\mathbb{e}_i^\mathsf{T} A \mathbb{e}_j \leq 1 - \big(1 - \mathbb{e}_j^\mathsf{T} A \mathbb{1}_n\big) A_{ij} < 1.
\end{aligned}
$$

This concludes the proof of statement *(2)*.

Regarding statement *(3)*, note that, if A is row-substochastic, then A^k is row-substochastic for any natural $k \geq 1$. Let S_k be the set of indices i such that the ith row-sum of A^k is strictly less than 1. Statement *(1)* implies $S_k \subseteq S_{k+1}$. Moreover, because of the existence of directed paths from every node to nodes with row-sum less than 1, we know that there exists k^* such that $S_k^* = \{1, \dots, n\}$. This proves statement *(3)*.

Next, define the maximum row-sum at time k^* by

$$\gamma = \max_{i \in \{1, \dots, n\}} \sum_{j=1}^{n} (A^{k^*})_{ij} < 1.$$

Given any natural number k, we can write $k = ak^* + b$ with a a positive integer and $b \in \{0, \dots, k^* - 1\}$. Note that

$$A^k \mathbb{1}_n \leq A^{ak^*} \mathbb{1}_n \leq \gamma^a \mathbb{1}_n.$$

The last inequality implies that, as $k \to \infty$ and therefore $a \to \infty$, the sequence A^k converges to 0. This fact proves statement *(4)* and, in turn, that the condition *(iii)a* implies the bound *(iii)b*.

Finally, we sketch the proof that the bound *(iii)b* implies the condition *(iii)a*. By contradiction, if condition *(iii)a* does not hold, then the condensation of G contains a sink whose corresponding row-sums in A are all equal to $\max(A \mathbb{1}_n)$. But to that sink corresponds an eigenvector of A whose eigenvalue is therefore $\max(A \mathbb{1}_n)$. We refer to Theorem *5.2* for a brief review of the properties of reducible non-negative matrix and leave to the reader the details of the proof. ∎

4.6 Historical notes and further reading

Standard books on algebraic graph theory are (Biggs, 1994; Godsil and Royle, 2001).

The proof for Theorem *4.7* is taken from (Bullo et al., 2009). For more information on the Frobenius number we refer to (Owens, 2003) and Wikipedia:Coin_Problem.

More results on spectral graph theory and, specifically, a review and recent results on bounding the spectral radius of an adjacency matrix are given, for example, by Nikiforov (2002) and Das and Kumar (2004).

4.7 Exercises

E4.1 **Directed paths and powers of the adjacency matrix.** Prove Lemma 4.2.

E4.2 **Edges and triangles in an undirected graph.** Let A be the binary adjacency matrix for an undirected graph G without self-loops. Recall that the trace of A is $\operatorname{trace}(A) = \sum_{i=1}^{n} a_{ii}$.

   (i) Show $\operatorname{trace}(A) = 0$.

   (ii) Show $\operatorname{trace}(A^2) = 2|E|$, where $|E|$ is the number of edges of G.

   (iii) Show $\operatorname{trace}(A^3) = 6|T|$, where $|T|$ is the number of triangles of G. (A triangle is a complete subgraph with three vertices.)

   (iv) Verify results (i)–(iii) on the matrix $A = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$.

E4.3 **A sufficient condition for primitivity.** Assume the square matrix A is non-negative and irreducible. Show that

   (i) if A has a positive diagonal element, then A is primitive,

   (ii) if A is primitive, then it is false that A must have a positive diagonal element.

E4.4 **Example row-stochastic matrices and associated digraph.** Consider the row-stochastic matrices

$$A_1 = \frac{1}{2} \begin{bmatrix} 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 \end{bmatrix}, \quad A_2 = \frac{1}{2} \begin{bmatrix} 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix}, \quad \text{and} \quad A_3 = \frac{1}{2} \begin{bmatrix} 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix}.$$

   (i) Draw the digraphs G_1, G_2 and G_3 associated with these three matrices.

Using only the original definitions and without relying on the characterizations in Theorems 4.3 and 4.7, show that:

   (ii) the matrices A_1, A_2 and A_3 are irreducible and primitive,

   (iii) the digraphs G_1, G_2 and G_3 are strongly connected and aperiodic, and

   (iv) the averaging algorithm defined by A_2 converges in a finite number of steps.

E4.5 **Primitive matrices are irreducible.** Prove Lemma 2.11, that is, show that a primitive matrix is irreducible. *Hint: You are allowed to use Theorem 4.3.*

E4.6 **Yet another equivalent definition of irreducibility.** Consider a non-negative matrix A of dimension n. From Theorem 4.3, we know that A is irreducible if and only if

   (i) there does not exist a permutation $P \in \{0,1\}^{n \times n}$ and $1 \leq r \leq n-1$ such that

$$PAP^{\mathsf{T}} = \left[\begin{array}{c|c} B_{r \times r} & C_{r \times (n-r)} \\ \hline \mathbb{0}_{(n-r) \times r} & D_{(n-r) \times (n-r)} \end{array} \right].$$

Consider now the following property of A:

   (ii) for any non-negative vector $y \in \mathbb{R}^n_{\geq 0}$ with $0 < k < n$ strictly positive components, the vector $(I_n + A)y$ has at least $k+1$ strictly positive components.

Prove that statement (i) implies statement (ii).

E4.7   **An example reducible or irreducible matrix.** Consider the binary matrix:

$$A = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

Prove that A is irreducible or prove that A is reducible by providing a permutation matrix P that transforms A into an upper block-triangular matrix, i.e., $P^{\mathsf{T}}AP = \begin{bmatrix} \star & \star \\ \mathbb{0} & \star \end{bmatrix}$.

E4.8   **The exponent of a primitive matrix.**

   (i) Let G be the digraph with nodes $\{1, \ldots, 3\}$ and edges $\{(1, 2), (2, 1), (2, 3), (3, 1)\}$. Explain if and why G is strongly connected and aperiodic.

   (ii) Recall a non-negative matrix A is primitive if there exists a number k such that $A^k > 0$; the smallest such number k is called the *exponent* of the primitive matrix A. Do one of the following:

       a) prove that the exponent of a primitive matrix $A \in \mathbb{R}^{n \times n}$ is less than or equal to n, or
       b) provide a counterexample.

E4.9   **Bounds on the spectral radius of irreducible non-negative matrices.** For a non-negative matrix A, complete the proof of Theorem 4.9(i), that is, show that

   (i) $\min(A\mathbb{1}_n) \leq \rho(A)$ and, therefore,

$$\min(A\mathbb{1}_n) \leq \rho(A) \leq \max(A\mathbb{1}_n).$$

Next, show that

   (ii) if A is irreducible and $\min(A\mathbb{1}_n) < \max(A\mathbb{1}_n)$, then

$$\min(A\mathbb{1}_n) < \rho(A) < \max(A\mathbb{1}_n).$$

E4.10 **Eigenvalue shifting for stochastic matrices.** Let $A \in \mathbb{R}^{n \times n}$ be an irreducible row-stochastic matrix. Let E be a diagonal matrix with diagonal elements $E_{ii} \in \{0, 1\}$, with at least one diagonal element equal to zero. Show that AE and EA are convergent.

E4.11 **Normalization of non-negative irreducible matrices.** Consider a strongly connected weighted digraph G with n nodes and with an irreducible adjacency matrix $A \in \mathbb{R}^{n \times n}$. The matrix A is not necessarily row-stochastic. Find a positive vector $v \in \mathbb{R}^n$ so that the normalized matrix

$$P = \frac{1}{\rho(A)}(\mathrm{diag}(v))^{-1}A\,\mathrm{diag}(v)$$

is non-negative, irreducible, and row-stochastic.

E4.12 **Characterization of indecomposable matrices.** Following (Wolfowitz, 1963), we say a non-negative matrix A is *indecomposable* if its associated digraph contains a globally reachable node. Generalizing the proof of Theorem 4.7, show that the following statements are equivalent:

   (i) A is indecomposable and the subgraph of globally reachable nodes is aperiodic, and
   (ii) there exists an index $h \in \mathbb{N}$ such that A^h has a positive column.

E4.13 **The Frobenius number.** Prove Lemma 4.8.
       Hint: Read up on the Frobenius number in (Owens, 2003).

E4.14 **Leslie population model.** The Leslie model is used in population ecology to model the changes in a population of organisms over a period of time; see the original reference (Leslie, 1945) and a comprehensive text (Caswell, 2006). In this model, the population is divided into n groups based on age classes; the indices i are ordered increasingly with the age, so that $i = 1$ is the class of the newborns. The variable $x_i(k)$, $i \in \{1, \ldots, n\}$, denotes the number of individuals in the age class i at time k; at every time step k the $x_i(k)$ individuals

- produce a number $\alpha_i x_i(k)$ of offsprings (i.e., individuals belonging to the first age class), where $\alpha_i \geq 0$ is a fecundity rate, and

- progress to the next age class with a survival rate $\beta_i \in [0, 1]$.

If $x(k)$ denotes the vector of individuals at time k, the Leslie population model reads

$$x(k + 1) = Ax(k) = \begin{bmatrix} \alpha_1 & \alpha_2 & \ldots & \alpha_{n-1} & \alpha_n \\ \beta_1 & 0 & \ldots & 0 & 0 \\ 0 & \beta_2 & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \ldots & \beta_{n-1} & 0 \end{bmatrix} x(k), \tag{E4.1}$$

where A is referred to as the *Leslie matrix*. Consider the following two independent sets of questions. First, assume $\alpha_i > 0$ for all $i \in \{1, \ldots, n\}$ and $0 < \beta_i \leq 1$ for all $i \in \{1, \ldots, n - 1\}$.

(i) Prove that the matrix A is primitive.

(ii) Let $p_i(k) = \frac{x_i(k)}{\sum_{i=1}^{n} x_i(k)}$ denote the percentage of the total population in class i at time k. Call $p(k)$ the *population distribution* at time k. Compute $\lim_{k \to +\infty} p(k)$ as a function of the spectral radius $\rho(A)$ and the parameters (α_i, β_i), $i \in \{1, \ldots, n\}$.

Hint: Obtain a recursive expression for the components of the right dominant eigenvector of A

(iii) Assume $\beta_i = \beta > 0$ and $\alpha_i = \frac{\beta}{n}$ for $i \in \{1, \ldots, n\}$. What percentage of the total population belongs to the eldest class asymptotically, that is, what is $\lim_{k \to \infty} p_n(k)$?

(iv) Find a sufficient condition on the parameters (α_i, β_i), $i \in \{1, \ldots, n\}$, so that the population will eventually become extinct.

Second, assume $\alpha_i \geq 0$ for $i \in \{1, \ldots, n\}$ and $0 \leq \beta_i \leq 1$ for all $i \in \{1, \ldots, n - 1\}$.

(v) Find a necessary and sufficient condition on the parameters $\alpha_1, \ldots, \alpha_n$, and $\beta_1, \ldots, \beta_{n-1}$, so that the Leslie matrix A is irreducible.

(vi) For an irreducible Leslie matrix (as in the previous point (v)), find a sufficient condition on the parameters (α_i, β_i), $i \in \{1, \ldots, n\}$, that ensures that the population will not go extinct.

E4.15 **Swiss railroads: continued.** From Exercise E3.7, consider the fictitious railroad map of Switzerland given in Figure E3.1. Write the unweighted adjacency matrix A of this transportation network and, relying upon A and its powers, answer the following questions:

(i) what is the number of links of the shortest path connecting St. Gallen to Zermatt?

(ii) is it possible to go from Bern to Chur using 4 links? And 5?

(iii) how many different routes, with strictly less then 9 links and possibly visiting the same station more than once, start from Zürich and end in Lausanne?

E4.16 **Tridiagonal Toeplitz matrices.** An $n \times n$ matrix A is *tridiagonal Toeplitz* if there exist scalar numbers a, b, and c, with $a \neq 0$ and $c \neq 0$ such that

$$A = \begin{bmatrix} b & a & 0 & \ldots & 0 \\ c & b & a & \ldots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \ldots & c & b & a \\ 0 & \ldots & 0 & c & b \end{bmatrix}.$$

Show that the eigenvalues and right eigenvectors of a tridiagonal Toeplitz A are, for $j \in \{1, \ldots, n\}$,

$$\lambda_j = b + 2a\sqrt{c/a}\cos\Big(\frac{j\pi}{n+1}\Big), \quad \text{and} \quad v_j = \begin{bmatrix} (c/a)^{1/2}\sin(1j\pi/(n+1)) \\ (c/a)^{2/2}\sin(2j\pi/(n+1)) \\ \vdots \\ (c/a)^{n/2}\sin(nj\pi/(n+1)) \end{bmatrix}.$$

E4.17 **Circulant matrices.** A matrix $C \in \mathbb{C}^{n \times n}$ is *circulant* if there exists scalar numbers $c_0, \ldots, c_{n-1}$ such that

$$C = \begin{bmatrix} c_0 & c_1 & \ldots & c_{n-1} \\ c_{n-1} & c_0 & \ldots & c_{n-2} \\ \vdots & \ddots & \ddots & \vdots \\ c_1 & c_2 & \ldots & c_0 \end{bmatrix}.$$

In other words, a circulant matrix is fully specified by its first row; the remaining row of C are cyclic permutations of the first row. A circulant matrix is Toeplitz. Show that

(i) the eigenvalues and eigenvectors C are, for $j \in \{0, \ldots, n-1\}$,

$$\lambda_j = c_0 + c_1\omega_j + c_2\omega_j^2 + \cdots + c_{n-1}\omega_j^{n-1}, \quad \text{and} \quad v_j = \begin{bmatrix} 1 \\ \omega_j \\ \vdots \\ \omega_j^{n-1} \end{bmatrix},$$

where $\omega_j = \exp\Big(\frac{2j\pi i}{n}\Big)$, $j \in \{0, \ldots, n-1\}$, are the nth complex roots of the number 1, and $i = \sqrt{-1}$.

(ii) for n even, $\kappa \in \mathbb{R}$, and $(c_0, c_1, \ldots, c_{n-1}) = (1 - 2\kappa, \kappa, 0, \ldots, 0, \kappa)$, the eigenvalues are, for $j \in \{1, \ldots, n\}$,

$$\lambda_j = 2\kappa \cos\frac{2\pi(j-1)}{n} + (1 - 2\kappa).$$

Note: *Circulant matrices enjoy numerous properties; e.g., if C_1 and C_2 are circulant, so are C_1^{T}, $C_1 + C_2$ and C_1C_2. Additional properties are discussed for example by* Davis (*1979*).

E4.18 **Adjacency spectrum of basic graphs.** Given the basic graphs in Example 4.1 and the properties of tridiagonal Toeplitz and circulant matrices in Exercises E4.16 and E4.17, prove the statements in Table 4.1. In other words, show that, for $n \geq 2$,

(i) for the *path graph* P_n, the adjacency matrix is Toeplitz tridiagonal and the adjacency spectrum is $\{2\cos(\pi i/(n+1)) \mid i \in \{1, \ldots, n\}\}$;

(ii) for the *cycle graph* C_n, the adjacency matrix is circulant and the adjacency spectrum is $\{2\cos(2\pi i/n)) \mid i \in \{1, \ldots, n\}\}$;

(iii) for the *star graph* S_n, the adjacency matrix is $e_1 e_{-1} + e_{-1} e_1$, where $e_{-i} = \mathbb{1}_n - e_i$, and the adjacency spectrum is $\{\sqrt{n-1}, 0, \ldots, 0, -\sqrt{n-1}\}$;

(iv) for the *complete graph* K_n, the adjacency matrix is $\mathbb{1}_n \mathbb{1}_n^\mathsf{T} - I_n$, and the adjacency spectrum is $\{(n-1), -1, \ldots, -1\}$; and

(v) for the *complete bipartite graph* $K_{n,m}$, the adjacency matrix is $\begin{bmatrix} \mathbb{0}_{n \times n} & \mathbb{1}_{n \times m} \\ \mathbb{1}_{m \times n} & \mathbb{0}_{m \times m} \end{bmatrix}$ and the adjacency spectrum $\{\sqrt{nm}, 0, \ldots, 0, -\sqrt{nm}\}$.

Discrete-time Averaging Systems

After discussing matrix and graph theory, we are ready to go back to the averaging model introduced in Chapter 1. Recall that the discrete-time averaging systems, as given in equation (1.2), is

$$x(k+1) = Ax(k), \tag{5.1}$$

where the matrix $A = [a_{ij}]$ is row-stochastic. Also recall from Chapter 1 the study of (i) opinion dynamics in social influence networks (given an arbitrary stochastic matrix, what do its powers converge to?) and (ii) averaging algorithms in wireless sensor networks (design an algorithm to compute the average of a collection numbers located at distinct nodes). Other related examples from the appendices of Chapter 1 include the study of robotic networks in cyclic pursuit and balancing and of more general design problems in wireless sensor networks.

Figure 5.1: Opinion averaging is believed to be a key mechanism in social influence network.

This chapter presents some convergence results for the averaging model 5.1 defined by stochastic matrices; we discuss primitive matrices and reducible matrices with a single or multiple sinks. We then the equal-neighbor and the Metropolis–Hastings models of row-stochastic matrices. Finally, we present some centrality notions from network science.

5.1 Averaging systems achieving consensus

We now bring together Perron-Frobenius theory with algebraic graph theory to provide necessary and sufficient conditions for an averaging system to achieve consensus.

Recall that a sufficient condition for convergence of the averaging model (5.1) is given in Corollary 2.15: if A is primitive, then each solution converges to consensus asymptotically. The following result is more general and also amounts to an extension to a class of reducible matrices of the Perron-Frobenius Theorem 2.12.

Theorem 5.1 (Consensus for row-stochastic matrices with a globally-reachable aperiodic strong-ly-connected component). *Let A be a row-stochastic matrix and let G be its associated digraph. The following statements are equivalent:*

(A1) *the eigenvalue 1 is simple and all other eigenvalues μ satisfy $|\mu| < 1$,*

(A2) *A is semi-convergent and $\lim_{k\to\infty} A^k = \mathbb{1}_n w^{\mathsf{T}}$, for some $w \in \mathbb{R}^n$, $w \geq 0$, and $\mathbb{1}_n^{\mathsf{T}} w = 1$,*

(A3) *G contains a globally reachable node and the subgraph of globally reachable nodes is aperiodic.*

If any, and therefore all, of the previous conditions are satisfied, then the matrix A is said to be indecomposable *and the following properties hold:*

(i) *$w \geq 0$ is the left dominant eigenvector of A and $w_i > 0$ if and only if node i is globally reachable;*

(ii) *the solution to the averaging model (5.1) $x(k+1) = Ax(k)$ satisfies*

$$\lim_{k\to\infty} x(k) = \left(w^{\mathsf{T}} x(0) \right) \mathbb{1}_n;$$

(iii) *if additionally A is doubly-stochastic, then $w = \frac{1}{n}\mathbb{1}_n$ (because $A^{\mathsf{T}}\mathbb{1}_n = \mathbb{1}_n$ and $\frac{1}{n}\mathbb{1}_n^{\mathsf{T}}\mathbb{1}_n = 1$) so that*

$$\lim_{k\to\infty} x(k) = \frac{\mathbb{1}_n^{\mathsf{T}} x(0)}{n}\mathbb{1}_n = \operatorname{average}\!\big(x(0)\big)\mathbb{1}_n.$$

An example indecomposable row-stochastic matrix with its associated digraph and spectrum is illustrated in Figure 5.2.

(a) A row-stochastic matrix; each row contains equal entries summing to 1.

(b) The corresponding digraph has an aperiodic subgraph of globally reachable nodes.

(c) The spectrum of the adjacency matrix includes a dominant eigenvalue.
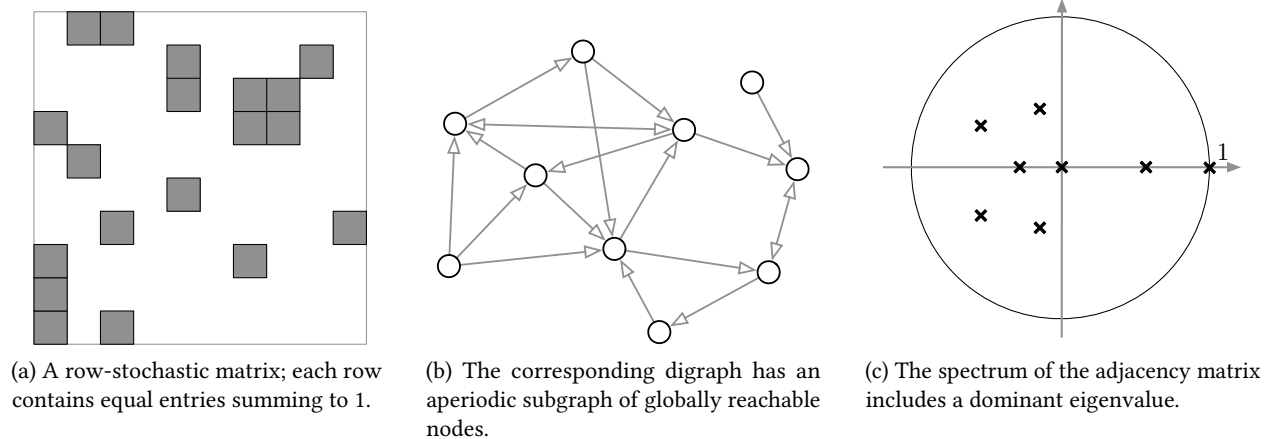
Figure 5.2: An example indecomposable row-stochastic matrix, its associated digraph consistent with Theorem 5.1(A2), and its spectrum consistent with Theorem 5.1(A1)

Note: The implication (A3) $\implies$ (ii) amounts to a result in which the structure of the network determines its function, i.e., the asymptotic behavior of the averaging system.

Note: as discussed in Section 2.3, statement (ii) implies that the limiting value is a weighted average of the initial conditions with relative weights given by the convex combination coefficients $w_1, \ldots, w_n$. Note that $w > 0$ if and only if the digraph associated to A is strongly connected. In digraphs that are not strongly connected, the initial values $x_i(0)$ of all nodes i which are not globally reachable have no effect on the final convergence value. In a social influence network, the coefficient w_i is regarded as the "social influence" of agent i. We illustrate this concept for the famous Krackhardt's advice network (Krackhardt, 1987); see Figure 5.3.
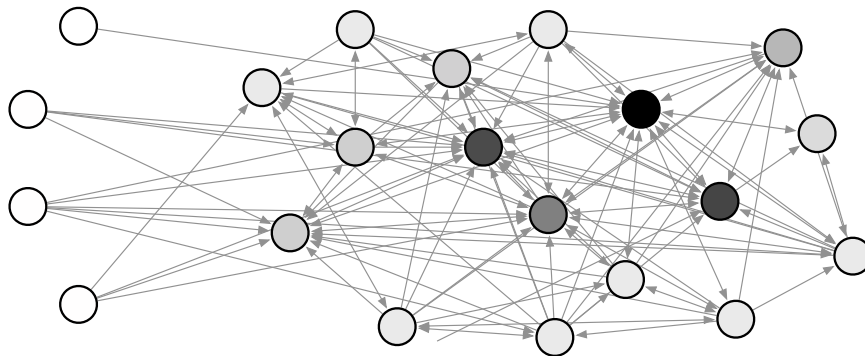
Figure 5.3: Krackhardt's advice network (Krackhardt, 1987) describing the interactions among 21 individuals. The social influence of each node is illustrated by its gray level. the adjacency matrix of this digraph is indecomposable, i.e., the digraph contains a subgraph of globally reachable nodes that is aperiodic.

Note: to clarify statement (A3) it is useful to review some properties of globally reachable nodes. We first recall a useful property from Lemma 3.1: G has a globally reachable node if and only if its condensation digraph has a globally reachable node (i.e., the condensation of G has a single sink). Second, it is easy to see that the set of globally reachable nodes induces a strongly connected component of G.

Proof of Theorem 5.1. The statement (A1) $\implies$ (A2) is precisely Theorem 2.14 with $\lambda = 1$ (whose proof is given in Section 2.3.4).

Next, we prove that (A2) $\implies$ (A3). The assumption $\mathbb{1}_n^\mathsf{T} w = 1$ implies that at least one element, say the jth element, of w is positive. Because $\lim_{k\to\infty} A^k = \mathbb{1}_n w^\mathsf{T}$, we know that the jth column of $\lim_{k\to\infty} A^k$ has all-positive elements. Thus, for sufficiently large K, the jth column of A^K has all-positive elements, so there is a path of length K from every node to the jth node. Thus, the jth node is globally reachable.

Now, we prove by contradiction that the strongly-connected component of globally reachable nodes is aperiodic: suppose this component is periodic with period $p > 1$. Pick j and K as in the previous paragraph so that there is a path of length K from the jth node to itself (a cycle of length K). Similarly, there must also be a path of length $K + 1$ from the jth node to itself (a cycle of length $K + 1$). Both cycles belong to the sub-graph of globally reachable nodes, since all the nodes on the cycles are globally reachable via the jth node. From the definition of period, both K and $K + 1$ must be divisible by p. This is not possible if $p > 1$, and is a contradiction. Hence, (A2) $\implies$ (A3).

Finally, we prove the implications (A3) $\implies$ (A1) and (A2). By assumption the condensation digraph of A contains a sink that is globally reachable, hence it is unique. Assuming $0 < n_1 < n$ nodes are globally reachable, a permutation of rows and columns (see Exercise E3.1), brings the matrix A into the

lower-triangular form

$$A = \begin{bmatrix} A_{11} & \mathbb{0}_{n_1 \times n_2} \\ A_{21} & A_{22} \end{bmatrix}, \tag{5.2}$$

where $A_{11} \in \mathbb{R}^{n_1 \times n_1}$, $A_{22} \in \mathbb{R}^{n_2 \times n_2}$, with $n_1 + n_2 = n$. The state vector x is correspondingly partitioned into $x_1 \in \mathbb{R}^{n_1}$ and $x_2 \in \mathbb{R}^{n_2}$ so that

$$x_1(k+1) = A_{11} x_1(k), \tag{5.3}$$
$$x_2(k+1) = A_{21} x_1(k) + A_{22} x_2(k). \tag{5.4}$$

In other words, x_1 and A_{11} are the variables and the matrix corresponding to the sink. Because the sink, as a subgraph of G, is strongly connected and aperiodic, A_{11} is primitive and row-stochastic and, by Corollary 2.15,

$$\lim_{k \to \infty} A_{11}^k = \mathbb{1}_{n_1} w_1^\mathsf{T},$$

where $w_1 > 0$ is the left eigenvector with eigenvalue 1 for A_{11} normalized so that $\mathbb{1}_{n_1}^\mathsf{T} w_1 = 1$.

   We next analyze the matrix A_{22} as follows. Recall from Corollary 4.11 that an irreducible row-substochastic matrix has spectral radius less than 1. Now, because A_{21} cannot be zero (otherwise the sink would not be globally reachable), the matrix A_{22} is row-substochastic. Moreover, (after appropriately permuting rows and columns of A_{22}) it can be observed that A_{22} is a lower-triangular matrix such that each diagonal block is row-substochastic and irreducible (corresponding to each node in the condensation digraph). Therefore, we know $\rho(A_{22}) < 1$ and, in turn, $I_{n_2} - A_{22}$ is invertible. Because A_{11} is primitive and $\rho(A_{22}) < 1$, A is semi-convergent and $\lim_{k \to \infty} x_2(k)$ exists. This establishes that (A3) $\implies$ (A1). Taking the limit as $k \to \infty$ in equation (5.4), some straightforward algebra shows that

$$\lim_{k \to \infty} x_2(k) = (I_{n_2} - A_{22})^{-1} A_{21} \left(\lim_{k \to \infty} x_1(k) \right) = (I_{n_2} - A_{22})^{-1} A_{21} \left(\mathbb{1}_{n_1} w_1^\mathsf{T} \right) x_1(0).$$

From the row-stochasticity of A, we know $A_{21} \mathbb{1}_{n_1} + A_{22} \mathbb{1}_{n_2} = \mathbb{1}_{n_2}$ and hence $(I_{n_2} - A_{22})^{-1} A_{21} \mathbb{1}_{n_1} = \mathbb{1}_{n_2}$. Collecting these results, we write

$$\lim_{k \to \infty} \begin{bmatrix} A_{11} & \mathbb{0}_{n_1 \times n_2} \\ A_{21} & A_{22} \end{bmatrix}^k = \begin{bmatrix} \mathbb{1}_{n_1} w_1^\mathsf{T} & \mathbb{0}_{n_1 \times n_2} \\ \mathbb{1}_{n_2} w_1^\mathsf{T} & \mathbb{0}_{n_2 \times n_2} \end{bmatrix} = \mathbb{1}_n \begin{bmatrix} w_1 \\ \mathbb{0}_{n_2} \end{bmatrix}^\mathsf{T}.$$

This establishes that (A3) $\implies$ (A2) and (A1) $\implies$ (i). The implications (A2) $\implies$ (ii) and (A2) $\implies$ (iii) are straightforward.   ∎

5.2   Averaging with reducible matrices with multiple sinks

In this section we now consider the general case of digraphs that do not contain globally reachable nodes, that is, digraphs whose condensation digraph has multiple sinks. Such an example digraph is the famous Sampson Monastery network (Sampson, 1969); see Figure 5.4.

   In the following statement we say that a node is *connected* with a sink of a digraph if there exists a directed path from the node to any node in the sink.
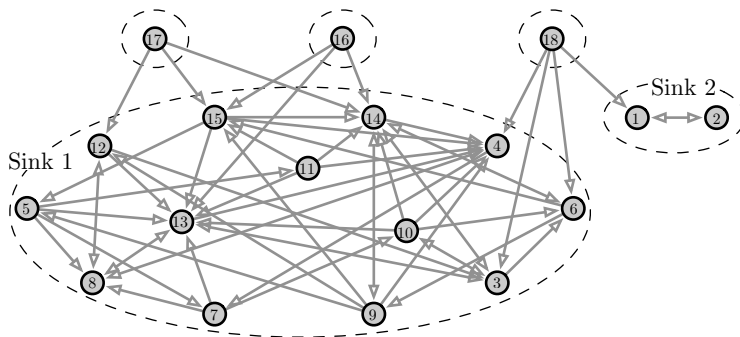
Figure 5.4: This image illustrates the Sampson Monastery dataset (Sampson, 1969). This dataset describes the of social relations among a set of 18 monk-novitiates in an isolated contemporary American monastery. This digraph contains two sinks in its condensation.

Theorem 5.2 (Convergence for row-stochastic matrices with multiple aperiodic sinks). *Let A be a row-stochastic matrix, let G be its associated digraph, and let $M \geq 2$ be the number of sinks in the condensation digraph $C(G)$. If each of the M sinks is aperiodic, then*

(i) *the semi-simple eigenvalue $\rho(A) = 1$ has multiplicity equal M and is strictly larger than the magnitude of all other eigenvalues, hence A is semi-convergent,*

(ii) *there exist M left eigenvectors of A, denoted by $w^m \in \mathbb{R}^n$, for $m \in \{1, \dots, M\}$, with the properties that: $w^m \geq 0$, $w_1^m + \cdots + w_n^m = 1$ and w_i^m is positive if and only if node i belongs to the m-th sink,*

(iii) *the solution to the averaging model $x(k+1) = Ax(k)$ with initial condition $x(0)$ satisfies*

$$
\lim_{k \to \infty} x_i(k) = \begin{cases} (w^m)^\mathsf{T} x(0), & \text{if node } i \text{ belongs to the } m\text{-th sink,} \\ (w^m)^\mathsf{T} x(0), & \text{if node } i \text{ is connected with the } m\text{-th sink and no other sink,} \\ \displaystyle\sum_{m=1}^{M} z_{i,m}\big((w^m)^\mathsf{T} x(0)\big), & \text{if node } i \text{ is connected to more than one sink,} \end{cases}
$$

where, for each node i connected to more than one sink, the coefficients $z_{i,m}, m \in \{1, \dots, S\}$, are convex combination coefficients and are strictly positive if and only if there exists a directed path from node i to the sink m.

Proof. Rather than treating the general case with heavy notation, we work out a significant example with the key ideas of the general proof, and refer the reader to (DeMarzo et al., 2003, Theorem 10) for the details. Assume the condensation digraph of A is composed of three nodes, two of which are sinks, as in the side figure.
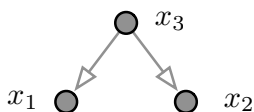
Therefore, after a permutation of rows and columns (see Exercise E3.1), A can be written as

$$A = \begin{bmatrix} A_{11} & 0 & 0 \\ 0 & A_{22} & 0 \\ A_{31} & A_{32} & A_{33} \end{bmatrix}$$

and the state vector x is correspondingly partitioned into the vectors x_1, x_2 and x_3. The state equations are:

$$x_1(k+1) = A_{11}x_1(k), \tag{5.5}$$
$$x_2(k+1) = A_{22}x_2(k), \tag{5.6}$$
$$x_3(k+1) = A_{31}x_1(k) + A_{32}x_2(k) + A_{33}x_3(k). \tag{5.7}$$

By the properties of the condensation digraph and the assumption of aperiodicity of the sinks, the digraphs associated to the row-stochastic matrices A_{11} and A_{22} are strongly connected and aperiodic. Therefore, we immediately conclude that

$$\lim_{k\to\infty} x_1(k) = \big(w_1^{\mathsf{T}} x_1(0)\big) \mathbb{1}_{n_1} \quad \text{and} \quad \lim_{k\to\infty} x_2(k) = \big(w_2^{\mathsf{T}} x_2(0)\big) \mathbb{1}_{n_2},$$

where w_1 (resp. w_2) is the left eigenvector of the eigenvalue 1 for matrix A_{11} (resp. A_{22}) with the usual normalization $\mathbb{1}_{n_1}^{\mathsf{T}} w_1 = \mathbb{1}_{n_2}^{\mathsf{T}} w_2 = 1$.

Regarding the matrix A_{33}, the same discussion as in the previous proof leads to $\rho(A_{33}) < 1$ and, in turn, to the statement that $I_{n_3} - A_{33}$ is nonsingular. By taking the limit as $k \to \infty$ in equation (5.7), some straightforward algebra shows that

$$\lim_{k\to\infty} x_3(k) = (I_{n_3} - A_{33})^{-1}\big(A_{31} \lim_{k\to\infty} x_1(k) + A_{32} \lim_{k\to\infty} x_2(k)\big)$$
$$= (w_1^{\mathsf{T}} x_1(0))\big((I_{n_3} - A_{33})^{-1} A_{31}\mathbb{1}_{n_1}\big) + (w_2^{\mathsf{T}} x_2(0))\big((I_{n_3} - A_{33})^{-1} A_{32}\mathbb{1}_{n_2}\big).$$

Moreover, because A is row-stochastic, we know

$$A_{31}\mathbb{1}_{n_1} + A_{32}\mathbb{1}_{n_2} + A_{33}\mathbb{1}_{n_3} = \mathbb{1}_{n_3},$$

and, using again the fact that $I_{n_3} - A_{33}$ is nonsingular,

$$\mathbb{1}_{n_3} = (I_{n_3} - A_{33})^{-1} A_{31}\mathbb{1}_{n_1} + (I_{n_3} - A_{33})^{-1} A_{32}\mathbb{1}_{n_2}.$$

This concludes our proof of Theorem 5.2 for the simplified case $C(G)$ having three nodes and two sinks. ∎

Note that: convergence does not occur to consensus (not all components of the state are equal) and the final value of all nodes is independent of the initial values at nodes which are not in the sinks of the condensation digraph.

We conclude this section with a figure providing a summary of the asymptotic behavior of discrete-time averaging systems and its relationships with properties of matrices and graphs; see Figure 5.5.

5.3   Appendix: Design of graphs weights

In this section we describe two widely-adopted algorithms to design weights for unweighted graphs.

Properties of $x(k+1) = Ax(k)$ Properties of row-stochastic matrix A Properties of associated digraph
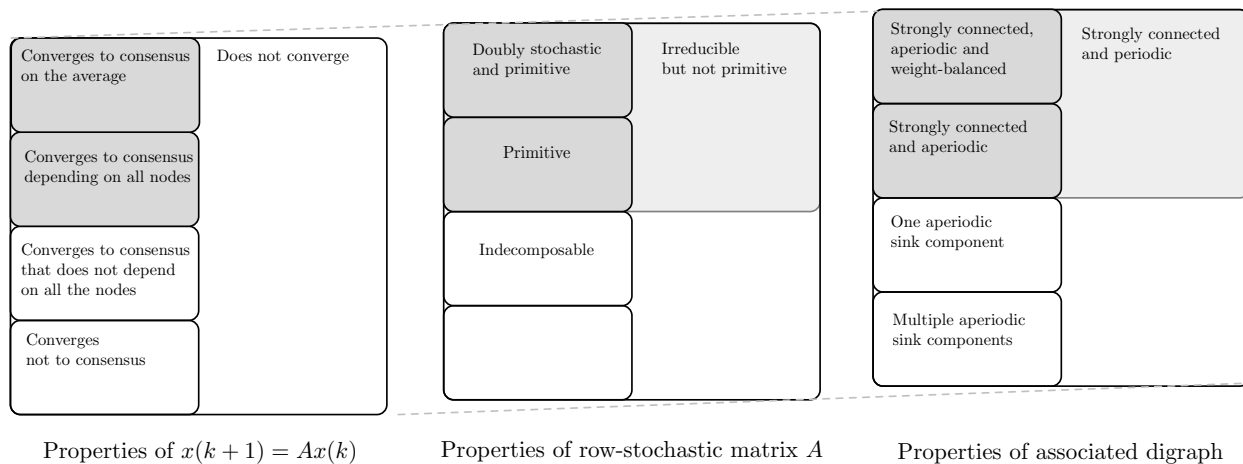
Figure 5.5: Corresponding properties for the discrete-time averaging dynamical system $x(k+1) = Ax(k)$, the row-stochastic matrix A and the associated weighted digraph.

5.3.1 The equal-neighbor model

Let G be a connected undirected graph, binary adjacency matrix A, and degree matrix $D = \text{diag}(d_1, \ldots, d_n)$, where $d_1, \ldots, d_n$ are the node degrees. Define the *equal-neighbor matrix*

$$A_{\text{equal-neighbor}} = D^{-1}A. \tag{5.8}$$

For example, consider the graph in Figure 5.6, for which we have:

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}, \quad D = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix} \implies A_{\text{equal-neighbor}} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1/3 & 0 & 1/3 & 1/3 \\ 0 & 1/2 & 0 & 1/2 \\ 0 & 1/2 & 1/2 & 0 \end{bmatrix}. \tag{5.9}$$
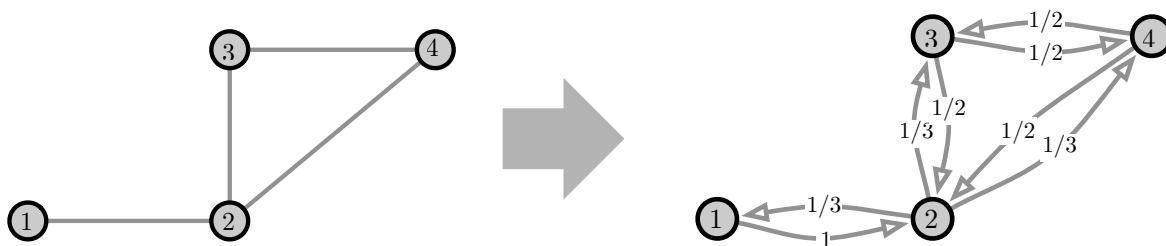
Figure 5.6: The equal-neighbor matrix

The following result is for the more general setting of weighted undirected graphs.

Lemma 5.3 (The equal-neighbor row-stochastic matrix). *Let G be a connected weighted graph with weighted adjacency matrix A and weighted degrees $d_1, \ldots, d_n$. For the equal-neighbor matrix $A_{\text{equal-neighbor}}$ defined as in (5.8),*

(i) $A_{\text{equal-neighbor}}$ *is well-defined, row-stochastic, and irreducible;*

(ii) *the left dominant eigenvector of* $A_{\text{equal-neighbor}}$, *normalized to have unit sum, is*

$$w_{\text{equal-neighbor}} = \frac{1}{\sum_{i=1}^n d_i} \begin{bmatrix} d_1 \\ \vdots \\ d_n \end{bmatrix},$$

so that, assuming G is aperiodic, the solution to the averaging model (5.1) $x(k+1) = Ax(k)$ *satisfies*

$$\lim_{k \to \infty} x_i(k) = \frac{1}{\sum_{i=1}^n d_i} \sum_{i=1}^n d_i x_i(0); \tag{5.10}$$

(iii) $A_{\text{equal-neighbor}}$ *is doubly-stochastic if and only if G is regular (i.e., all nodes have the same degree).*

For example, for the equal-neighbor matrix in equation (5.9) and Figure 5.6, one can easily verify that the dominant eigenvector is $\begin{bmatrix} 1 & 3 & 2 & 2 \end{bmatrix}^{\mathsf{T}} /8$.

Proof of Lemma 5.3. Because G is connected, each node degree is strictly positive, the degree matrix is invertible, and $A_{\text{equal-neighbor}}$ is well-defined. Because G is connected and because the zero/positive pattern of $A_{\text{equal-neighbor}}$ is the same as that of A, we know $A_{\text{equal-neighbor}}$ is irreducible. Next, we note a simple fact: any $v \in \mathbb{R}^n$ with non-zero entries satisfies $\text{diag}(v)^{-1}v = \mathbb{1}_n$. Let $d = A\mathbb{1}_n$ denote the vector of node degrees so that $D = \text{diag}(d)$. Statement (i) follows from

$$A_{\text{equal-neighbor}}\mathbb{1}_n = \text{diag}(d)^{-1}(A\mathbb{1}_n) = \text{diag}(d)^{-1}d = \mathbb{1}_n.$$

Statement (ii) follows from

$$A_{\text{equal-neighbor}}^{\mathsf{T}} w_{\text{equal-neighbor}} = A\,\text{diag}(d)^{-1}\Big(\frac{1}{\mathbb{1}_n^{\mathsf{T}} d}d\Big) = \frac{1}{\mathbb{1}_n^{\mathsf{T}} d}A\mathbb{1}_n = \frac{1}{\mathbb{1}_n^{\mathsf{T}} d}d = w_{\text{equal-neighbor}},$$

where we used the fact that A is symmetric. Statement (iii) is an immediate consequence of (ii). ■

We conclude this section by reviewing the distributed averaging algorithm introduced in Section 1.2.

Example 5.4 (Averaging in wireless sensor networks). As in equation (1.1), assume each node of a wireless sensor network contains a value x_i and repeatedly executes:

$$x_i(k+1) := \text{average}\big(x_i(k), \{x_j(k), \text{ for all neighbor nodes } j\}\big), \tag{5.11}$$

or, more explicitly, $x_i(k+1) = \frac{1}{1+d_i}(x_i(k) + \sum_{j \in \mathcal{N}(i)} x_j(k))$. Algorithm (5.11) can be written as:

$$x(k+1) = \begin{bmatrix} 1/2 & 1/2 & 0 & 0 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 0 & 1/3 & 1/3 & 1/3 \\ 0 & 1/3 & 1/3 & 1/3 \end{bmatrix} x(k) =: A_{\text{wsn}} x(k),$$

where the matrix A_{wsn} is defined as in Section 1.2 and where it is easy to verify that

$$A_{\text{wsn}} = (D + I_4)^{-1}(A + I_4).$$

Clearly, $A + I_4$ is the adjacency matrix of a graph that is equal to the graph in figure with the addition of a self-loop at each node; this new graph has degree matrix $D + I_4$. Therefore, the matrix A_{wsn} is an equal-neighbor matrix for the graph with added self-loops. We illustrate this observation in Figure 5.7. From Lemma 5.3 we know that the left dominant eigenvector of A_{wsn} is
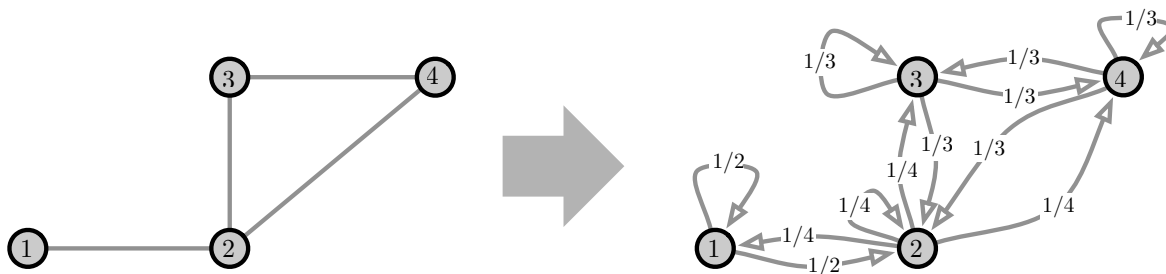
Figure 5.7: The equal-neighbor matrix for an undirected graph with added self-loops

$$w_{\text{equal-neighbor+selfloops}} = \frac{1}{n + \sum_i d_i} \begin{bmatrix} d_1 + 1 \\ \vdots \\ d_n + 1 \end{bmatrix} = \begin{bmatrix} 1/6 \\ 1/3 \\ 1/4 \\ 1/4 \end{bmatrix},$$

because $(d_1, d_2, d_3, d_4) = (1, 3, 2, 2)$ and $n = 4$. This result is consistent with the numerically-computed eigenvector in Example 2.5. $\bullet$

5.3.2 The Metropolis–Hastings model

Next, we suggest a second way of assigning weights to a graph for the purpose of designing an averaging algorithm (that achieves average consensus). Given an undirected unweighted graph G with with edge set E and degrees $d_1, \ldots, d_n$, define the weighted adjacency matrix $A_{\text{Metropolis-Hastings}}$, called the *Metropolis–Hastings matrix*, by

$$(A_{\text{Metropolis-Hastings}})_{ij} = \begin{cases} \dfrac{1}{1 + \max\{d_i, d_j\}}, & \text{if } \{i, j\} \in E \text{ and } i \neq j, \\ 1 - \displaystyle\sum_{\{i,h\}\in E} (A_{\text{Metropolis-Hastings}})_{ih}, & \text{if } i = j, \\ 0, & \text{otherwise.} \end{cases}$$

In our example,

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}, D = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix} \implies A_{\text{Metropolis-Hastings}} = \begin{bmatrix} 3/4 & 1/4 & 0 & 0 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 0 & 1/4 & 5/12 & 1/3 \\ 0 & 1/4 & 1/3 & 5/12 \end{bmatrix}.$$
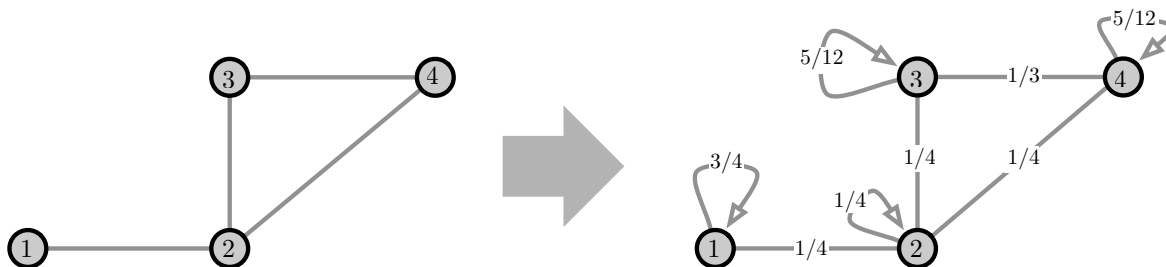
Figure 5.8: The Metropolis–Hastings model

One can verify that the Metropolis–Hastings weights have the following properties:

(i) $(A_{\text{Metropolis-Hastings}})_{ij} > 0$ if $\{i,j\} \in E$, $(A_{\text{Metropolis-Hastings}})_{ii} > 0$ for all $i \in \{1,\dots,n\}$, and $(A_{\text{Metropolis-Hastings}})_{ij} = 0$ else;

(ii) $A_{\text{Metropolis-Hastings}}$ is symmetric and doubly-stochastic;

(iii) $A_{\text{Metropolis-Hastings}}$ is primitive if and only if G is connected; and

(iv) the averaging model (5.1) $x(k+1) = Ax(k)$ achieves average consensus.

5.4 Appendix: Design and computation of centrality measures

In network science it is of interest to determine the relative importance of a node in a network. There are many ways to do so and they are referred to as centrality measures or *centrality scores*. This section presents six centrality notions based on the adjacency matrix. We treat the general case of a weighted digraph G with weighted adjacency matrix A (warning: many articles in the literature deal with undirected graphs only.) The matrix A is non-negative, but not necessarily row stochastic. From the Perron-Frobenius theory, recall the following facts:

(i) if G is strongly connected, then the spectral radius $\rho(A)$ is an eigenvalue of maximum magnitude and its corresponding left eigenvector can be selected to be strictly positive and with unit sum (see Theorem 2.12); and

(ii) if G contains a globally reachable node, then the spectral radius $\rho(A)$ is an eigenvalue of maximum magnitude and its corresponding left eigenvector is non-negative and has positive entries corresponding to each globally reachable node (see Theorem 5.1).

Degree centrality For an arbitrary weighted digraph G, the *degree centrality* $c_{\text{degree}}(i)$ of node i is its in-degree:

$$c_{\text{degree}}(i) = d_{\text{in}}(i) = \sum_{j=1}^{n} a_{ji}, \tag{5.12}$$

that is, the number of in-neighbors (if G is unweighted) or the sum of the weights of the incoming edges. Degree centrality is relevant, for example, in (typically unweighted) citation networks whereby articles are ranked on the basis of their citation records. (Warning: the notion that a high citation count is an indicator of quality is clearly a fallacy.)

Eigenvector centrality One problem with degree centrality is that each in-edge has unit count, even if the in-neighbor has negligible importance. To remedy this potential drawback, one could define the importance of a node to be proportional to the weighted sum of the importance of its in-neighbors (see (Bonacich, 1972b) for an early reference). This line of reasoning leads to the following definition.

For a weighted digraph G with globally reachable nodes (or for an undirected graph that is connected), define the *eigenvector centrality* vector, denoted by c_{ev}, to be the left dominant eigenvector of the adjacency matrix A associated with the dominant eigenvalue and normalized to satisfy $\mathbb{1}_n^\top c_{\text{ev}} = 1$.

Note that the eigenvector centrality satisfies

$$A^\top c_{\text{ev}} = \frac{1}{\alpha} c_{\text{ev}} \quad \Longleftrightarrow \quad c_{\text{ev}}(i) = \alpha \sum_{j=1}^{n} a_{ji} c_{\text{ev}}(j). \tag{5.13}$$

where $\alpha = \frac{1}{\rho(A)}$ is the only possible choice of scalar coefficient in equation (5.13) ensuring that there exists a unique solution and that the solution, denoted c_{ev}, is strictly positive in a strongly connected digraph and non-negative in a digraph with globally reachable nodes. Note that this connectivity property may be restrictive in some cases.
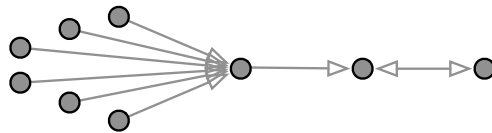
Figure 5.9: Comparing degree centrality versus eigenvector centrality: the node with maximum in-degree has zero eigenvector centrality in this graph

Katz centrality For a weighted digraph G, pick an attenuation factor $\alpha < 1/\rho(A)$ and define the *Katz centrality* vector (see (Katz, 1953)), denoted by c_{K}, by the following equivalent formulations:

$$c_{\text{K}}(i) = \alpha \sum_{j=1}^{n} a_{ji}(c_{\text{K}}(j) + 1), \tag{5.14}$$

or

$$c_{\text{K}}(i) = \sum_{k=1}^{\infty} \sum_{j=1}^{n} \alpha^k (A^k)_{ji}. \tag{5.15}$$

Katz centrality has therefore two interpretations:

(i) the importance of a node is an attenuated sum of the importance and of the number of the in-neighbors – note indeed how equation (5.14) is a combination of equations (5.12) and (5.13), and

(ii) the importance of a node is α times number of length-1 paths into i (i.e., the in-degree) plus α^2 times the number of length-2 paths into i, etc. (From Lemma 4.2, recall that, for an unweighted digraph, $(A^k)_{ji}$ is equal to the number of directed paths of length k from j to i.)

Note how, for $\alpha < 1/\rho(A)$, equation (5.14) is well-posed and equivalent to

$$
\begin{aligned}
c_{\mathrm{K}} &= \alpha A^{\mathsf{T}}(c_{\mathrm{K}} + \mathbb{1}_n) \\
\iff\ c_{\mathrm{K}} + \mathbb{1}_n &= \alpha A^{\mathsf{T}}(c_{\mathrm{K}} + \mathbb{1}_n) + \mathbb{1}_n \\
\iff\ (I_n - \alpha A^{\mathsf{T}})(c_{\mathrm{K}} + \mathbb{1}_n) &= \mathbb{1}_n \\
\iff\ c_{\mathrm{K}} &= (I_n - \alpha A^{\mathsf{T}})^{-1}\mathbb{1}_n - \mathbb{1}_n \\
\iff\ c_{\mathrm{K}} &= \sum_{k=1}^{\infty} \alpha^k (A^{\mathsf{T}})^k \mathbb{1}_n,
\end{aligned}
\tag{5.16}
$$

where we used the identity $(I_n - \mathcal{A})^{-1} = \sum_{k=0}^{\infty} \mathcal{A}^k$ valid for any matrix $\mathcal{A}$ with $\rho(\mathcal{A}) < 1$; see Exercise E2.12.

There are two simple ways to compute the Katz centrality. According to equation (5.16), for limited size problems, one can invert the matrix $(I_n - \alpha A^{\mathsf{T}})$. Alternatively, one can show that the following iteration converges to the correct value: $c_{\mathrm{K}}^+ := \alpha A^{\mathsf{T}}(c_{\mathrm{K}} + \mathbb{1}_n)$.

Figure 5.10: The pattern in figure displays the so-called hyperlink matrix, i.e., the transpose of the adjacency matrix, for a collection of websites at the Lincoln University in New Zealand from the year 2006. Light blue points are nonzero entries of the adjacency matrix; dark blue points are outgoing links toward dangling nodes. Each empty column corresponds to a webpage without any outgoing link, that is, to a so-called dangling node. This network has 3756 nodes with 31,718 links. A fairly large portion of the nodes are dangling nodes: in this example, there are 3255 dangling nodes, which is over 85% of the total. Image courtesy of Roberto Tempo from data described in (Ishii and Tempo, 2014).
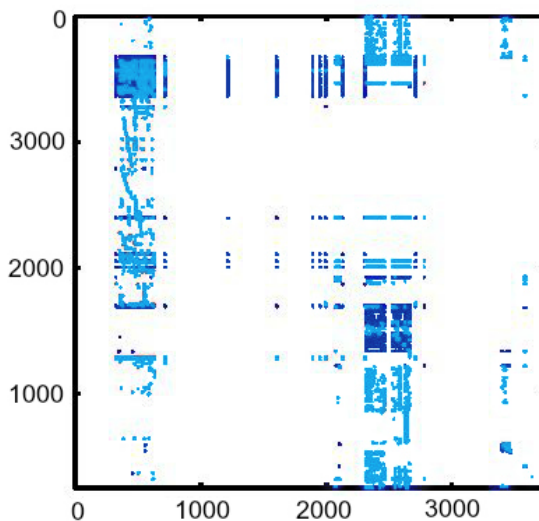
PageRank centrality   For a weighted digraph G with row-stochastic adjacency matrix (i.e., unit out-degree for each node), pick a convex combination coefficient $\alpha \in\ ]0, 1[$ and define the *PageRank* centrality

vector, denoted by c_{pr}, as the unique positive solution to

$$c_{\mathrm{pr}}(i) = \alpha \sum_{j=1}^{n} a_{ji} c_{\mathrm{pr}}(j) + \frac{1-\alpha}{n}, \tag{5.17}$$

or, equivalently, to

$$c_{\mathrm{pr}} = M c_{\mathrm{pr}}, \quad \mathbb{1}_n^\mathsf{T} c_{\mathrm{pr}} = 1, \quad \text{where } M = \alpha A^\mathsf{T} + \frac{1-\alpha}{n} \mathbb{1}_n \mathbb{1}_n^\mathsf{T}. \tag{5.18}$$

(To establish the equivalence between these two definitions, the only non-trivial step is to notice that if c_{pr} solves equation (5.17), then it must satisfy $\mathbb{1}_n^\mathsf{T} c_{\mathrm{pr}} = 1$.)

Note that, for arbitrary unweighted digraphs and binary adjacency matrices $A_{0,1}$, it is natural to compute the PageRank vector with $A = D_{\mathrm{out}}^{-1} A_{0,1}$. We refer to (Ishii and Tempo, 2014; Gleich, 2015) for the important interpretation of the PageRank score as the stationary distribution of the so-called random surfer of an hyperlinked document network — it is under this disguise that the PageRank score was conceived by the Google co-founders and a corresponding algorithm led to the establishment of the Google search engine. In the Google problem it is customary to set $\alpha \approx .85$.

Closeness and betweenness centrality (based on shortest paths) Degree, eigenvector, Katz and PageRank centrality are presented using the adjacency matrix. Next we present two centrality measures based on the notions of shortest path and geodesic distance; these two notions belong to the class of *radial* and *medial* centrality measures (Borgatti and Everett, 2006).

We start by introducing some additional graph theory. For a weighted digraph with n nodes, the *length* of a directed path is the sum of the weights of edges in the directed path. For $i, j \in \{1, \ldots, n\}$, a *shortest path* from a node i to a node j is a directed path of smallest length. Note: it is easy to construct examples with multiple shortest paths, so that the shortest path is not unique. The *geodesic distance* $d_{i \to j}$ from node i to node j is the length of a shortest path from node i to node j; we also stipulate that the geodesic distance $d_{i \to j}$ takes the value zero if $i = j$ and is infinite if there is no path from i to j. Note: in general $d_{i \to j} \neq d_{j \to i}$. Finally, for $i, j, k \in \{1, \ldots, n\}$, we let $g_{i \to k \to j}$ denote the number of shortest paths from a node i to a node j that pass through node k.

For a strongly-connected weighted digraph, the *closeness* of node $i \in \{1, \ldots, n\}$ is the inverse sum over the geodesic distances $d_{i \to j}$ from node i to all other nodes $j \in \{1, \ldots, n\}$, that is:

$$c_{\mathrm{closeness}}(i) = \frac{1}{\sum_{j=1}^{n} d_{i \to j}}. \tag{5.19}$$

For a strongly-connected weighted digraph, the *betweenness* of node $i \in \{1, \ldots, n\}$ is the fraction of all shortest paths $g_{k \to i \to j}$ from any node k to any other node j passing through node i, that is:

$$c_{\mathrm{betweenness}}(i) = \frac{\sum_{j,k=1}^{n} g_{k \to i \to j}}{\sum_{h=1}^{n} \sum_{j,k=1}^{n} g_{k \to h \to j}}. \tag{5.20}$$

Summary   To conclude this section, in Table 5.1, we summarize the various centrality definitions for a weighted directed graph.

Measure	Definition	Assumptions
degree centrality	$c_{\text{degree}} = A^{\mathsf{T}} \mathbb{1}_n$	
eigenvector centrality	$c_{\text{ev}} = \alpha A^{\mathsf{T}} c_{\text{ev}}$	$\alpha = \dfrac{1}{\rho(A)}$, G has a globally reachable node
PageRank centrality	$c_{\text{pr}} = \alpha A^{\mathsf{T}} c_{\text{pr}} + \dfrac{1-\alpha}{n} \mathbb{1}_n$	$\alpha < 1$, $A\mathbb{1}_n = \mathbb{1}_n$
Katz centrality	$c_{\text{K}} = \alpha A^{\mathsf{T}}(c_{\text{K}} + \mathbb{1}_n)$	$\alpha < \dfrac{1}{\rho(A)}$
closeness centrality	$c_{\text{closeness}}(i) = \dfrac{1}{\sum_{j=1}^{n} d_{i \to j}}$	G strongly connected
betweenness centrality	$c_{\text{betweenness}}(i) = \dfrac{\sum_{j,k=1}^{n} g_{k \to i \to j}}{\sum_{h=1}^{n} \sum_{j,k=1}^{n} g_{k \to h \to j}}$	G strongly connected

Table 5.1: Definitions of centrality measures for a weighted digraph G with adjacency matrix A

Figure 5.11 illustrates some centrality notions on a small instructive example due to Brandes (2006). Note that a different node is the most central one in each metric; this variability is naturally expected and highlights the need to select a centrality notion relevant to the specific application of interest.

(a) degree centrality   (b) eigenvector centrality

(c) closeness centrality   (d) betweenness centrality
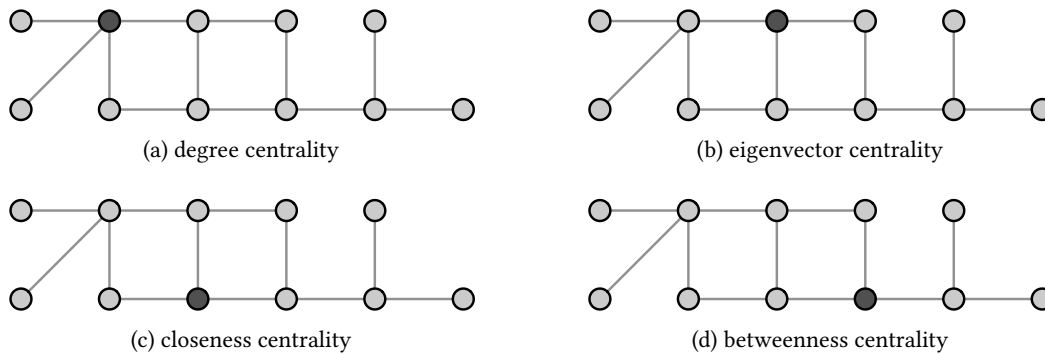
Figure 5.11: Degree, eigenvector, closeness, and betweenness centrality for an undirected unweighted graph. The dark node is the most central node in the respective metric; a different node is the most central one in each metric.

5.5   Historical notes and further reading

For references on social influence networks and opinion dynamics we refer to Chapter 1. An early reference for Theorem 5.2 is (DeMarzo et al., 2003, Appendix C and, specifically, Theorem 10). An early reference to the study of indecomposable stochastic matrices is (Wolfowitz, 1963).

On the topic of computing row-stochastic matrices, we postpone to Chapter 10 the study of related optimization problems.

A standard modern treatment of centrality notions is (Newman, 2010, Chapter 7); see also (Easley and Kleinberg, 2010, Chapter 14) for an introductory discussion. We also refer to (Brandes and Erlebach, 2005) for a comprehensive review of network analysis metrics and related computational algorithms, beyond centrality measures. Historically, centrality measures were originally studied in sociology, An incomplete list of early references and historical reviews in sociology includes (Bavelas, 1950) on closeness centrality, (Katz, 1953) on Katz centrality, (Freeman, 1977) on betweeness centrality, and (Bonacich, 1972a,b) on eigenvector centrality. Kleinberg (1999) generalizes centrality notions to networks with hubs and authorities; see Exercise E5.15.

PageRank is a centrality measure that has received tremendous recent attention due to the success of the Google search engines; this notion was popularized by (Brin and Page, 1998; Page, 2001), but see also the previous work (Friedkin, 1991) on total effective centrality and its relationship with PageRank (Friedkin and Johnsen, 2014). We refer to (Ishii and Tempo, 2014; Gleich, 2015; Nesterov, 2012) for recent works on PageRank and its multiple extentions and applications; we refer to (Ishii and Tempo, 2010; Zhao et al., 2013) for randomized distributed algorithms for PageRank computation.

5.6   Exercises

E5.1   **A sample DeGroot panel.** A conversation between 5 panelists is modeled according to the DeGroot model
by an averaging system $x^+ = A_{\text{panel}}x$, where

$$
A_{\text{panel}} = \begin{bmatrix}
0.15 & 0.15 & 0.1 & 0.2 & 0.4 \\
0 & 0.55 & 0 & 0 & 0.45 \\
0.3 & 0.05 & 0.05 & 0 & 0.6 \\
0 & 0.4 & 0.1 & 0.5 & 0 \\
0 & 0.3 & 0 & 0 & 0.7
\end{bmatrix}.
$$

Assuming that the panel has sufficiently long deliberations, answer the following:

   (i)   Draw the condensation of the associated digraph.
   (ii) Do the panelists finally agree on a common decision?
   (iii) In the event of agreement, does the initial opinion of any panelists get rejected? If so, which ones?
   (iv) Assume the panelists' initial opinions are their self-appraisals (i.e., the self-weights $a_{11}, \ldots, a_{55}$) and
         compute the final opinion via elementary calculations.

E5.2   **Three DeGroot panels.** Recall the DeGroot model introduced in Chapter 1. Denote by $x_i(0)$ the initial
opinion of each individual, and $x_i(k)$ its updated opinion after k communications with its neighbors. Then
the vector of opinions evolves over time according to $x(k + 1) = Ax(k)$ where the coefficient $a_{ij} \in [0, 1]$ is
the influence of the opinion of individual j on the update of the opinion of agent i, subject to the constraint
$\sum_j a_{ij} = 1$. Consider the following three scenarios:

   (i)   Everybody gives the same weight to the opinion of everybody else.
   (ii) There is a distinct agent (suppose the agent with index $i = 1$) that weights equally the opinion of all the
         others, and the remaining agents compute the mean between their opinion and the one of first agent.
   (iii) All the agents compute the mean between their opinion and the one of the first agent. Agent 1 does not
         change her opinion.

In each case, derive the averaging matrix A, show that the opinions converge asymptotically to a final opinion
vector, and characterize this final opinion vector.

E5.3   **The equal-neighbor row-stochastic matrix for weighted directed graphs.** Let G be a weighted digraph
with n nodes, weighted adjacency matrix A and weighted out-degree matrix D_{out}. Define the *equal-neighbor
matrix*

$$
A_{\text{equal-neighbor}} = (I_n + D_{\text{out}})^{-1}(I_n + A).
$$

Show that

   (i) $A_{\text{equal-neighbor}}$ is row-stochastic;
   (ii) $A_{\text{equal-neighbor}}$ is primitive if and only if G is strongly connected; and
   (iii) $A_{\text{equal-neighbor}}$ is doubly-stochastic if G is weight-balanced and the weighted degree is constant for all
         nodes (i.e., $D_{\text{out}} = D_{\text{in}} = dI_n$ for some $d \in \mathbb{R}_{>0}$).

E5.4   **Reversible primitive row-stochastic matrices.** Let A be a primitive row-stochastic $n \times n$ matrix and w
be its left dominant eigenvector. The matrix A is *reversible* if

$$
w_i A_{ij} = A_{ji} w_j, \quad \text{for all } i, j \in \{1, \ldots, n\}, \tag{E5.1}
$$

or, equivalently,

$$
\text{diag}(w)A = A^T \text{diag}(w).
$$

Prove the following statements:

(i) if A is reversible, then its associated digraph is undirected, that is, if (i, j) is an edge, then so is (j, i),

(ii) if A is reversible, then $\operatorname{diag}(w)^{1/2} A \operatorname{diag}(w)^{-1/2}$ is symmetric and, hence, A has n real eigenvalues and n eigenvectors, and

(iii) if A is an equal-neighbor matrix for an unweighted undirected graph, then A is reversible.

Recall that, for $w = (w_1, \ldots, w_n) > 0$, the following definitions hold: $\operatorname{diag}(w)^{1/2} = \operatorname{diag}(\sqrt{w_1}, \ldots, \sqrt{w_n})$ and $\operatorname{diag}(w)^{-1/2} = \operatorname{diag}(1/\sqrt{w_1}, \ldots, 1/\sqrt{w_n})$.

E5.5 **A stubborn agent.** Pick $\alpha \in\ ]0, 1[$, and consider the discrete-time averaging algorithm

$$x_1(k + 1) = x_1(k),$$
$$x_2(k + 1) = \alpha x_1(k) + (1 - \alpha)x_2(k).$$

Perform the following tasks:

(i) compute the matrix A representing this algorithm and verify it is row-stochastic,

(ii) compute the eigenvalues and eigenvectors of A,

(iii) draw the directed graph G representing this algorithm and discuss its connectivity properties,

(iv) compute the condensation digraph of G,

(v) compute the final value of this algorithm as a function of the initial values in two alternate ways:

    a) invoking Exercise E2.10, and
    b) invoking Theorem 5.1.

E5.6 **Agents with self-confidence levels.** Consider 2 agents, labeled $+1$ and -1, described by the self-confidence levels s_{+1} and s_{-1}. Assume $s_{+1} \geq 0$, $s_{-1} \geq 0$, and $s_{+1} + s_{-1} = 1$. For $i \in \{+1, -1\}$, define

$$x_i^+ := s_i x_i + (1 - s_i)x_{-i}.$$

Perform the following tasks:

(i) compute the matrix A representating this algorithm and verify it is row-stochastic,

(ii) compute A^2,

(iii) compute the eigenvalues, the right eigenvectors, and the left eigenvectors of A,

(iv) compute the final value of this algorithm as a function of the initial values and of the self-confidence levels. Is it true that an agent with higher self-confidence makes a larger contribution to the final value?

E5.7 **Persistent disagreement and the Friedkin-Johnsen model of opinion dynamics (Friedkin and Johnsen, 1999).** Let A be a row-stochastic matrix whose associated digraph describes an *interpersonal influence network*. Let each individual possess an *openness level* $\lambda_i \in [0, 1]$, $i \in \{1, \ldots, n\}$, descring how open is the individual to changing her initial opinion about a subject; set $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$. Consider the *Friedkin-Johnsen model* of opinion dynamics

$$x(k + 1) = \Lambda A x(k) + (I_n - \Lambda)x(0). \tag{E5.2}$$

In other words, in this model, each individual i exhibits an attachment $(1 - \lambda_i)$ to its initial opinion $x_i(0)$, $x_i(k)$ represents the current opinion and $x_i(0)$ represents a prejudice by individual i. Consider the following two assumptions:

(A1) at least one individual has a strictly positive attachment to its initial opinion, that is, $\lambda_i < 1$ for at least one individual i; and

(A2) the interpersonal influence network contains directed paths from each individual with openness level equal to 1 to an individual with openness level less than 1.

Note that, if Assumption (A1) is not satisfied and therefore $\Lambda = I_n$, then we recover the DeGroot opinion dynamics model introduced in Section 1.1 and analyzed in this chapter. In what follows, let Assumption (A1) hold.

   (i) Show that the matrix ΛA is convergent if and only if Assumption (A2) holds.
       Hint: *Recall Corollary 4.11*

Next, under Assumption (A2), perform the following tasks:

  (ii) show that the matrix $V = (I_n - \Lambda A)^{-1}(I_n - \Lambda)$ is well-defined and row-stochastic,
     Hint: *Review Exercises E2.10 and E2.12*
  (iii) show that the limiting opinions satisfy $\lim_{k \to +\infty} x(k) = V x(0)$,
  (iv) write the n-dimensional Friedkin-Johnsen model (E5.2) as a $2n$-dimensional averaging model $\bar{x}(k+1) = \bar{A}\bar{x}(k)$, for an appropriate row-stochastic matrix $\bar{A} \in \mathbb{R}^{2n \times 2n}$,
  (v) show that A and V have the same left dominant eigenvector when $\Lambda = \lambda I_n$, for $0 < \lambda < 1$,
  (vi) compute the matrix V and state whether two agents will achieve consensus or mantain persistent disagreement for the following pairs of matrices:

$$A_1 = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}, \text{ and } \Lambda_1 = \mathrm{diag}(1/2, 1),$$

$$A_2 = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}, \text{ and } \Lambda_2 = \mathrm{diag}(1/4, 3/4).$$

Note: *Friedkin and Johnsen (1999, 2011) make the additional assumption that $\lambda_i = 1 - a_{ii}$, for $i \in \{1, \ldots, n\}$; this assumption couples the openness level with the interpersonal influences and has the effect of enhancing stubborness of the individuals. This assumption is not needed here. The model (E5.2) is also referred to the averaging model with stubborn agents. Other properties of this model are studied in (Bindel et al., 2015; Friedkin et al., 2016; Ravazzi et al., 2015).*

E5.8   **Necessary and sufficient conditions for semi-convergence.** Theorem 5.2 provides a sufficient condition for a row-stochastic matrix to be semi-convergent. We now provide a necessary a sufficient counterpart.
    Let A be a row-stochastic matrix with M consended sinks. Prove that the following statements are equivalent:

  (i) the eigenvalue 1 is semi-simple with multiplicity M and all other eigenvalues have magnitude strictly smaller than 1,
  (ii) A is semi-convergent,
  (iii) each sink in the condensation digraph associated to A is aperiodic.

Note: *Gantmacher (1959) calls "regular" the semi-convergent row-stochastic matrices and "fully regular" the semi-convergent row-stochastic matrices whose limiting matrix has rank one, i.e., the indecomposable row-stochastic matrices.*

E5.9   **Average consensus via the parallel averaging algorithm.** Let G be a weighted graph with weighted adjacency matrix A and weighted degrees $d_1, \ldots, d_n$. Assume G is connected and aperiodic and consider the equal-neighbor matrix $A_{\mathrm{en}} = \mathrm{diag}(d_1, \ldots, d_n)^{-1}A$. Assign a value $x_i \in \mathbb{R}$ to each node i and consider the *parallel averaging algorithm:*
    1: each node i sets $y_i(0) = 1/d_i$ and $z_i(0) = x_i/d_i$
    2: the nodes run the averaging algorithms $y(k+1) = A_{\mathrm{en}}y(k)$ and $z(k+1) = A_{\mathrm{en}}z(k)$ for $k \in \mathbb{Z}_{\geq 0}$
    3: each node i sets $x_i(k) = z_i(k)/y_i(k)$ at each $k \in \mathbb{Z}_{\geq 0}$
  Show that the parallel averaging algorithm

(i) is well posed, i.e., $y_i(k)$ does not vanish for any $i \in \{1, \ldots, n\}$ and $k \in \mathbb{Z}_{\geq 0}$, and

(ii) achieves average consensus, that is, $\lim_{k \to \infty} x(k) = \operatorname{average}(x_1, \ldots, x_n)\mathbb{1}_n$.

Note: This algorithm is also referred to as the push sum iteration, *because it may implemented over directional communication by "summing the pushed variables." This algorithm was originally introduced by Kempe et al. (2003) and later studied in (Olshevsky and Tsitsiklis, 2009; Benezit et al., 2010).*

E5.10 **Computing centrality.** Write in your favorite programming language algorithms to compute degree, eigenvector, Katz and pagerank centralities. Compute these four centralities for the following undirected unweighted graphs (without self-loops):

(i) the cycle graph with 5 nodes;

(ii) the star graph with 5 nodes;

(iii) the line graph with 5 nodes; and

(iv) the Zachary karate club network dataset. This dataset can be downloaded for example from: `http://konect.uni-koblenz.de/networks/ucidata-zachary`

To compute Katz centrality of a matrix A, select $\alpha = 1/(2\rho(A))$. For pagerank, use $\alpha = 1/2$.

Hint: Recall that pagerank centrality is well-defined for a row-stochastic matrix.

E5.11 **Central nodes in example graph.** For the unweighted undirected graph in Figure 5.11, verify (with the aid of a computational package) that the dark nodes have indeed the largest degree, eigenvector, closeness and betweenness centrality as stated in the figure caption.

E5.12 **Iterative computation of Katz centrality.** Given a graph with adjacency matrix A, show that the solution to the iteration $x(k+1) := \alpha A^\mathsf{T}(x(k) + \mathbb{1}_n)$ with $\alpha < 1/\rho(A)$ converges to the Katz centrality vector c_K, for all initial conditions $x(0)$.

E5.13 **Move away from your nearest neighbor and reducible averaging.** Consider $n \geq 3$ robots with positions $p_i \in \mathbb{R}$, $i \in \{1, \ldots, n\}$, dynamics $p_i(t+1) = u_i(t)$, where $u_i \in \mathbb{R}$ is a steering control input. For simplicity, assume that the robots are indexed according to their initial position: $p_1(0) \leq p_2(0) \leq p_3(0) \leq \cdots \leq p_n(0)$. Consider two walls at the positions $p_0 \leq p_1(0)$ and $p_{n+1} \geq p_n(0)$ so that all robots are contained between the walls. The walls are stationary, that is, $p_0(t+1) = p_0(t) = p_0$ and $p_{n+1}(t+1) = p_{n+1}(t) = p_{n+1}$.

Consider the following coordination law: robots $i \in \{2, \ldots, n-1\}$ (each having two neighbors) move to the centroid of the local subset $\{p_{i-1}, p_i, p_{i+1}\}$. The robots $\{1, n\}$ (each having one robotic neighbor and one neighboring wall) move to the centroid of the local subsets $\{p_0, p_1, p_2\}$ and $\{p_{n-1}, p_n, p_{n+1}\}$, respectively. Hence, the closed-loop robot dynamics are

$$p_i(t+1) = \frac{1}{3}(p_{i-1}(t) + p_i(t) + p_{i+1}(t)), \quad i \in \{1, \ldots, n\}.$$

Show that the robots become uniformly spaced on the interval $[p_0, p_{n+1}]$ using Theorem 5.2.

E5.14 **The role of the nodal degree in averaging systems.** Let G be a connected undirected graph without self-loops. Consider the averaging dynamics:

$$x(k+1) = Ax(k),$$

where $A = D^{-1}A_{01}$, D is the degree matrix, and A_{01} is the binary adjacency matrix of G.

(i) Under which conditions on G will the system converge to a final consensus state, i.e., an element of $\operatorname{span}\{\mathbb{1}_n\}$?

(ii) Assuming each state converges to a final consensus value, what is this steady state value?

(iii) Let $e(k) = x(k) - \lim_{k \to \infty} x(k)$ be the disagreement error at time instant k. Show that the error dynamics is linear, that is, of the form $e(k+1) = Be(k)$ and determine the matrix B.

(iv) Find a function $f(k, \lambda_2, \ldots, \lambda_n, d_1, \ldots, d_n)$ depending on the time step k, the eigenvalues $\lambda_2, \ldots, \lambda_n$ of A, and the degrees of the nodes $d_1, \ldots, d_n$ such that

$$\|e(k)\|_2 \leq f(k, \lambda_2, \ldots, \lambda_n, d_1, \ldots, d_n)\|e(0)\|_2.$$

E5.15 **Hubs and authorities (Kleinberg, 1999).** Let G be a digraph with vertex set $V = \{1, \ldots, n\}$ and edge set E. Assume G has a globally reachable node and the subgraph of globally reachable nodes is aperiodic.

We define two scores for each vertex $j \in \{1, \ldots, n\}$: the *hub score* $h_j \in \mathbb{R}$ and the *authority score* $a_j \in \mathbb{R}$. We initialize these scores with positive values and updated them simultaneously for all vertices according to the following mutually reinforcing relation: the hub score of vertex j is set equal to the sum of the authority scores of all vertices pointed to by j, and, similarly, the authority score of vertex j is set equal to the sum of the hub scores of all vertices pointing to j. In concise formulas, for $k \in \mathbb{N}$,

$$\begin{cases} h_j(k+1) = \sum_{i: (j,i) \in E} a_i, \\ a_j(k+1) = \sum_{i: (i,j) \in E} h_i. \end{cases} \tag{E5.3}$$

(i) Let $x(k) = \begin{bmatrix} h(k)^{\mathsf{T}} & a(k)^{\mathsf{T}} \end{bmatrix}^{\mathsf{T}}$ denote the stacked vector of hub and authority scores. Provide an update equation for the hub and authority scores of the form

$$x(k+1) = Mx(k),$$

for some matrix $M \in \mathbb{R}^{2n \times 2n}$.

(ii) Will the sequence $x(k)$ converge as $k \to \infty$?

In what follows, we consider the modified iteration

$$y(k+1) = \frac{My(k)}{\|My(k)\|_2},$$

where M is defined as in statement (i) above.

(iii) Will the sequence $y(k)$ converge as $k \to \infty$?

(iv) Show that the two subsequences of even and odd iterates, $k \mapsto y(2k)$ and $k \mapsto y(2k+1)$, converge, that is,

$$\lim_{k \to \infty} y(2k) = y_{\text{even}}(y_0), \quad \lim_{k \to \infty} y(2k+1) = y_{\text{odd}}(y_0),$$

where $y_0 = x(0)$ is the stacked vector of initial hub and authority scores.

(v) Provide expressions for $y_{\text{even}}(y_0)$ and $y_{\text{odd}}(y_0)$.

E5.16 **Maximum entropy random walk (Burda et al., 2009).** Let G be an unweighted connected graph with binary adjacency matrix $A \in \{0,1\}^n$. Let (λ, v) be the dominant eigenpair, i.e, $Av = \lambda v$ and $\mathbb{1}_n^{\mathsf{T}} v = 1$. Similarly to E4.11, define the square matrix P by

$$p_{ij} = \frac{1}{\lambda} \frac{v_j}{v_i} a_{ij}, \quad \text{for } i, j \in \{1, \ldots, n\}.$$

Perform the following tasks:

(i) show that P is well defined, row stochastic, and irreducible,

(ii) pick $i, j \in \{1, \ldots, n\}$ and $k \geq 1$. Assuming there exists a path of length k from i to j, let $c_{ij}^{[k]}$ denote the product of the edge weights along the path and show that

$$c_{ij}^{[k]} = \frac{1}{\lambda^k} \frac{v_j}{v_i},$$

(iii) let $w > 0$ be the left dominant eigenvector of P, normalized so that $\mathbb{1}_n^\mathsf{T} w = 1$, and show that

$$w_i = \frac{1}{\|v\|_2^2} v_i^2.$$

The Laplacian Matrix

The previous chapters studied adjacency matrices and their application to discrete-time averaging dynamics. This chapter introduces and characterizes a second relevant matrix associated to a digraph, called the Laplacian matrix. Laplacian matrices appear in numerous applications and enjoy numerous useful properties.

6.1 The Laplacian matrix

Definition 6.1 (Laplacian matrix of a digraph). *Given a weighted digraph G with adjacency matrix A and out-degree matrix D_{out}, the* Laplacian matrix *of G is*

$$L = D_{\text{out}} - A.$$

In components $L = (\ell_{ij})_{i,j \in \{1,\dots,n\}}$

$$\ell_{ij} = \begin{cases} -a_{ij}, & \text{if } i \neq j, \\ \displaystyle\sum_{h=1,h\neq i}^{n} a_{ih}, & \text{if } i = j, \end{cases}$$

or, for an unweighted undirected graph,

$$\ell_{ij} = \begin{cases} -1, & \text{if } \{i,j\} \text{ is an edge and not self-loop,} \\ d(i), & \text{if } i = j, \\ 0, & \text{otherwise.} \end{cases}$$

An example is illustrated in Figure 6.1.

Note:

(i) the sign pattern of L is important — diagonal elements are non-negative (zero or positive) and off-diagonal elements are non-positive (zero or negative);

(ii) the Laplacian matrix L of a digraph G does not depend upon the existence and values of self-loops in G;

$$L = \begin{bmatrix} 7.4 & -3.7 & -3.7 & 0 & 0 \\ -8.9 & 10.1 & 0 & -1.2 & 0 \\ 0 & 0 & 6.0 & -3.7 & -2.3 \\ 0 & 0 & 0 & 0 & 0 \\ -4.4 & 0 & 0 & -2.3 & 6.7 \end{bmatrix}.$$
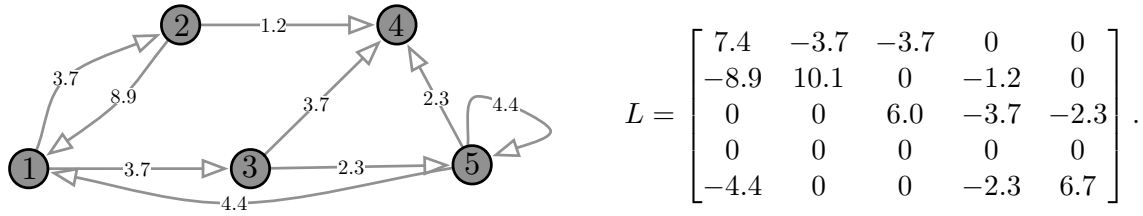
Figure 6.1: A weighted digraph and its Laplacian matrix

(iii) the graph G is undirected (i.e., symmetric adjacency matrix) if and only if L is symmetric. In this case, $D_{\text{out}} = D_{\text{in}} = D$ and $A = A^{\mathsf{T}}$;

(iv) in a directed graph, $\ell_{ii} = 0$ (instead of $\ell_{ii} > 0$) if and only if node i has zero out-degree;

(v) L is said to be *irreducible* if G is strongly connected.

We conclude this section with some useful equalities. By the way, obviously

$$(Ax)_i = \sum_{j=1}^n a_{ij} x_j. \tag{6.1}$$

First, for $x \in \mathbb{R}^n$,

$$(Lx)_i = \sum_{j=1}^n \ell_{ij} x_j = \ell_{ii} x_i + \sum_{j=1,j\neq i}^n \ell_{ij} x_j = \Big(\sum_{j=1,j\neq i}^n a_{ij} \Big) x_i + \sum_{j=1,j\neq i}^n (-a_{ij}) x_j$$

$$= \sum_{j=1,j\neq i}^n a_{ij}(x_i - x_j) = \sum_{j\in\mathcal{N}^{\text{out}}(i)} a_{ij}(x_i - x_j) \tag{6.2}$$

$$\stackrel{\text{for unit weights}}{=} d_{\text{out}}(i)\big(x_i - \text{average}(\{x_j, \text{ for all out-neighbors } j\})\big).$$

Second, assume $L = L^{\mathsf{T}}$ (i.e., $a_{ij} = a_{ji}$) and compute:

$$x^{\mathsf{T}} Lx = \sum_{i=1}^n x_i (Lx)_i = \sum_{i=1}^n x_i \Big(\sum_{j=1,j\neq i}^n a_{ij}(x_i - x_j) \Big)$$

$$= \sum_{i,j=1}^n a_{ij} x_i (x_i - x_j) = \Big(\frac{1}{2} + \frac{1}{2}\Big) \sum_{i,j=1}^n a_{ij} x_i^2 - \sum_{i,j=1}^n a_{ij} x_i x_j$$

$$\stackrel{\text{by symmetry}}{=} \frac{1}{2} \sum_{i,j=1}^n a_{ij} x_i^2 + \frac{1}{2} \sum_{i,j=1}^n a_{ij} x_j^2 - \sum_{i,j=1}^n a_{ij} x_i x_j$$

$$= \frac{1}{2} \sum_{i,j=1}^n a_{ij}(x_i - x_j)^2 \tag{6.3}$$

$$= \sum_{\{i,j\}\in E} a_{ij}(x_i - x_j)^2. \tag{6.4}$$

These equalities are useful because it is common to encounter the "array of differences" Lx and the quadratic "error" or "disagreement" function $x^\mathsf{T} Lx$. They provide the correct intuition for the definition of the Laplacian matrix. In some literature, the function $x \mapsto x^\mathsf{T} Lx$ is referred to as the *Laplacian potential function*, because of the energy and power interpetation we present in the next two examples.

6.1.1 The Laplacian in mechanical networks of springs
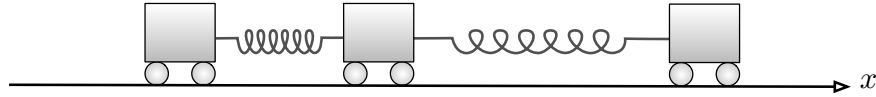
Let $x_i \in \mathbb{R}$ denote the displacement of the ith rigid body. Assume that each spring is ideal linear-elastic and let a_{ij} be the spring constant for the spring connecting the ith and jth bodies.

Define a graph as follows: the nodes are the rigid bodies $\{1, \dots, n\}$ with locations $x_1, \dots, x_n$, and the edges are the springs with weights a_{ij}. Each node i is subject to a force

$$\mathrm{F}_i = \sum_{j \neq i} a_{ij}(x_j - x_i) = -(Lx)_i,$$

where L is the Laplacian for the network of springs (modeled as an undirected weighted graph). Moreover, recalling that the spring $\{i, j\}$ stores the quadratic energy $\frac{1}{2}a_{ij}(x_i - x_j)^2$, the total elastic energy is

$$\mathrm{E}_{\mathrm{elastic}} = \frac{1}{2} \sum_{\{i,j\} \in E} a_{ij}(x_i - x_j)^2 = \frac{1}{2}x^\mathsf{T} Lx.$$

In this role, the Laplacian matrix is referred to as the stiffness matrix. Stiffness matrices can be defined for spring networks in arbitrary dimensions (not only on the line) and with arbitrary topology (not only a chain graph, or line graph, as in figure). More complex spring networks can be found, for example, in finite-element discretization of flexible bodies and finite-difference discretization of diffusive media.

6.1.2 The Laplacian in electrical networks of resistors
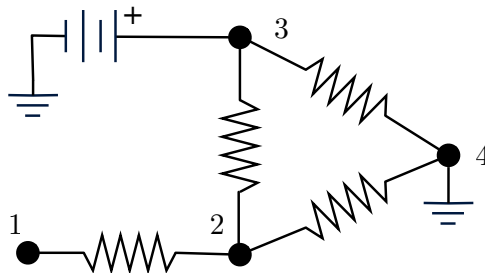
Suppose the graph is an *electrical network* with only pure resistors and ideal voltage sources: (i) each graph vertex $i \in \{1, \dots, n\}$ is possibly connected to an ideal voltage source, (ii) each edge is a resistor, say with resistance r_{ij} between nodes i and j. (This is an undirected weighted graph.)

Ohm's law along each edge $\{i, j\}$ gives the current flowing from i to j as

$$c_{i \to j} = (v_i - v_j)/r_{ij} = a_{ij}(v_i - v_j),$$

where v_i is the voltage at node i and a_{ij} is the inverse resistance, called conductance. We set $a_{ij} = 0$ whenever two nodes are not connected by a resistance and let L denote the Laplacian matrix of conductances. Kirchhoff's current law says that at each node i:

$$c_{\text{injected at } i} = \sum_{j=1, j \neq i}^{n} c_{i \to j} = \sum_{j=1, j \neq i}^{n} a_{ij}(v_i - v_j).$$

Hence, the vector of injected currents c_{injected} and the vector of voltages at the nodes v satisfy

$$c_{\text{injected}} = L v.$$

Moreover, the power dissipated on resistor $\{i, j\}$ is $c_{i \to j}(v_i - v_j)$, so that the total dissipated power is

$$P_{\text{dissipated}} = \sum_{\{i,j\} \in E} a_{ij}(v_i - v_j)^2 = v^{\mathsf{T}} L v.$$

6.2 Properties of the Laplacian matrix

Lemma 6.2 (Zero row-sums). *Let G be a weighted digraph with Laplacian L and n nodes. Then*

$$L \mathbb{1}_n = \mathbb{0}_n.$$

In equivalent words, 0 is an eigenvalue of L with eigenvector $\mathbb{1}_n$.

Proof. For all rows i, the ith row-sum is zero:

$$\sum_{j=1}^{n} \ell_{ij} = \ell_{ii} + \sum_{j=1, j \neq i}^{n} \ell_{ij} = \Big(\sum_{j=1, j \neq i}^{n} a_{ij} \Big) + \sum_{j=1, j \neq i}^{n} (-a_{ij}) = 0.$$

Equivalently, in vector format (remembering the weighted out-degree matrix D_{out} is diagonal and contains the row-sums of A):

$$L \mathbb{1}_n = D_{\text{out}} \mathbb{1}_n - A \mathbb{1}_n = \begin{bmatrix} d_{\text{out}}(1) \\ \vdots \\ d_{\text{out}}(n) \end{bmatrix} - \begin{bmatrix} d_{\text{out}}(1) \\ \vdots \\ d_{\text{out}}(n) \end{bmatrix} = \mathbb{0}_n.$$

$\blacksquare$

Based on this lemma, we now extend the notion of Laplacian matrix to a setting in which there is no digraph to start with.

Definition 6.3 (Laplacian matrix). *A matrix $L \in \mathbb{R}^{n \times n}$, $n \geq 2$, is* Laplacian *if*

(i) *its row-sums are zero,*

(ii) *its diagonal entries are non-negative, and*

(iii) *its non-diagonal entries are non-positive.*

A Laplacian matrix L induces a weighted digraph G without self-loops in the natural way, that is, by letting (i, j) be an edge of G if and only if $\ell_{ij} > 0$. With this definition, L is the Laplacian matrix of G.

Lemma 6.4 (Zero column-sums). *Let G be a weighted digraph with Laplacian L and n nodes. The following statements are equivalent:*

(i) *G is weight-balanced, i.e., $D_{\text{out}} = D_{\text{in}}$; and*

(ii) *$\mathbb{1}_n^\mathsf{T} L = \mathbb{0}_n^\mathsf{T}$.*

Proof. Pick $j \in \{1, \ldots, n\}$ and compute

$$(\mathbb{1}_n^\mathsf{T} L)_j = (L^\mathsf{T} \mathbb{1}_n)_j = \sum_{i=1}^n \ell_{ij} = \ell_{jj} + \sum_{i=1, j \neq i}^n \ell_{ij} = d_{\text{out}}(j) - d_{\text{in}}(j),$$

where the last equality follows from

$$\ell_{jj} = d_{\text{out}}(j) - a_{jj} \quad \text{and} \quad \sum_{i=1, j \neq i}^n \ell_{ij} = -(d_{\text{in}}(j) - a_{jj}).$$

In summary, we know that $\mathbb{1}_n^\mathsf{T} L = \mathbb{0}_n^\mathsf{T}$ if and only if $D_{\text{out}} = D_{\text{in}}$. ∎

Lemma 6.5 (Spectrum of the Laplacian matrix). *Given a weighted digraph G with Laplacian L, the eigenvalues of L different from 0 have strictly-positive real part.*

Proof. Recall $\ell_{ii} = \sum_{j=1, j \neq i}^n a_{ij} \geq 0$ and $\ell_{ij} = -a_{ij} \leq 0$ for $i \neq j$. By the Geršgorin Disks Theorem 2.8, we know that each eigenvalue of L belongs to at least one of the disks

$$\left\{ z \in \mathbb{C} \mid |z - \ell_{ii}| \leq \sum_{j=1, j \neq i}^n |\ell_{ij}| \right\} = \left\{ z \in \mathbb{C} \mid |z - \ell_{ii}| \leq \ell_{ii} \right\}.$$
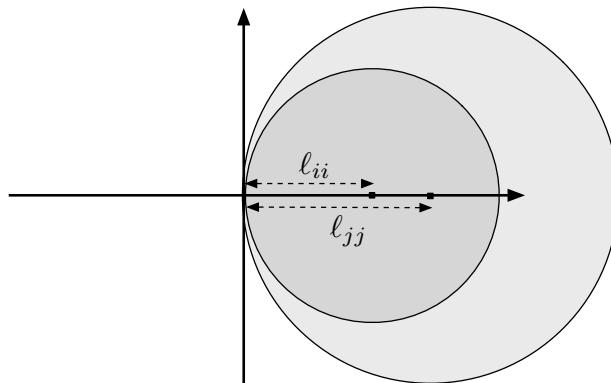
These disks, with radius equal to the center, contain the origin and complex numbers with positive real part.   ∎

For an undirected graph without self-loops and with symmetric adjacency matrix $A = A^\mathsf{T}$, we know that L is symmetric and positive semidefinite, i.e., all eigenvalues of L are real and non-negative and that $d(i) = \ell_i$. In this case, by convention, we write these eigenvalues as

$$0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n.$$

Note:

- the second smallest eigenvalue λ_2 is called the *Fiedler eigenvalue* or the *algebraic connectivity*, due to the early work by Fiedler (1973); and

- we refer the reader to Exercise E6.3 for a lower bound and an upper bound on λ_n based on the maximum degree.

6.3   Graph connectivity and the rank of the Laplacian

Theorem 6.6 (Rank of the Laplacian). *Let L be the Laplacian matrix of a weighted digraph G with n nodes. Let d be the number of sinks in the condensation digraph of G. Then*

$$\mathrm{rank}(L) = n - d.$$

This theorem has the following immediate consequences:

(i) a digraph G contains a globally reachable vertex if and only if $\mathrm{rank}(L) = n - 1$ (also recall the properties of $C(G)$ from Lemma 3.1); and

(ii) for the case of undirected graphs, we have the following two results: the rank of L is equal to n minus the number of connected components of G and an undirected graph G is connected if and only if $\lambda_2 > 0$.

Proof. We start by simplifying the problem. Define a new weighted digraph $\bar{G}$ by modifying G as follows: at each node, add a self-loop with unit weight if no self-loop is present, or increase the weight of the self-loop by 1 if a self-loop is present. Also, define another weighted digraph $\bar{\bar{G}}$ by modyfing $\bar{G}$ as follows: for each node, divide the weights of its out-going edges by its out-degree, so that the out-degree of each node is 1. In other words, define $\bar{A} = A + I$ and $\bar{L} = L$, and define $\bar{\bar{A}} = \bar{D}_{\mathrm{out}}^{-1}\bar{A}$ and $\bar{\bar{L}} = \bar{D}_{\mathrm{out}}^{-1}\bar{L} = I_n - \bar{\bar{A}}$. Clearly, the rank of L is equal to the rank of $\bar{\bar{L}}$. Therefore, without loss of generality, we consider in what follows only digraphs with row-stochastic adjacency matrices.

Because the condensation digraph $C(G)$ has d sinks, after a renumbering of the nodes, that is, a permutation of rows and columns (see Exercise E3.1), the adjacency matrix A can be written in block lower

tridiagonal form as

$$
A = \begin{bmatrix}
A_{11} & 0 & 0 & \cdots & 0 & 0 \\
0 & A_{22} & 0 & \ddots & \ddots & 0 \\
0 & 0 & \ddots & \ddots & \ddots & \vdots \\
\vdots & \ddots & \ddots & \ddots & \ddots & 0 \\
0 & \ddots & \ddots & 0 & A_{dd} & 0 \\
A_{1o} & A_{2o} & \cdots & \cdots & A_{do} & A_{\text{others}}
\end{bmatrix} \in \mathbb{R}^{n \times n}.
$$

where the state vector x is correspondingly partitioned into the vectors $x_1, \dots, x_d$ and x_{others} of dimensions $n_1, \dots, n_d$ and $n - (n_1 + \cdots + n_d)$ respectively, corresponding to the d sinks and all other nodes.

Each sink of $C(G)$ is a strongly connected and, therefore, the non-negative square matrices $A_{11}, \dots, A_{dd}$ are irreducible. The Perron–Frobenius Theorem for irreducible matrices 2.12 implies that the number 1 is a simple eigenvalue for each of them.

The square matrix A_{others} is non-negative and it can itself be written as a block lower triangular matrix, whose diagonal block matrices, say $(A_{\text{others}})_1, \dots, (A_{\text{others}})_N$ are non-negative and irreducible. Moreover, each of these diagonal block matrices must be row-substochastic because (1) each row-sum for each of these matrices is at most 1, and (2) at least one of the row-sums of each of these matrices must be smaller than 1, otherwise that matrix would correspond to a sink of $C(G)$. In summary, because the matrices $(A_{\text{others}})_1, \dots, (A_{\text{others}})_N$ are irreducible and row-substochastic, the matrix A_{others} has spectral radius $\rho(A_{\text{others}}) < 1$.

We now write the Laplacian matrix $L = I_n - A$ with the same block lower triangular structure:

$$
L = \begin{bmatrix}
L_{11} & 0 & 0 & \cdots & 0 & 0 \\
0 & L_{22} & 0 & \ddots & \ddots & 0 \\
0 & 0 & \ddots & \ddots & \ddots & \vdots \\
\vdots & \ddots & \ddots & \ddots & \ddots & 0 \\
0 & \ddots & \ddots & 0 & L_{dd} & 0 \\
-A_{1o} & -A_{2o} & \cdots & \cdots & -A_{do} & L_{\text{others}}
\end{bmatrix}, \tag{6.5}
$$

where, for example, $L_{11} = I_{n_1} - A_{11}$. Because the number 1 is a simple eigenvalue of A_{11}, the number 0 is a simple eigenvalue of L_{11}. Therefore, $\mathrm{rank}(L_{11}) = n_1 - 1$. This same argument establishes that the rank of L is at most $n - d$ because each one of the matrices $L_{11}, \dots, L_{dd}$ is of rank $n_1 - 1, \dots, n_d - 1$, respectively. Finally, we note that the rank of L_{others} is maximal, because $L_{\text{others}} = I - A_{\text{others}}$ and $\rho(A_{\text{others}}) < 1$ together imply that 0 is not an eigenvalue for L_{others}. ∎

6.4 Appendix: Community detection via algebraic connectivity

As just presented, the algebraic connectivity λ_2 of an undirected and weighted graph G is positive if and only if G is connected. We build on this insight and show that the algebraic connectivity does not only

provide a binary connectivity measure, but it also quantifies the "bottleneck" of the graph. To develop this intuition, we study the problem of *community detection* in a large-scale undirected graph. This problem arises, for example, when identifying group of friends in a social network by means of the interaction graph.

Specifically, we consider the problem of partitioning the vertices V of an undirected connected graph G in two sets V_1 and V_2 so that

$$V_1 \cup V_2 = V, \ V_1 \cap V_2 = \emptyset, \ \text{ and } V_1, V_2 \neq \emptyset.$$

Of course, there are many such partitions. We measure the quality of a partition by the sum of the weights of all edges that need to be *cut* to separate the vertices V_1 and V_2 into two disconnected components. Formally, the *size of the cut* separating V_1 and V_2 is

$$J = \sum_{i \in V_1, j \in V_2} a_{ij} \, .$$

We are interested in finding the cut with minimal size that identifies the two groups of nodes that are most loosely connected. The problem of minimizing the cut size J is combinatorial and computationally hard since we need to consider all possible partitions of the vertex set V. We present here a tractable approach based on a so-called relaxation step. First, define a vector $x \in \{-1, +1\}^n$ with entries $x_i = 1$ for $i \in V_1$ and $x_i = -1$ for $i \in V_2$. Then the cut size J can be rewritten via the Laplacian potential as

$$J = \frac{1}{4} \sum_{i,j=1}^{n} a_{ij}(x_i - x_j)^2 = \frac{1}{2}x^{\mathsf{T}}Lx$$

and the minimum cut size problem is:

$$\underset{x \in \{-1,1\}^n \setminus \{-\mathbb{1}_n, \mathbb{1}_n\}}{\text{minimize}} x^{\mathsf{T}}Lx.$$

(Here we exclude the cases $x \in \{-\mathbb{1}_n, \mathbb{1}_n\}$ because they correspond to one of the two groups being empty.) Second, since this problem is still computationally hard, we relax the problem from binary decision variables $x_i \in \{-1, +1\}$ to continuous decision variables $y_i \in [-1, 1]$ (or $\|y\|_\infty \leq 1$), where we exclude $y \in \text{span}(\mathbb{1}_n)$ (corresponding to one of the two groups being empty). Then the minimization problem becomes

$$\underset{y \in \mathbb{R}^n, y \perp \mathbb{1}_n, \|y\|_\infty = 1}{\text{minimize}} y^{\mathsf{T}}Ly.$$

As a third and final step, we consider a 2-norm constraint $\|y\|_2 = 1$ instead of an ∞-norm constraint $\|y\|_\infty = 1$ (recall that $\|y\|_\infty \leq \|y\|_2 \leq \sqrt{n}\|y\|_\infty$) to obtain the following heuristic:

$$\underset{y \in \mathbb{R}^n, y \perp \mathbb{1}_n, \|y\|_2 = 1}{\text{minimize}} y^{\mathsf{T}}Ly.$$

Notice that $y^{\mathsf{T}}Ly \geq \lambda_2 \|y\|^2$ and this inequality holds true with equality whenever $y = v_2$, the normalized eigenvector associated to λ_2. Thus, the unique minimum of the relaxed optimization problem is λ_2 and the minimizer is $y = v_2$. We can then use as a heuristic $x = \text{sign}(v_2)$ to find the desired partition $\{V_1, V_2\}$. Hence, the algebraic connectivity λ_2 is an estimate for the size of the minimum cut, and the signs of the

entries of v_2 identify the associated partition in the graph. For these reasons λ_2 and v_2 can be interpreted as the size and the location of a "bottleneck" in a graph.

To illustrate the above concepts, we borrow an example problem with the corresponding Matlab code from (Gleich, 2006). we construct a randomly generated graph as follows. First, we partition $n = 1000$ nodes in two groups V_1 and V_2 of sizes 450 and 550 nodes, respectively. Second, we connect any pair of nodes in the set V_1 (respectively V_2) with probability 0.3 (respectively 0.2). Third and finally, any two nodes in distinct groups, $i \in V_1$ and $j \in V_2$, are connected with a probability of 0.1. The sparsity pattern of the associated adjacency matrix is shown in the left panel of Figure 6.2. No obvious partition is visible at first glance since the indices are not necessarily sorted, that is, V_1 is not necessarily $\{1, \ldots, 450\}$. The second panel displays the entries of the eigenvector v_2 sorted according to their magnitude showing a sharp transition between positive and negative entries. Finally, the third panel displays the correspondingly sorted adjacency matrix $\tilde{A}$ clearly indicating the partition $V = V_1 \cup V_2$.

The Matlab code to generate Figure 6.2 can be found below. For additional analysis of this problem, we refer the reader to (Gleich, 2006).

A $\qquad$ $\tilde{v}_2$ $\qquad$ $\tilde{A}$
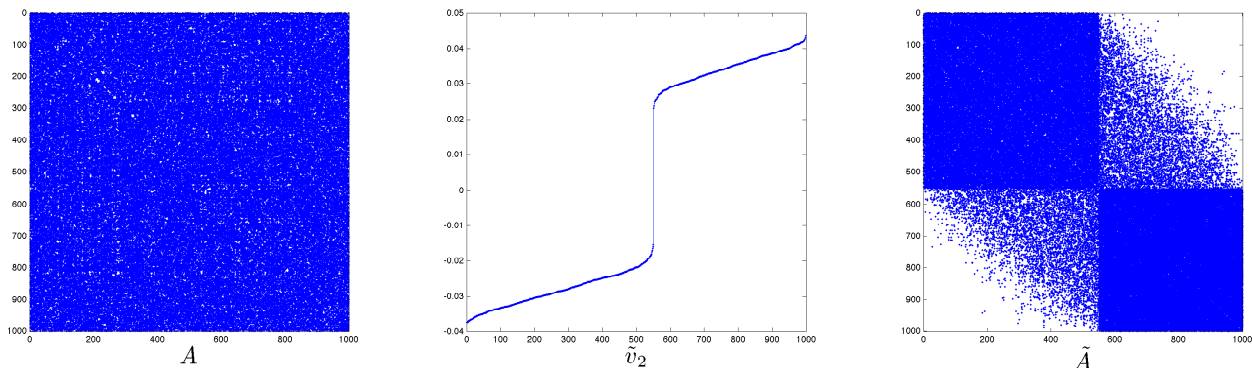
Figure 6.2: The first panel shows a randomly-generated sparse adjacency matrix A for a graph with 1000 nodes. The second panel displays the eigenvector $\tilde{v}_2$ which is identical to the normalized eigenvector v_2 after sorting the entries according to their magnitude, and the third panel displays the correspondingly sorted adjacency matrix $\tilde{A}$.

```
1   % choose a graph size
2   n = 1000;
3
4   % randomly assign the nodes to two grous
5   x = randperm(n);
6   group_size = 450;
7   group1 = x(1:group_size);
8   group2 = x(group_size+1:end);
9
10  % assign probabilities of connecting nodes
11  p_group1 = 0.3;
12  p_group2 = 0.2;
13  p_between_groups = 0.1;
14
15  % construct adjacency matrix
```

```
16  A(group1, group1) = rand(group_size,group_size) < p_group1;
17  A(group2, group2) = rand(n−group_size,n−group_size) < p_group2;
18  A(group1, group2) = rand(group_size, n−group_size) < p_between_groups;
19  A = triu(A,1); A = A + A';
20
21  % can you see the groups?
22  subplot(1,3,1); spy(A);
23  xlabel('$A$', 'Interpreter','latex','FontSize',28);
24
25  % construct Laplacian and its spectrum
26  L = diag(sum(A))−A;
27  [V D] = eigs(L, 2, 'SA');
28
29  % plot the components of the algebraic connectivity sorted by magnitude
30  subplot(1,3,2); plot(sort(V(:,2)), '.−');
31  xlabel('$\tilde v_2$', 'Interpreter','latex','FontSize',28);
32
33  % partition the matrix accordingly and spot the communities
34  [ignore p] = sort(V(:,2));
35  subplot(1,3,3); spy(A(p,p));
36  xlabel('$\tilde A$', 'Interpreter','latex','FontSize',28);
```

6.5 Appendix: Control design for clock synchronization

In this section we consider an idealized network of heterogeneous clocks and design a control strategy to ensure they achieve synchronization.

Consider n simplified clocks modeled as discrete-time integrators: $x_i(k + 1) = x_i(k) + d_i$. The initial value $x_i(0)$ is called the *initial offset* and d_i is called the *clock speed* (or skew); see Figure 6.3. Assume that we can control each clock according to

$$x(k + 1) = x(k) + d + u(k). \tag{6.6}$$

Define the average clock speed by $d_{\mathrm{ave}} = \mathrm{average}(d) = \mathbb{1}_n^{\mathsf{T}} d/n$.
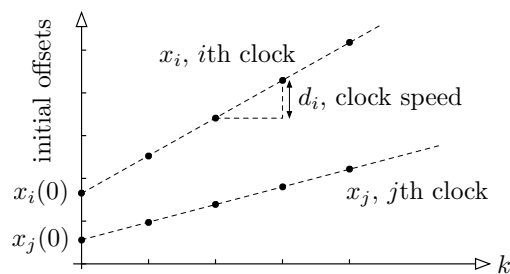
Figure 6.3: Two clocks with different initial offset $x_i(0) \neq x_j(0)$ and speeds $d_i \neq d_j$.

The *clock synchronization problem* is to design a control law u such that, for all clocks for all j and j,

$$\lim_{k \to \infty} x_i(k) - x_j(k) = 0.$$

Proportional/averaging control Suppose the clocks are interconnected by an connected undirected graph so that that each node i can measure the errors $(x_j(k) - x_i(k))$ for some neighbors j. For each edge $\{i, j\}$, let $\kappa_{ij} = \kappa_{ji} > 0$ be a control gain (and set $\kappa_{pq} = 0$ whenever $\{p, q\}$ is not an edge), and select the *proportional/averaging control law*

$$x_i(k + 1) = x_i(k) + d_i + \sum_{j=1}^{n} \kappa_{ij}(x_j(k) - x_i(k)).$$

To analyse this control design, we proceed as follows. First, if $L = L^\mathsf{T}$ denotes the Laplacian matrix defined by these control gains, then the control is $u(k) = -Lx(k)$ and the closed-loop system is

$$x(k + 1) = (I_n - L)x(k) + d.$$

For $\max_{i \in \{1,\dots,n\}} \sum_{j=1}^{n} \kappa_{ij} < 1$, the matrix $I_n - L$ is non-negative and therefore row-stochastic.

Note: we now see that the closed-loop system is an averaging system with a forcing term; this is the reason we call this control action proportional/averaging.

Second, we define $y(k) = x(k) - kd_{\mathrm{ave}}\mathbb{1}_n$. One can show that $y(k+1) = (I_n - L)y(k) + (d - d_{\mathrm{ave}}\mathbb{1}_n)$ and that this system is precisely an affine averaging system as studied in Exercise E2.11. According to Exercise E2.11(iii), we know that, generically, $y(k) \to y_{\mathrm{final}} \notin \mathrm{span}\{\mathbb{1}_n\}$ so that

$$\lim_{k \to \infty} x_i(k) - x_j(k) = \lim_{k \to \infty} y_i(k) - y_j(k) \neq 0.$$

In other words, proportional control keeps the errors bounded (they would naturally diverge without it), but does not achieve vanishing errors and therefore does not solve the clocks synchronization problem.

Proportional/averaging and integral control We now introduce a so-called integrator state w_i at each node, pick an *integral control gain* γ, and design the *proportional/averaging integral control* as

$$u(k) = -Lx(k) - w(k),$$
$$w(k + 1) = w(k) + \gamma Lx(k),$$

so that the closed-loop system dynamics is

$$\begin{aligned} x(k + 1) &= (I_n - L)x(k) - w(k) + d, \\ w(k + 1) &= w(k) + \gamma Lx(k). \end{aligned} \tag{6.7}$$

The rationale for integral control is that, when in steady state with $w(k + 1) = w(k)$, the integral equation in (6.7) enforces $\mathbb{0}_n = Lx(k)$. Hence, if the closed loop (6.7) admits a steady state, then necessarily all clocks must be synchronized. It is natural to assume a zero initial state for the initial integral state $w(0) = \mathbb{0}_n$.

Lemma 6.7 (Asymptotic clock synchronization). *Consider n clocks (6.6) with heterogeneous initial offsets $x_i(0)$, speeds d_i, and average speed $d_{\mathrm{ave}} = \mathrm{average}(d)$. Assume the undirected communication graph among them is connected. Select proportional/averaging gains κ_{ij} for all edges $\{i, j\}$ and an integral control gain γ satisfying*

$$\max_{i \in \{1,\dots,n\}} \sum_{j=1}^{n} \kappa_{ij} < 1, \quad \text{and} \quad 0 < \gamma < 1. \tag{6.8}$$

Then the proportional/averaging integral control ensures that, in the closed loop, the clocks synchronize and

$$\lim_{k\to\infty} \big(x(k) - (d_{\text{ave}}k + x_{\text{ave}}(0))\mathbb{1}_n\big) = \mathbb{0}_n.$$

In other words, the clocks asymptotically synchronizes and their time grows linearly with a speed equal to the average clock speed.

Proof. We start by studying the evolution of the affine dynamical system (6.7) using the modal decomposition as illustrated in Section 2.1. Being a symmetric Laplacian matrix, L has real eigenvalues $0 = \lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$ with corresponding orthonormal eigenvectors $v_1 = \mathbb{1}_n/\sqrt{n}, v_2, \ldots, v_n$. By left-multiplying the closed-loop system dynamics (6.7) by v_α^{T}, $\alpha \in \{1, \ldots, n\}$, we obtain the following n decoupled 2-dimensional systems:

$$\begin{bmatrix} x_\alpha(k+1) \\ w_\alpha(k+1) \end{bmatrix} = \begin{bmatrix} 1 - \lambda_\alpha & -1 \\ \gamma\lambda_\alpha & 1 \end{bmatrix} \begin{bmatrix} x_\alpha(k) \\ w_\alpha(k) \end{bmatrix} + \begin{bmatrix} d_\alpha \\ 0 \end{bmatrix}, \quad \alpha \in \{1, \ldots, n\}, \tag{6.9}$$

where $x_\alpha(k) = v_\alpha^{\mathsf{T}}x(k)$, $w_\alpha(k) = v_\alpha^{\mathsf{T}}w(k)$, and $d_\alpha = v_\alpha^{\mathsf{T}}d$. From this decomposition, the full state can be reconstructed by

$$x(k) = \sum_{\alpha=1}^{n} x_\alpha(k)v_\alpha = x_{\text{ave}}(k)\mathbb{1}_n + \sum_{\alpha=2}^{n} x_\alpha(k)v_\alpha,$$

$$w(k) = \sum_{\alpha=1}^{n} w_\alpha(k)v_\alpha = w_{\text{ave}}(k)\mathbb{1}_n + \sum_{\alpha=2}^{n} w_\alpha(k)v_\alpha.$$

where $x_{\text{ave}}(k) = \text{average}(x(k))$ and $w_{\text{ave}}(k) = \text{average}(w(k))$.

For $\alpha = 1$, after a simple rescaling, equation (6.9) reads

$$\begin{bmatrix} x_{\text{ave}}(k+1) \\ w_{\text{ave}}(k+1) \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_{\text{ave}}(k) \\ w_{\text{ave}}(k) \end{bmatrix} + \begin{bmatrix} d_{\text{ave}} \\ 0 \end{bmatrix}.$$

Because $w(0) = \mathbb{0}_n$, we compute $w(k) = \mathbb{0}_n$ and $x_{\text{ave}}(k) = d_{\text{ave}}k + x_{\text{ave}}(0)$.

It now suffices to show that the solutions to the $n-1$ equations (6.9), for $\alpha \in \{2, \ldots, n\}$, satisfy $\lim_{k\to\infty} x_\alpha(k) = 0$. Simple calculations show that the only equilibrium solutions to the $n-1$ equations (6.9), for $\alpha \in \{2, \ldots, n\}$, are $x_\alpha^* = 0$ and $w_\alpha^* = -d_\alpha$. Hence, it suffices to show that all eigenvalues of the $n-1$ matrices of dimension 2×2 have magnitude strictly less than 1. For $\alpha \in \{2, \ldots, n\}$, the $n-1$ characteristic equations are

$$(z-1)^2 + \lambda_\alpha(z - 1 + \gamma) = 0.$$

We claim that these polynomials have both roots strictly inside the unit circle if and only if, for all $\alpha \in \{2, \ldots, n\}$,

$$0 < \gamma < 1, \quad \text{and} \quad 0 < \lambda_\alpha < 4/(2\gamma). \tag{6.10}$$

Recall from the proof of, and the discussion following, Lemma 6.5 that

$$\lambda_i \le \lambda_n < 2 \max_{i\in\{1,\ldots,n\}} \sum_{j=1}^{n} \kappa_{ij}.$$

But by the assumption (6.8) we know $\max_{i \in \{1,\dots,n\}} \sum_{j=1}^{n} \kappa_{ij} < 1$, hence $\lambda_n < 2 \times 1 < 4/(2 - \gamma)$ for all $0 < \gamma < 1$. Hence, the inequalities (6.10) are satisfied.

To verify that the inequalities (6.10) imply that all roots have magnitude less than 1, we use the so-called bilinear transform method. This method is based on the equivalence between the following two properties: the original polynomial has roots strictly inside the unit disk and the transformed polynomial has roots with strictly negative real part. We proceed as follows: we take $z = (1 + s)/(1 - s)$ and substitute it into the polynomial $(z - 1)^2 + \lambda_\alpha(z - 1 + \gamma)$ so that, removing the denominator, we obtain the polynomial $(4 - 2\lambda_\alpha + \lambda_\alpha\gamma)s^2 - \lambda_\alpha(2\gamma - 2)s + \lambda_\alpha\gamma$. By the Routh-Hurwitz stability criterion, this polynomial has roots with negative real part if and only if all three coefficients are strictly positive or strictly negative. Some elementary calculations show that all three coefficients may never be negative and that all three coefficients are positive if and only if the inequalities (6.10) hold. ∎

6.6 Historical notes and further reading

Standard books on algebraic graph theory with extensive characterizations of adjacency and Laplacian matrices include (Biggs, 1994) and (Godsil and Royle, 2001). Two surveys about Laplacian matrices are (Mohar, 1991; Merris, 1994). Of particular interest for further reading is Kirchhoff's Matrix Tree Theorem.

The rank of the Laplacian, as characterized in Theorem 6.6, was studied as early as in (Fife, 1972; Foster and Jacquez, 1975). A mathematical approach is given in (Agaev and Chebotarev, 2000) which features the first necessary and sufficient characterization. We also refer to the more recent (Lin et al., 2005; Ren and Beard, 2005) for the specific case of $\mathrm{rank}(L) = n - 1$.

The generalized inverse of the Laplacian matrix appears in some applications and is studied in Gutman and Xiao (2004).

The ground-breaking work in (Fiedler, 1973) established the use of the eigenvalues of the Laplacian matrix for example as a way to quantify graph connectivity and to perform clustering, as illustrated in Section 6.4. For surveys on community detection we refer to (Porter et al., 2009; Fortunato, 2010).

The example on clock synchronization via proportional/averaging and integral control in Section 6.5 is taken from (Carli et al., 2008a). More realistic settings are studied in (Schenato and Fiorentin, 2011; Carli and Zampieri, 2014). Surveys include (Sundararaman et al., 2005; Sivrikaya and Yener, 2004; Simeone et al., 2008).

Complex-valued graphs, adjacency and Laplacian matrices are studied in (Reff, 2012); see also (Lin et al., 2013; Dong and Qiu, 2014) for some related applications.

6.7   Exercises

E6.1   **The spectra of Laplacian and row-stochastic adjacency matrices.** Consider a row-stochastic matrix $A \in \mathbb{R}^{n \times n}$. Let L be the Laplacian matrix of the digraph associated to A. Compute the spectrum of L as a function of the spectrum $\mathrm{spec}(A)$ of A.

E6.2   **Basic properties of a symmetric Laplacian matrix.** Let G be a weighted undirected graph with symmetric Laplacian matrix $L \in \mathbb{R}^{n \times n}$.

    (i) Prove, without relying on the Geršgorin Disks Theorem 2.8, that L is symmetric positive semidefinite. (Note that the proof of Lemma 6.5 relies on Geršgorin Disks Theorem 2.8).

Assume G is connected. Let λ_2 be the smallest non-zero eigenvalue of L with eigenvector v (unique up to rescaling). Show that

    (ii) $v \perp \mathbb{1}_n$ and $v^\mathsf{T} L v = \lambda_2 \|v\|_2^2$, and

    (iii) for any $x \in \mathbb{R}^n$,

$$x^\mathsf{T} L x \geq \lambda_2 \left\| x - \frac{1}{n}(\mathbb{1}_n^\mathsf{T} x)\mathbb{1}_n \right\|_2^2,$$

        with equality only when x is parallel to v.

E6.3   **Upper and lower bound on largest Laplacian eigenvalue.** Let G be an undirected graph with symmetric Laplacian matrix $L = L^\mathsf{T} \in \mathbb{R}^{n \times n}$, Laplacian eigenvalues $0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$, and maximum degree $d_{\max} = \max_{i \in \{1, \dots, n\}} d_i$. Show that the maximum eigenvalue λ_n satisfies:

$$d_{\max} \leq \lambda_n \leq 2d_{\max}.$$

Hint: For the upper bound review the proof of Lemma 6.5.

E6.4   **Examples in spectral graph theory.** Let $G^\star$ be a graph with 8 nodes and with Laplacian matrix $L(G^\star) \in \mathbb{R}^{8 \times 8}$. For $\mathrm{i} = \sqrt{-1}$, assume the spectrum of $L(G^\star)$ is

$$\mathrm{spec}(L(G^\star)) = \{0, 0, 0.5104, 1.6301, 2, 2.2045 - 1.0038\mathrm{i}, 2.2045 + 1.0038\mathrm{i}, 2.8646\}.$$

Consider the graphs G_1, G_2, and G_3 shown below. Argue why the following statements are true:

    (i) G_1 cannot be $G^\star$,

    (ii) G_2 cannot be $G^\star$, and

    (iii) G_3 cannot be $G^\star$.
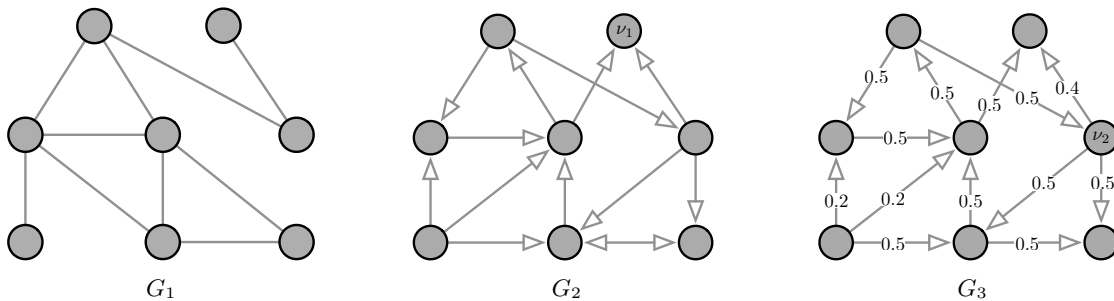
Figure E6.1: Example graphs and digraphs with 8 noses

E6.5   **The Laplacian matrix plus its transpose.** Let G be a weighted digraph with Laplacian matrix L. Prove the following statements are equivalent:

   (i) G is weight-balanced,

   (ii) $L + L^\mathsf{T}$ is positive semidefinite.

   Next, assume G is weight-balanced with adjacency matrix A, show that

   (iii) $L + L^\mathsf{T}$ is the Laplacian matrix of the digraph associated to the symmetric adjacency matrix $A + A^\mathsf{T}$, and

   (iv) $(L + L^\mathsf{T})\mathbb{1}_n = \mathbb{0}_n$,

E6.6   **Scaled Laplacian matrices.** Let $L = L^\mathsf{T} \in \mathbb{R}^{n \times n}$ be the Laplacian matrix of a connected, undirected, and symmetrically weighted graph. Given scalars $d_1, \ldots, d_n$, define the matrices A and B by

$$A := \mathrm{diag}\{d_1, \ldots, d_n\}L \quad \text{and} \quad B := L\,\mathrm{diag}\{d_1, \ldots, d_n\}.$$

   (i) Give necessary and sufficient conditions on $\{d_1, \ldots, d_n\}$ for A to be a Laplacian matrix.

   (ii) Give necessary and sufficient conditions on $\{d_1, \ldots, d_n\}$ for B to be a Laplacian matrix.

   (iii) Give a sufficient condition on $\{d_1, \ldots, d_n\}$ for A and B to be symmetric.

   (iv) Assuming $d_i \neq 0$, $i \in \{1, \ldots, n\}$, do A and B possess a zero eigenvalue? If so, what are the corresponding right and left eigenvectors for A and B?

E6.7   **The disagreement function in a directed graph (Gao et al., 2008).** Recall that the *quadratic form* associated with a symmatric matrix $B \in \mathbb{R}^{n \times n}$ is the function $x \mapsto x^\mathsf{T} B x$. Let G be a weighted digraph G with n nodes and define the *quadratic disagreement function* $\Phi_G : \mathbb{R}^n \to \mathbb{R}$ by

$$\Phi_G(x) = \frac{1}{2} \sum_{i,j=1}^{n} a_{ij}(x_j - x_i)^2.$$

   Show that:

   (i) Φ_G is the quadratic form associated with the symmetric positive-semidefinite matrix

$$P = \frac{1}{2}(D_\mathrm{out} + D_\mathrm{in} - A - A^\mathsf{T}),$$

   (ii) $P = \frac{1}{2}\big(L + L(\mathrm{rev})\big)$, where the Laplacian of the reverse digraph is $L(\mathrm{rev}) = D_\mathrm{in} - A^\mathsf{T}$.

E6.8   **The pseudoinverse Laplacian matrix.** The *Moore-Penrose pseudoinverse* (or simply the *pseudoinverse*) of an $n \times m$ matrix M is the unique $m \times n$ matrix $M^\dagger$ with the following properties:

   (i) $MM^\dagger M = M$,

   (ii) $M^\dagger M M^\dagger = M^\dagger$, and

   (iii) $MM^\dagger$ is symmetric and $M^\dagger M$ is symmetric.

   Now, let L be the Laplacian matrix of a weighted connected undirected graph with n nodes. Let $U \in \mathbb{R}^{n \times n}$ be an orthonormal matrix of eigenvectors of L such that

$$L = U \begin{bmatrix} 0 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{bmatrix} U^\mathsf{T}.$$

   Show that

(i) $L^\dagger = U \begin{bmatrix} 0 & 0 & \dots & 0 \\ 0 & 1/\lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1/\lambda_n \end{bmatrix} U^\mathsf{T},$

(ii) $LL^\dagger = L^\dagger L = I_n - \dfrac{1}{n} \mathbb{1}_n \mathbb{1}_n^\mathsf{T}$, and

(iii) $L^\dagger \mathbb{1}_n = \mathbb{0}_n$.

E6.9   **The regularized Laplacian matrix.** Let L be the Laplacian matrix of a weighted connected undirected graph with n nodes. Given a scalar $\beta \in \mathbb{R}$, define the *regularized Laplacian matrix* $L_{\mathrm{reg},\beta} = L + \frac{\beta}{n} \mathbb{1}_n \mathbb{1}_n^T$. Show that

   (i) $L_{\mathrm{reg},\beta}$ is nonsingular for $\beta \neq 0$,

   (ii) $L_{\mathrm{reg},\beta}$ is positive definite for $\beta > 0$, and

   (iii) the inverse of $L_{\mathrm{reg},\beta}$ satisfies

$$L_{\mathrm{reg},\beta}^{-1} = \left(L + \frac{\beta}{n} \mathbb{1}_n \mathbb{1}_n^T \right)^{-1} = L^\dagger + \frac{1}{\beta\,n} \mathbb{1}_n \mathbb{1}_n^T \,.$$

   Hint: *Make use of the singular value decomposition in Exercise E6.8.*

E6.10   **The Green matrix of a Laplacian matrix.** Assume L is the Laplacian matrix of a weighted connected undirected graph with n nodes. Show that

   (i) the matrix $L + \frac{1}{n} \mathbb{1}_n \mathbb{1}_n^\mathsf{T}$ is positive definite,

   (ii) the so-called *Green matrix*

$$X = \left(L + \frac{1}{n} \mathbb{1}_n \mathbb{1}_n^\mathsf{T} \right)^{-1} - \frac{1}{n} \mathbb{1}_n \mathbb{1}_n^\mathsf{T} \tag{E6.1}$$

   is the unique solution to the system of equations:

$$\begin{cases} LX = I_n - \frac{1}{n} \mathbb{1}_n \mathbb{1}_n^\mathsf{T}, \\ \mathbb{1}_n^\mathsf{T} X = \mathbb{0}_n^\mathsf{T}, \end{cases}$$

   (iii) $X = L^\dagger$, where $L^\dagger$ is defined in Exercise E6.8. In other words, the Green matrix formula (E6.1) is an alternative definition of the pseudoinverse Laplacian matrix.

E6.11   **Monotonicity of Laplacian eigenvalues.** Consider a symmetric Laplacian matrix $L \in \mathbb{R}^{n \times n}$ associated to a weighted and undirected graph $G = \{V, E, A\}$. Assume G is connected and let $\lambda_2(G) > 0$ be its algebraic connectivity, i.e., the second-smallest eigenvalue of L. Show that

   (i) $\lambda_2(G)$ is a monotonically non-decreasing function of each weight a_{ij}, $\{i, j\} \in E$; and

   (ii) $\lambda_2(G)$ is monotonically non-decreasing function in the edge set in the following sense: $\lambda_2(G) \leq \lambda_2(G')$ for any graph $G' = (V, E', A')$ with $E \subset E'$ and $a_{ij} = a'_{ij}$ for all $\{i, j\} \in E$.

   Hint: *Use the disagreement function.*

E6.12   **Gaussian elimination and Laplacian matrices.** Consider an undirected and connected graph and its associated Laplacian matrix $L \in \mathbb{R}^{n \times n}$. Consider the associated linear Laplacian equation $y = Lx$, where $x \in \mathbb{R}^n$ is unknown and $y \in \mathbb{R}^n$ is a given vector. Verify that an elimination of x_n from the last row of this

equation yields the following reduced set of equations:

$$
\begin{bmatrix} y_1 \\ \vdots \\ y_{n-1} \end{bmatrix} + \underbrace{\begin{bmatrix} -L_{1n}/L_{nn} \\ \vdots \\ -L_{n-1,n}/L_{nn} \end{bmatrix}}_{=A} y_n = \underbrace{\begin{bmatrix} \ddots & \vdots & \cdot^{\cdot^{\cdot}} \\ \dots & L_{ij} - \frac{L_{in} \cdot L_{jn}}{L_{nn}} & \dots \\ \cdot_{\cdot_{\cdot}} & \vdots & \ddots \end{bmatrix}}_{=L_{\mathrm{red}}} \begin{bmatrix} x_1 \\ \vdots \\ x_{n-1} \end{bmatrix},
$$

where the (i,j)-element of L_{red} is given by $L_{ij} - L_{in} \cdot L_{jn}/L_{nn}$. Show that the matrices $A \in \mathbb{R}^{n-1 \times 1}$ and $L \in \mathbb{R}^{(n-1) \times (n-1)}$ obtained after Gaussian elimination have the following properties:

  (i) A is non-negative and column-stochastic matrix with at least one strictly positive element; and

  (ii) L_{red} is a symmetric and irreducible Laplacian matrix.

Hint: *To show the irreducibility of L_{red}, verify the following property regarding the* fill-in *of the matrix L_{red}: The graph associated to the Laplacian L_{red} has an edge between nodes i and j if and only if (i) either $\{i,j\}$ was an edge in the original graph associated to L, (ii) or $\{i,n\}$ and $\{j,n\}$ were edges in the original graph associated to L.*

E6.13 **Thomson's principle and energy routing.** Consider a connected and undirected resistive electrical network with n nodes, with external nodal current injections $c \in \mathbb{R}^n$ satisfying the balance condition $\mathbb{1}_n^\top c = 0$, and with resistances $R_{ij} > 0$ for every undirected edge $\{i,j\} \in E$. For simplicity, we set $R_{ij} = \infty$ if there is no edge connecting i and j. As shown earlier in this chapter, Kirchhoff's and Ohm's laws lead to the network equations

$$
c_{\text{injected at } i} = \sum_{j \in \mathcal{N}(i)} c_{j \to i} = \sum_{j \in \mathcal{N}(i)}^n \frac{1}{R_{ij}} (v_i - v_j),
$$

where v_i is the potential at node i and $c_{j \to i} = 1/R_{ij} \cdot (v_i - v_j)$ is the current flow from node i to node j. Consider now a more general set of current flows $f_{i \to j}$ (for all $i, j \in \mathbb{R}^n$) "routing energy through the network" and compatible with the following basic assumptions:

  (i) Skew-symmetry: $f_{i \to j} = -f_{j \to i}$ for all $i, j \in \mathbb{R}^n$;
  (ii) Consistency: $f_{i \to j} = 0$ if $\{i,j\} \notin E$;
  (iii) Conservation: $c_{\text{injected at } i} = \sum_{j \in \mathcal{N}(i)} f_{j \to i}$ for all $i \in \mathbb{R}^n$.

Show that among all possible current flows $f_{i \to j}$, the physical current flow $f_{i \to j} = c_{i \to j} = 1/R_{ij} \cdot (v_j - v_i)$ uniquely minimizes the energy dissipation:

$$
\begin{aligned}
\underset{f_{i \to j}, \, i,j \in \{1,\dots,n\}}{\text{minimize}} \quad & J = \frac{1}{2} \sum_{i,j=1}^n R_{ij} f_{i \to j}^2 \\
\text{subject to} \quad & f_{i \to j} = -f_{j \to i} && \text{for all } i,j \in \mathbb{R}^n , \\
& f_{i \to j} = 0 && \text{for all } \{i,j\} \notin E , \\
& c_{\text{injected at } i} = \sum_{j \in \mathcal{N}(i)} f_{j \to i} && \text{for all } i \in \mathbb{R}^n .
\end{aligned}
$$

Hint: *The solution requires knowledge of the Karush-Kuhn-Tucker (KKT) conditions for optimality; this is a classic topic in nonlinear constrained optimization discussed in numerous textbooks, e.g., in (Luenberger and Ye, 2008).*

E6.14 **Linear spring networks with loads.** Consider the two (connected) spring networks with n moving masses in figure. For the right network, assume one of the masses is connected with a single stationary object with a

spring. Refer to the left spring network as *free* and to the right network as *grounded*. Let F_{load} be a load force applied to the n moving masses.

For the left network, let $L_{\text{free},n}$ be the $n \times n$ Laplacian matrix describing the free spring network among the n moving masses, as defined in Section 6.1.1. For the right network, let $L_{\text{free},n\,+\,1}$ be the $(n+1) \times (n+1)$ Laplacian matrix for the spring network among the n masses and the stationary object. Let L_{grounded} be the $n \times n$ *grounded Laplacian* of the n masses constructed by removing the row and column of $L_{\text{free},n\,+\,1}$ corresponding to the stationary object.

For the free spring network subject to F_{load},

   (i) do equilibrium displacements exist for arbitrary loads?

  (ii) if the load force F_{load} is *balanced* in the sense that $\mathbb{1}_n^{\mathsf{T}} F_{\text{load}} = 0$, is the resulting equilibrium displacement unique?

 (iii) compute the equilibrium displacement if unique, or the set of equilibrium displacements otherwise, assuming a balanced force profile is applied.

For the grounded spring network,

 (iv) derive an expression relating L_{grounded} to $L_{\text{free},n}$,

  (v) show that L_{grounded} is invertible,

 (vi) compute the displacement for the "grounded" spring network for arbitrary load forces.

We refer to Exercise E9.10 for a comprehensive treatment of grounded Laplacian matrices.

E6.15 **From algebraic to vertex connectivity.** Consider an unweighted undirected graph $G = (V, E)$ with second smallest eigenvalue $\lambda_2(G)$. Given a subset of nodes $S \subseteq V$, we define a graph $G' = (V', E')$ by *deleting* the nodes in S from G as follows: we let $V' = V \setminus S$ and E' contain all the edges in E except for those connected to a node in S. The *vertex connectivity* $\kappa(G)$ of G is defined by

$$\kappa(G) = \begin{cases} 0, & \text{if } G \text{ is disconnected,} \\ \text{minimum number of nodes whose deletion disconnects } G, & \text{otherwise.} \end{cases}$$

Show that

   (i) $0 \leq \lambda_2(G) \leq \lambda_2(G') + |S|$, where $|S|$ is the cardinality of S,

  (ii) $\lambda_2(G) \leq \kappa(G)$.

Hint: *Let $z \in \mathbb{R}^{|V'|}$, $\|z\|_2 = 1$, denote the eigenvector of the Laplacian $L(G')$ associated with $\lambda_2(G')$. You may find it useful to define $q \in \mathbb{R}^{|V|}$ such that $q_i = z_i$ for every $i \in V'$ and $q_i = 0$ for every $i \in S$.*

E6.16 **Maximum power dissipation.** As in Subsection 6.1.2, consider an electrical network composed by three voltage sources (v_1, v_2, v_3) connected by three resistors (each with unit resistance in an undirected ring topology. Let L be the Laplacian matrix of conductances. Recall that the total power dissipated by the circuit is

$$\mathrm{P}_{\text{dissipated}} = v^{\mathsf{T}} L v.$$

What is the maximum dissipated power if the voltages v are such that $\|v\|_2 = 1$?
Hint: *Recall the notion of induced 2-norm.*

E6.17 **Distributed averaging-based PI control.** Consider a set of n controllable agents governed by the second-order dynamics

$$\dot{x}_i = y_i, \tag{E6.2a}$$

$$\dot{y}_i = u_i + \eta_i, \tag{E6.2b}$$

where $i \in \{1, \ldots, n\}$ is the index set, $u_i \in \mathbb{R}$ is a control input to agent i, and $\eta_i \in \mathbb{R}$ is an unknown disturbance affecting agent i. Given an undirected, connected, and weighted graph $G = (V, E, A)$ with node set $V = \{1, \ldots, n\}$, edge set $E \subset V \times V$, and adjacency matrix $A = A^\mathsf{T} \in \mathbb{R}^{n \times n}$, we assume each agent can measure its velocity $y_i \in \mathbb{R}$ as well as the relative position $x_i - x_j$ for each neighbor $\{i, j\} \in E$. Based on these measurements, consider now the distributed averaging-based proportional-integral (PI) controller

$$u_i = -\sum_{j=1}^{n} a_{ij}(x_i - x_j) - y_i - q_i, \tag{E6.3a}$$

$$\dot{q}_i = y_i - \sum_{j=1}^{n} a_{ij}(q_i - q_j), \tag{E6.3b}$$

where $q_i \in \mathbb{R}$ is a dynamic control state for each agent $i \in \{1, \ldots, n\}$. Your tasks are as follows:

(i) show that the average state $\frac{1}{n} \sum_{i=1}^{n} x_i(t)$ is bounded for all $t \geq 0$,

(ii) characterize the set of equilibria $(x^\star, y^\star, q^\star)$ of the closed-loop system (E6.2)-(E6.3), and

(iii) show that all trajectories converge to these closed-loop equilibria.

<div align="right">

Chapter **7**

</div>

Continuous-time Averaging Systems

In this chapter we consider averaging algorithms in which the variables evolve in continuous time, instead of discrete time. In other words, we consider a certain class of differential equations and show when their asymptotic behavior is the emergence of consensus.

7.1 Example systems

We present here some simple examples of continuous-time averaging systems.

7.1.1 Example #1: Continuous-time opinion dynamics

This example is a continuous-time version of the discrete-time averaging models we have studied in details, starting from Section 1.1. We start by considering the opinion change of individual i from time k to time $k+1$ in the discrete-time averaging model (1.2):

$$\Delta x_i(k) = x_i(k+1) - x_i(k) = \Big(\sum_{j=1}^{n} a_{ij} x_j(k) \Big) - x_i(k) = \sum_{j=1}^{n} a_{ij}(x_j(k) - x_i(k)),$$

where the last step follows from the equality $\sum_{j=1}^{n} a_{ij} = 1$. Using the equality (6.2), we can obtain

$$\Delta x(k) = -Lx(k),$$

where $L = I_n - A$ is the Laplacian of the adjacency matrix A. We now assume that the opinion change occurs infinitesimally slowly. Specifically, we assume there exists a time period $\Delta t \ll 1$ such that the time indexes k and $k+1$ correspond to real times t and $t + \Delta t$ respectively, and that $L = \overline{L}\Delta t$, for an appropriate Laplacian matrix L. In summary, this assumption implies

$$\frac{\Delta x(t)}{\Delta t} = -\overline{L}x(t),$$

and, taking the limit as $\Delta t \to 0^+$, we obtain the Abelson's continuous-time opinion dynamics model:

$$\dot{x}(t) = -\overline{L}x(t).$$

We refer to this equation as to the *Laplacian flow*. In this model, the edge weights $\overline{a}_{ij}$ of the Laplacian $\overline{L}$ are *contact rates* between the individuals.

   Note: As this dynamics models continuous-time averaging, we expect to see consensus value emerge for certain classes of digraphs.

7.1.2   Example #2: Flocking behavior for a group of animals

Next, we are interested in swarming and flocking behavior that many animal species exhibit from decentralized interactions, e.g., see Figure 7.1. To model this behavior we consider a simple "alignment rule" for each
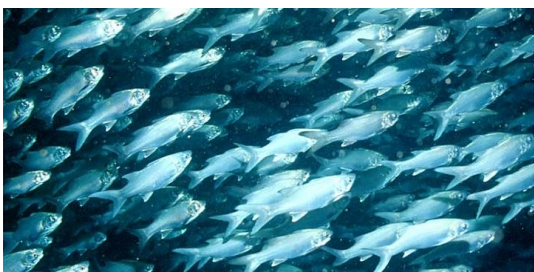
(a) A swarm of pacific threadfins (*Polydactylus sexfilis*). Public domain image from the U.S. National Oceanic and Atmospheric Administration.

(b) A flock of snow geese (*Chen caerulescens*). Public domain image from the U.S. Fish and Wildlife Service.

Figure 7.1: Examples of animal flocking behaviors

agent to steer towards the average heading of its neighbors; see Figure 7.2. This alignment rule amounts to
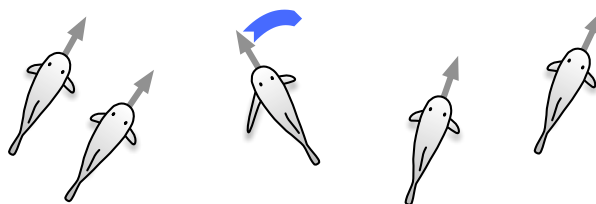
Figure 7.2: Alignment rule: the center fish rotates clockwise to align itself with the average heading of its neighbors.

a "spring-like" attractive force, described as follows:

$$\dot{\theta}_i = \begin{cases} (\theta_j - \theta_i), & \text{if } i\text{th agent has one neighbor} \\ \frac{1}{2}(\theta_{j_1} - \theta_i) + \frac{1}{2}(\theta_{j_2} - \theta_i), & \text{if } i\text{th agent has two neighbors} \\ \frac{1}{m}(\theta_{j_1} - \theta_i) + \cdots + \frac{1}{m}(\theta_{j_m} - \theta_i), & \text{if } i\text{th agent has } m \text{ neighbors} \end{cases}$$
$$= \text{average}\big(\{\theta_j, \text{ for all neighbors } j\}\big) - \theta_i.$$

This interaction law can be written as

$$\dot{\theta} = -L\theta$$

where L is the Laplacian of an appropriate weighted digraph G: each bird is a node and each directed edge (i,j) has weight $1/d_{\text{out}}(i)$. Here it is useful to recall the interpretation of $-(Lx)_i$ as a force perceived by node i in a network of springs.

Note: it is mathematically ill-posed to compute averages on a circle, but we will not worry about this matter in this chapter.

Note: this incomplete model does not concern itself with positions. In other words, we do not discuss collision avoidance and formation/cohesion maintenance. Moreover, note that the graph G should be really state dependent. For example, we may assume that two birds see each other and interact if and only if their pairwise Euclidean distance is below a certain threshold.

7.1.3 Example #3: A simple RC circuit
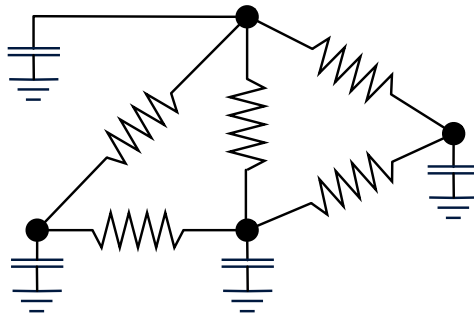
Finally, we consider an electrical network with only pure resistors and with pure capacitors connecting each node to ground. From the previous chapter, we know the vector of injected currents c_{injected} and the vector of voltages at the nodes v satisfy

$$c_{\text{injected}} = L\,v,$$

where L is the Laplacian for the graph with coefficients $a_{ij} = 1/r_{ij}$. Additionally, assuming C_i is the capacitance at node i, and keeping proper track of the current into each capacitor, we have

$$C_i \frac{d}{dt}v_i = -c_{\text{injected at }i}$$

so that, with the shorthand $C = \operatorname{diag}(C_1, \ldots, C_n)$,

$$\frac{d}{dt}v = -C^{-1}L\,v. \tag{7.1}$$

Note: $C^{-1}L$ is again a Laplacian matrix (for a directed weighted graph).

Note: it is physically intuitive that after some transient all nodes will have the same potential. This intuition will be proved later in the chapter.

7.1.4 Example #4: Discretization of partial differential equations

The name Laplacian matrix is inherited from the *Laplacian operator* in the diffusion partial differential equation (PDEs) named after the French mathematician Pierre-Simon Laplace.

(a) Spatial domain

(b) Discretization through a mesh grid
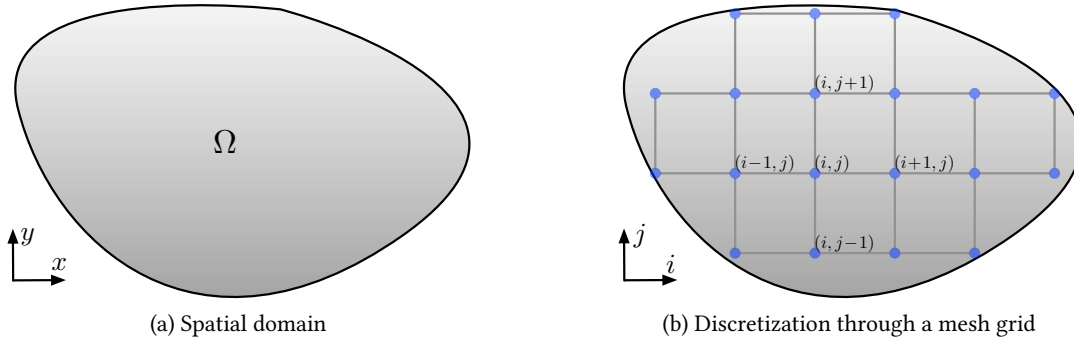
Consider a closed bounded spatial domain $\Omega \subset \mathbb{R}^2$ and a spatio-temporal function $u(t, x, y)$ denoting the temperature at a point $(x, y) \in \Omega$ at time $t \in \mathbb{R}_{\geq 0}$. The evolution of the temperature $u(t, x, y)$ in time and space is governed by the *heat equation*

$$\frac{\partial u}{\partial t} + c\, \Delta u = 0 \,, \tag{7.2}$$

where $c > 0$ is the *thermal diffusivity* (which we assume constant) and the *Laplacian differential operator* is

$$\Delta u(t, x, y) = \frac{\partial^2 u}{\partial x^2}(t, x, y) + \frac{\partial^2 u}{\partial y^2}(t, x, y).$$

To approximately solve this PDE, we introduce a finite-difference approximation of (7.2). First, we discretize the spatial domain Ω through a mesh grid (specifically, a Cartesian grid) with discrete coordinates indexed by (i, j) and where neighboring grid points are a distance $h > 0$ apart. Second, we approximate the Laplacian operator via the finite-difference approximation:

$$\Delta u(t, x_i, y_j) \approx -\frac{4u(t, x_i, y_j) - u(t, x_{i-1}, y_j) - u(t, x_{i+1}, y_j) - u(t, x_i, y_{j-1}) - u(t, x_i, y_{j+1})}{h^2}.$$

(This is the correct expansion for an interior point; similar approximations can be written for boundary points, assuming the boundary conditions are free.) Now, the key observation is that the finite-difference approximation renders the heat equation to a Laplacian flow. Specifically, if u_{discrete} denotes the vector of values of u at the nodes, we have

$$\frac{d}{dt} u_{\text{discrete}} + L u_{\text{discrete}} = 0 \,,$$

where L is the Laplacian matrix of the mesh grid with weights c/h^2.

Another standard PDE involving the Laplacian operator is the *wave equation*

$$\frac{\partial^2 u}{\partial t^2} + c^2\, \Delta u = 0 \,, \tag{7.3}$$

modeling the displacement $u(t, x, y)$ of an elastic surface on Ω with *wave propagation speed* $c > 0$. In this case, a finite-difference approximation gives rise to the *second-order Laplacian flow*

$$\frac{d^2}{dt^2} u_{\text{discrete}} + L u_{\text{discrete}} = 0.$$

We study both first and second-order Laplacian flows in this chapter.

7.2 Continuous-time linear systems and their convergence properties

In Section 2.1 we presented discrete-time linear systems and their convergence properties; here we present their continuous-time analogous.

A *continuous-time linear system* is

$$\dot{x}(t) = Ax(t). \tag{7.4}$$

Its *solution* $t \mapsto x(t), t \in \mathbb{R}_{\geq 0}$ from an initial confition $x(0)$ satisfies $x(t) = \mathrm{e}^{At} x(0)$, where the *matrix exponential* of a square matrix A is defined by

$$\mathrm{e}^{A} = \sum_{k=0}^{\infty} \frac{1}{k!} A^{k}.$$

The matrix exponential is a remarkable operation with numerous properties; we ask the reader to review a few basic ones in Exercise E7.1. A matrix $A \in \mathbb{R}^{n \times n}$ is

  (i) *continuous-time semi-convergent* if $\lim_{t \to +\infty} \mathrm{e}^{At}$ exists, and

  (ii) *continuous-time convergent* (Hurwitz) if $\lim_{t \to +\infty} \mathrm{e}^{At} = \mathbb{0}_{n \times n}$.

The *spectral abscissa* of a square matrix A is the maximum of the real parts of the eigenvalues of A, that is,

$$\mu(A) = \max\{\Re(\lambda) \mid \lambda \in \mathrm{spec}(A)\}.$$

Theorem 7.1 (Convergence and spectral abscissa). *For a square matrix A, the following statements hold:*

  (i) A is continuous-time convergent (Hurwitz) $\iff \mu(A) < 0$,

  (ii) A is continuous-time semi-convergent $\iff \mu(A) \leq 0$, no eigenvalue has zero real part other than possibly the number 0, and if 0 is an eigenvalue, then it is semisimple.

We leave the proof of this theorem to the reader and mention that most required steps are similar to the dicussion in Section 2.1 and are discussed later in this chapter.

7.3 The Laplacian flow

Let G be a weighted directed graph with n nodes and Laplacian matrix L. The *Laplacian flow* on $\mathbb{R}^n$ is the dynamical system

$$\dot{x} = -Lx, \tag{7.5}$$

or, equivalently in components,

$$\dot{x}_i = \sum_{j=1}^{n} a_{ij}(x_j - x_i) = \sum_{j \in \mathcal{N}^{\mathrm{out}}(i)} a_{ij}(x_j - x_i).$$

7.3.1 Matrix exponential of a Laplacian matrix

Before analyzing the Laplacian flow, we provide some results on the matrix exponential of (minus) a Laplacian matrix. We show how such an exponential matrix is row-stochastic and has properties analogous to those for adjacency matrices studied in Section 4.2.

Theorem 7.2 (The matrix exponential of a Laplacian matrix). *Let L be an $n \times n$ Laplacian matrix with associated digraph G and with maximum diagonal entry $\ell_{\max} = \max\{\ell_{11}, \ldots, \ell_{nn}\}$. Then*

(i) $\exp(-L) \geq \mathrm{e}^{-\ell_{\max}} I_n \geq 0,$ *for any digraph G,*

(ii) $\exp(-L)\mathbb{1}_n = \mathbb{1}_n,$ *for any digraph G,*

(iii) $\mathbb{1}_n^{\mathsf{T}} \exp(-L) = \mathbb{1}_n^{\mathsf{T}},$ *for a weight-balanced G (i.e., $\mathbb{1}_n^{\mathsf{T}} L = \mathbb{0}_n^{\mathsf{T}}$),*

(iv) $\exp(-L)\, \mathrm{e}_j > 0,$ *for a digraph G whose j-th node is globally reachable, and*

(v) $\exp(-L) > 0,$ *for a strongly connected digraph G (i.e., for an irreducible L).*

Note that properties (i) and (ii) together imply that $\exp(-L)$ is row-stochastic.

Proof. From the equality $L\mathbb{1}_n = \mathbb{0}_n$ and the definition of matrix exponential, we compute

$$\exp(-L)\mathbb{1}_n = \Big(I_n + \sum_{k=1}^{\infty} \frac{(-1)^k}{k!} L^k\Big)\mathbb{1}_n = \mathbb{1}_n.$$

Similarly, if $\mathbb{1}_n^{\mathsf{T}} L = \mathbb{0}_n^{\mathsf{T}}$, we compute

$$\mathbb{1}_n^{\mathsf{T}} \exp(-L) = \mathbb{1}_n^{\mathsf{T}}\Big(I_n + \sum_{k=1}^{\infty} \frac{(-1)^k}{k!} L^k\Big) = \mathbb{1}_n^{\mathsf{T}}.$$

These calculations establish statements (ii) and (iii).

Next, we define a non-negative matrix A_L by

$$A_L = -L + \ell_{\max} I_n \qquad \Longleftrightarrow \qquad -L = -\ell_{\max} I_n + A_L.$$

Because $A_L I_n = I_n A_L$, we know

$$\exp(-L) = \exp(-\ell_{\max} I_n) \exp(A_L) = \mathrm{e}^{-\ell_{\max}} \exp(A_L).$$

Here we used the following properties of the matrix exponential operation: $\exp(A + B) = \exp(A)\exp(B)$ if $AB = BA$ and $\exp(a I_n) = \mathrm{e}^a I_n$. Next, because $A_L \geq 0$, we know that $\exp(A_L) = \sum_{k=0}^{\infty} A_L^k / k!$ is lower bounded by the first $n-1$ terms of the series so that

$$\exp(-L) = \mathrm{e}^{-\ell_{\max}} \exp(A_L) \geq \mathrm{e}^{-\ell_{\max}} \sum_{k=0}^{n-1} \frac{1}{k!} A_L^k. \tag{7.6}$$

Next, we derive two useful lower bounds on $\exp(-L)$ based on the inequality (7.6). First, by keeping just the first term, we establish statement (i):

$$\exp(-L) \geq \mathrm{e}^{-\ell_{\max}} I_n \geq 0.$$

Second, we lower bound the coefficients $1/k!$ and write:

$$\exp(-L) \geq e^{-\ell_{max}} \sum_{k=0}^{n-1} \frac{1}{k!} A^k \geq \frac{e^{-\ell_{max}}}{(n-1)!} \sum_{k=0}^{n-1} A^k.$$

Notice now that the digraph G associated to L is the same as that associated to A_L (we do not need to worry about self-loops here). Hence, if node j is globally reachable in G, then Lemma 4.4 implies that the jth column of $\sum_{k=0}^{n-1} A_L^k$ is positive and, by inequality (7.6), also the jth column of $\exp(-L)$ is positive. This statement establishes (iv). Moreover, if L irreducible, then A_L is irreducible, that is, A_L satisfies $\sum_{k=0}^{n-1} A^k > 0$ so that also $\exp(-L) > 0$. This establishes statement (v).   ∎

7.3.2 Equilibria and convergence of the Laplacian flow

We can now focus on the Laplacian flow dynamics.

Lemma 7.3 (Equilibrium points). *If G contains a globally reachable node, then the set of equilibrium points of the Laplacian flow (7.5) is $\{\alpha \mathbb{1}_n \mid \alpha \in \mathbb{R}\}$.*

Proof. A point x is an equilibrium for the Laplacian flow if $Lx = \mathbb{0}_n$. Hence, any point in the kernel of the matrix L is an equilibrium. From Theorem 6.6, if G contains a globally reachable node, then $\text{rank}(L) = n-1$. Hence, the dimension of the kernel space is 1. The lemma follows by recalling that $L\mathbb{1}_n = \mathbb{0}_n$.   ∎

We are now interested in characterizing the solution of the Laplacian flow (7.5). To build some intuition, we first consider an undirected graph G and write the modal decomposition of the solution as in Remark 2.3 for a discrete-time linear system. We proceed in two steps. First, because G is undirected, the matrix L is symmetric and has real eigenvalues $0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ with corresponding orthonormal (i.e., orthogonal and unit-length) eigenvectors $v_1, \ldots, v_n$. Define $y_i(t) = v_i^\mathsf{T} x(t)$ and left-multiply $\dot{x} = -Lx$ by v_i:

$$\frac{d}{dt} y_i(t) = -\lambda_i y_i(t), \qquad y_i(0) = v_i^\mathsf{T} x(0).$$

These n decoupled ordinary differential equations are immediately solved to give

$$x(t) = y_1(t)v_1 + y_2(t)v_2 + \cdots + y_n(t)v_n$$
$$= e^{-\lambda_1 t}(v_1^\mathsf{T} x(0))v_1 + e^{-\lambda_2 t}(v_2^\mathsf{T} x(0))v_2 + \cdots + e^{-\lambda_n t}(v_n^\mathsf{T} x(0))v_n.$$

Second, recall that $\lambda_1 = 0$ and $v_1 = \mathbb{1}_n/\sqrt{n}$ because L is a symmetric Laplacian matrix ($L\mathbb{1}_n = \mathbb{0}_n$). Therefore, we compute $(v_1^\mathsf{T} x(0))v_1 = \text{average}(x(0))\mathbb{1}_n$ and substitute

$$x(t) = \text{average}(x(0))\mathbb{1}_n + e^{-\lambda_2 t}(v_2^\mathsf{T} x(0))v_2 + \cdots + e^{-\lambda_n t}(v_n^\mathsf{T} x(0))v_n.$$

Now, let us assume that G is connected so that its second smallest eigenvalue λ_2 is strictly positive. In this case, we can infer that

$$\lim_{t \to \infty} x(t) = \text{average}(x(0))\mathbb{1}_n,$$

or, defining a *disagreement vector* $\delta(t) = x(t) - \text{average}(x(0))\mathbb{1}_n$, we infer

$$\delta(t) = \mathrm{e}^{-\lambda_2 t}(v_2^\mathsf{T} x(0))v_2 + \cdots + \mathrm{e}^{-\lambda_n t}(v_n^\mathsf{T} x(0))v_n.$$

In summary, we discovered that, for a connected undirected graph, the disagreement vector converges to zero with an exponential rate λ_2. In what follows, we state a more general convergence to consensus result for the continuous-time Laplacian flow. This result is parallel to Theorem 5.1.

Theorem 7.4 (Consensus for Laplacian matrices with globally reachable node). *If a Laplacian matrix L has associated digraph G with a globally reachable node, then*

(i) *the eigenvalue 0 of $-L$ is simple and all other eigenvalues of $-L$ have negative real part,*

(ii) $\lim_{t\to\infty} \mathrm{e}^{-Lt} = \mathbb{1}_n w^\mathsf{T}$, *where w is the left eigenvector of L with eigenvalue 0 satisfying $\mathbb{1}_n^\mathsf{T} w = 1$,*

(iii) $w_i \geq 0$ *for all nodes i and $w_i > 0$ if and only if node i is globally reachable,*

(iv) *the solution to $\frac{d}{dt}x(t) = -Lx(t)$ satisfies*

$$\lim_{t\to\infty} x(t) = \left(w^\mathsf{T} x(0)\right)\mathbb{1}_n,$$

(v) *if additionally G is weight-balanced, then G is strongly connected, $\mathbb{1}_n^\mathsf{T} L = \mathbb{0}_n^\mathsf{T}$, $w = \frac{1}{n}\mathbb{1}_n$, and*

$$\lim_{t\to\infty} x(t) = \frac{\mathbb{1}_n^\mathsf{T} x(0)}{n}\mathbb{1}_n = \text{average}\left(x(0)\right)\mathbb{1}_n.$$

Note: as a corollary to the statement (iii), the left eigenvector $w \in \mathbb{R}^n$ associated to the 0 eigenvalue has strictly positive entries if and only if G is strongly connected.

Proof. Because the associated digraph has a globally reachable node, Theorem 6.6 establishes that L has rank $n-1$ and that all eigenvalues of L have non-negative real part. Therefore, also remembering the property $L\mathbb{1}_n = \mathbb{0}_n$, we conclude that 0 is a simple eigenvalue with right eigenvector $\mathbb{1}_n$ and that all other eigenvalues of L have positive real part. This concludes the proof of (i). In what follows we let w denote the left eigenvector associated to the eigenvalue 0, that is, $w^\mathsf{T} L = \mathbb{0}_n^\mathsf{T}$, normalized so that $\mathbb{1}_n^\mathsf{T} w = 1$.

To prove statement (ii), we proceed in three steps. First, we write the Laplacian matrix in its Jordan normal form:

$$L = PJP^{-1} = P \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & J_2 & \ddots & 0 \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & J_m \end{bmatrix} P^{-1}, \tag{7.7}$$

where $m \leq n$ is the number of Jordan blocks, the first block is the scalar 0 (being the only eigenvalue we know), the other Jordan blocks $J_2, \ldots, J_m$ (unique up to re-ordering) are associated with eigenvalues with strictly positive real part, and where the columns of P are the generalized eigenvectors of L (unique up to rescaling).

Second, using some properties from Exercise E7.1, we compute the limit as $t \to \infty$ of $\mathrm{e}^{-Lt} = P\,\mathrm{e}^{-Jt}\,P^{-1}$ as

$$\lim_{t\to\infty} \mathrm{e}^{-Lt} = P \lim_{t\to\infty} \mathrm{e}^{-Jt}\, P^{-1} = P \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \ddots & 0 \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 0 \end{bmatrix} P^{-1} = (P\mathbb{e}_1)(\mathbb{e}_1^{\mathsf{T}} P^{-1}) = c_1 r_1,$$

where c_1 is the first column of P and r_1 is the first row of P^{-1}. The contributions of the Jordan blocks $J_2, \dots, J_m$ vanish because their eigenvalues have negative real part; e.g., for more details see (Hespanha, 2009).

Third and final, we characterize c_1 and r_1. By definition, the first column of P (unique up to rescaling) is a right eigenvector of the eigenvalue 0 for the matrix L, that is, $c_1 = \alpha \mathbb{1}_n$ for some scalar α since we know $L\mathbb{1}_n = \mathbb{0}_n$. Of course, it is convenient to define $c_1 = \mathbb{1}_n$. Next, equation (7.7) can be rewritten as $P^{-1}L = JP^{-1}$, whose first row is $r_1 L = \mathbb{0}_n^{\mathsf{T}}$. This equality implies $r_1 = \beta w^{\mathsf{T}}$ for some scalar β. Finally, we note that $P^{-1}P = I_n$ implies $r_1 c_1 = 1$, that is, $\beta w^{\mathsf{T}} \mathbb{1}_n = 1$. Since we know $w^{\mathsf{T}} \mathbb{1}_n = 1$, we infer that $\beta = 1$ and that $r_1 = w^{\mathsf{T}}$. This concludes the proof of statement (ii).

Next, we prove statement (iii). Pick a positive constant $\varepsilon < 1/d_{\max}$, where the maximum out-degree is $d_{\max} = \max\{d_{\mathrm{out}}(1), \dots, d_{\mathrm{out}}(n)\}$. Define $B = I_n - \varepsilon L$. It is easy to show that B is non-negative, row-stochastic, and has strictly positive diagonal elements. Moreover, $w^{\mathsf{T}}L = \mathbb{0}_n^{\mathsf{T}}$ implies $w^{\mathsf{T}}B = w^{\mathsf{T}}$ so that w is the left eigenvector with unit eigenvalue for B. Now, note that the digraph $G(L)$ associated to L (without self-loops) is identical to the digraph $G(B)$ associated to B, except for the fact that B has self-loops at each node. By assumption $G(L)$ has a globally reachable node and therefore so does $G(B)$, where the subgraph induced by the set of globally reachable nodes is aperiodic (due to the self-loops). Therefore, statement (iii) is now an immediate transcription of the same statement for row-stochastic matrices established in Theorem 5.1 (statement (i)).

Statements (iv) and (v) are straightforward and left as Exercise E7.3.   ∎

7.4   Second-order Laplacian flows

In this section we assume each node of the network obeys a so-called *double-integrator dynamic* (also referred to as *second-order dynamic*):

$$\ddot{x}_i = u_i, \qquad \text{or, in first-order equivalent form,} \qquad \begin{cases} \dot{x}_i &= v_i, \\ \dot{v}_i &= u_i, \end{cases} \tag{7.8}$$

where u_i is an appropriate control input signal to be designed.

We assume a weighted digraph describes the sensing and/or communication interactions among the agents with adjacency matrix A and Laplacian L. We also introduce constants $k_{\mathrm{p}}, k_{\mathrm{d}} \geq 0$ describing a so-called spring and damping coefficients respectively, as well as constants $\gamma_{\mathrm{p}}, \gamma_{\mathrm{d}} \geq 0$ describing position-averaging and velocity-averaging coefficients. Given the following law:

$$u_i = -k_{\mathrm{p}} x_i - k_{\mathrm{d}} \dot{x}_i + \sum_{j=1}^{n} a_{ij}\big(\gamma_{\mathrm{p}}(x_j - x_i) + \gamma_{\mathrm{d}}(\dot{x}_j - \dot{x}_i)\big),$$

the corresponding closed-loop systems, called the *second-order Laplacian flow*, is

$$\ddot{x}(t) + (k_\mathrm{d}I_n + \gamma_\mathrm{d}L)\dot{x}(t) + (k_\mathrm{p}I_n + \gamma_\mathrm{p}L)x(t) = \mathbb{0}_n. \tag{7.9}$$

By introducing the *second-order Laplacian matrix* $\mathcal{L} \in \mathbb{R}^{2n \times 2n}$, we write the system in first-order form:

$$\begin{bmatrix} \dot{x}(t) \\ \dot{v}(t) \end{bmatrix} = \begin{bmatrix} \mathbb{0}_{n \times n} & I_n \\ -k_\mathrm{p}I_n - \gamma_\mathrm{p}L & -k_\mathrm{d}I_n - \gamma_\mathrm{d}L \end{bmatrix} \begin{bmatrix} x(t) \\ v(t) \end{bmatrix} =: \mathcal{L} \begin{bmatrix} x(t) \\ v(t) \end{bmatrix}.$$

Name	Dynamics	References
Second-order consensus protocol ($k_\mathrm{p} = k_\mathrm{d} = 0, \gamma_\mathrm{d} = 1, \gamma_\mathrm{p} > 0$)	$\ddot{x}(t) + L\dot{x}(t) + \gamma_\mathrm{p}Lx(t) = \mathbb{0}_n$	(Ren and Atkins, 2005; Ren, 2008a; Yu et al., 2010)
Harmonic oscillators coupled via velocity averaging ($k_\mathrm{d} = \gamma_\mathrm{p} = 0, \gamma_\mathrm{d} = 1, k_\mathrm{p} > 0$)	$\ddot{x}(t) + L\dot{x}(t) + k_\mathrm{p}x(t) = \mathbb{0}_n$	(Ren, 2008b)
Position-averaging with absoluted velocity damping ($k_\mathrm{p} = \gamma_\mathrm{d} = 0, \gamma_\mathrm{p} = 1, k_\mathrm{d} > 0$)	$\ddot{x}(t) + k_\mathrm{d}\dot{x}(t) + Lx(t) = \mathbb{0}_n$	Exercise E7.12
Arbitrary-sign gains and digraphs (possibly with $L \neq L^\mathsf{T}$)	equation (7.9)	(Zhu et al., 2009). See (Zhang and Tian, 2009) for discrete-time setting.

Table 7.1: Classification of second-order Laplacian flows

It turns out that it is possible to compute the eigenvalues of the second-order Laplacian matrix; we refer to Exercise E7.11 for its eigenvectors.

Theorem 7.5 (Eigenvalues of second-order Laplacian matrices). *Given a Laplacian matrix L and coefficients $k_\mathrm{p}, k_\mathrm{d}, \gamma_\mathrm{p}, \gamma_\mathrm{d} \in \mathbb{R}$,*

(i) *the characteristic polynomial of $\mathcal{L}$ is*

$$\det(\eta I_{2n} - \mathcal{L}) = \det\big(\eta^2 I_n + \eta(k_\mathrm{d}I_n + \gamma_\mathrm{d}L) + (k_\mathrm{p}I_n + \gamma_\mathrm{p}L)\big);$$

(ii) *given the eigenvalues $\lambda_i, i \in \{1, \ldots, n\}$, of L, the $2n$ eigenvalues $\eta_{i,\pm}, i \in \{1, \ldots, n\}$, of $\mathcal{L}$ are solutions to*

$$\eta^2 + (k_\mathrm{d} + \gamma_\mathrm{d}\lambda_i)\eta + (k_\mathrm{p} + \gamma_\mathrm{p}\lambda_i) = 0, \quad i \in \{1, \ldots, n\}. \tag{7.10}$$

Proof. Regarding statement (i), we recall equality (E2.1b) from Exercise E2.16 and compute the characteristic polynomial of $\mathcal{L}$ as:

$$\begin{aligned}
\det(\eta I_{2n} - \mathcal{L}) &= \det\begin{bmatrix} \eta I_n & -I_n \\ k_\mathrm{p}I_n + \gamma_\mathrm{p}L & (\eta + k_\mathrm{d})I_n + \gamma_\mathrm{d}L \end{bmatrix} \\
&= \det\big((\eta I_n)((\eta + k_\mathrm{d})I_n + \gamma_\mathrm{d}L) - (-I_n)(k_\mathrm{p}I_n + \gamma_\mathrm{p}L)\big) \\
&= \det\big(\eta^2 I_n + \eta(k_\mathrm{d}I_n + \gamma_\mathrm{d}L) + (k_\mathrm{p}I_n + \gamma_\mathrm{p}L)\big).
\end{aligned}$$

Regarding statement (ii), let J_L be the Jordan normal form of L, i.e., let $L = TJ_LT^{-1}$ for an appropriate invertible T, and note

$$\det(\eta I_{2n} - \mathcal{L}) = \det\big(\eta^2 I_n + \eta(k_\mathrm{d}I_n + \gamma_\mathrm{d}J_L) + (k_\mathrm{p}I_n + \gamma_\mathrm{p}J_L)\big)$$
$$= \prod_{i=1}^{n}\big(\eta^2 + (k_\mathrm{d} + \gamma_\mathrm{d}\lambda_i)\eta + (k_\mathrm{p} + \gamma_\mathrm{p}\lambda_i)\big).$$

Therefore, the $2n$ solutions to the characteristic equation $\det(\eta I_{2n} - \mathcal{L}) = 0$ are n pairs of solutions $\eta_{2i,2i-i}$, $i \in \{1, \dots, n\}$, for the second-order equations (7.10). This concludes task (ii). ∎

Next, we provide a necessary and sufficient characterization of a so-called asymptotic second-order consensus concept.

Theorem 7.6 (Asymptotic second-order consensus). *Consider the second-order Laplacian flow* (7.9). *The following statements are equivalent:*

(i) *the second-order Laplacian flow achieves* asymptotic second-order consensus, *that is,* $|x_i - x_j| \to 0$ *and* $|\dot{x}_i - \dot{x}_j| \to 0$ *as* $t \to \infty$ *for all* $i, j \in \{1, \dots, n\}$, *and*

(ii) *the* $2(n-1)$ *eigenvalues* $\eta_{i,\pm}$, $i \in \{2, \dots, n\}$, *of the second-order Laplacian matrix* $\mathcal{L}$ *have strictly negative real part.*

Proof. We introduce the following change of coordinates: $Tx(t) = \begin{bmatrix} x_\mathrm{ave}(t) \\ \delta(t) \end{bmatrix}$, where $x_\mathrm{ave}(t) = \mathrm{average}(x(t))$, $\delta(t) \in \mathbb{R}^{n-1}$, and, from Exercise E2.3, $T = \begin{bmatrix} 1/n & 1/n & \dots & 1/n \\ -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{bmatrix}$. Correspondingly, we also have

$T\dot{x}(t) = \begin{bmatrix} \dot{x}_\mathrm{ave}(t) \\ \dot{\delta}(t) \end{bmatrix}$. To write the system in the new coordinates, we observe $T\mathbb{1}_n = e_1$ and compute

$$TLT^{-1}e_1 = TLT^{-1}(T\mathbb{1}_n) = TL\mathbb{1}_n = \mathbb{0}_n,$$

where the last equality follows from $L\mathbb{1}_n = \mathbb{0}_n$. This implies that the first column of TLT^{-1} is $\mathbb{0}_n$, that is,

$$TLT^{-1} = \begin{bmatrix} 0 & c^\mathsf{T} \\ \mathbb{0}_{n-1} & L_\mathrm{red} \end{bmatrix}, \text{ for } L_\mathrm{red} \in \mathbb{R}^{(n-1)\times(n-1)} \text{ and } c \in \mathbb{R}^{n-1}, \tag{7.11}$$

so that $\mathrm{spec}(L) = \{0\} \cup \mathrm{spec}(L_\mathrm{red})$. Next, we compute

$$\begin{bmatrix} T & \mathbb{0}_{n\times n} \\ \mathbb{0}_{n\times n} & T \end{bmatrix}\begin{bmatrix} \mathbb{0}_{n\times n} & I_n \\ -k_\mathrm{p}I_n - \gamma_\mathrm{p}L & -k_\mathrm{d}I_n - \gamma_\mathrm{d}L \end{bmatrix}\begin{bmatrix} T^{-1} & \mathbb{0}_{n\times n} \\ \mathbb{0}_{n\times n} & T^{-1} \end{bmatrix}$$
$$= \begin{bmatrix} \mathbb{0}_{n\times n} & T \\ -k_\mathrm{p}T - \gamma_\mathrm{p}TL & -k_\mathrm{d}T - \gamma_\mathrm{d}TL \end{bmatrix}\begin{bmatrix} T^{-1} & \mathbb{0}_{n\times n} \\ \mathbb{0}_{n\times n} & T^{-1} \end{bmatrix} = \begin{bmatrix} \mathbb{0}_{n\times n} & I_n \\ -k_\mathrm{p}I_n - \gamma_\mathrm{p}TLT^{-1} & -k_\mathrm{d}I_n - \gamma_\mathrm{d}TLT^{-1} \end{bmatrix}.$$
$$\tag{7.12}$$

Based on equations (7.11) and (7.12), we write the system in these new coordinates as

$$
\frac{d}{dt}\begin{bmatrix} x_{\mathrm{ave}} \\ \delta \\ \dot{x}_{\mathrm{ave}} \\ \dot{\delta} \end{bmatrix} = \begin{bmatrix} 0 & \mathbb{0}_{n-1}^{\mathsf{T}} & 1 & 0 \\ \mathbb{0}_{n-1} & \mathbb{0}_{(n-1)\times(n-1)} & \mathbb{0}_{n-1} & I_{n-1} \\ -k_{\mathrm{p}} & -\gamma_{\mathrm{p}}c^{\mathsf{T}} & -k_{\mathrm{d}} & -\gamma_{\mathrm{d}}c^{\mathsf{T}} \\ \mathbb{0}_{n-1} & -k_{\mathrm{p}}I_{n-1} - \gamma_{\mathrm{p}}L_{\mathrm{red}} & \mathbb{0}_{n-1} & -k_{\mathrm{d}}I_{n-1} - \gamma_{\mathrm{d}}L_{\mathrm{red}} \end{bmatrix} \begin{bmatrix} x_{\mathrm{ave}} \\ \delta \\ \dot{x}_{\mathrm{ave}} \\ \dot{\delta} \end{bmatrix}.
$$

We reorder the variables to obtain a block-diagonal matrix, whose eigenvalues are the eigenvalues of the diagonal blocks:

$$
\frac{d}{dt}\begin{bmatrix} x_{\mathrm{ave}} \\ \dot{x}_{\mathrm{ave}} \\ \delta \\ \dot{\delta} \end{bmatrix} = \begin{bmatrix} 0 & 1 & \mathbb{0}_{n-1}^{\mathsf{T}} & 0 \\ -k_{\mathrm{p}} & -k_{\mathrm{d}} & -\gamma_{\mathrm{p}}c^{\mathsf{T}} & -\gamma_{\mathrm{d}}c^{\mathsf{T}} \\ \mathbb{0}_{n-1} & \mathbb{0}_{n-1} & \mathbb{0}_{(n-1)\times(n-1)} & I_{n-1} \\ \mathbb{0}_{n-1} & \mathbb{0}_{n-1} & -k_{\mathrm{p}}I_{n-1} - \gamma_{\mathrm{p}}L_{\mathrm{red}} & -k_{\mathrm{d}}I_{n-1} - \gamma_{\mathrm{d}}L_{\mathrm{red}} \end{bmatrix} \begin{bmatrix} x_{\mathrm{ave}} \\ \dot{x}_{\mathrm{ave}} \\ \delta \\ \dot{\delta} \end{bmatrix}.
$$

We are now ready to conclude the proof: asymptotic second-order consensus is achieved if and only if $\delta \to \mathbb{0}_{n-1}$ and $\dot{\delta} \to \mathbb{0}_{n-1}$ as $t \to \infty$ if and only if all eigenvalues of $\begin{bmatrix} 0 & I_{n-1} \\ -k_{\mathrm{p}}I_{n-1} - \gamma_{\mathrm{p}}L_{\mathrm{red}} & -k_{\mathrm{d}}I_{n-1} - \gamma_{\mathrm{d}}L_{\mathrm{red}} \end{bmatrix}$ have strictly negative real part. But these eigenvalues are precisely the $2(n-1)$ eigenvalues $\eta_{i,\pm}$, $i \in \{2, \ldots, n\}$, of the second-order Laplacian matrix $\mathcal{L}$. ∎

Finally, we restrict out attention to undirected graphs and positive gains and present convergence results for this setting.

Theorem 7.7 (Asymptotic convergence of second-order Laplacian flows). *Consider the second-order Laplacian flow* (7.9). *Assume L is symmetric and irreducible (i.e., its associated digraph is undirected and connected). Define the* state average *and its time derivative by: $x_{\mathrm{ave}}(t) = \mathrm{average}\big(x(t)\big)$ and $\dot{x}_{\mathrm{ave}}(t) = \mathrm{average}\big(\dot{x}(t)\big)$. Then the state averages satisfy*

$$
\frac{d}{dt}\begin{bmatrix} x_{\mathrm{ave}}(t) \\ \dot{x}_{\mathrm{ave}}(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -k_{\mathrm{p}} & -k_{\mathrm{d}} \end{bmatrix} \begin{bmatrix} x_{\mathrm{ave}}(t) \\ \dot{x}_{\mathrm{ave}}(t) \end{bmatrix},
\tag{7.13}
$$

and, moreover,

(i) *for the second-order consensus protocol ($k_{\mathrm{p}} = k_{\mathrm{d}} = 0$, $\gamma_{\mathrm{d}} = 1$, $\gamma_{\mathrm{p}} > 0$), asymptotic consensus on a ramp signal is achieved, that is, as $t \to \infty$,*

$$
x(t) \to \Big(x_{\mathrm{ave}}(0) + \dot{x}_{\mathrm{ave}}(0)t\Big)\mathbb{1}_n;
$$

(ii) *for the harmonic oscillators coupled via velocity averaging ($k_{\mathrm{d}} = \gamma_{\mathrm{p}} = 0$, $\gamma_{\mathrm{d}} = 1$, $k_{\mathrm{p}} > 0$), asymptotic synchronization on an harmonic signal with frequency $\sqrt{k_{\mathrm{p}}}$ is achieved, that is, as $t \to \infty$,*

$$
x(t) \to \Big(x_{\mathrm{ave}}(0)\cos(\sqrt{k_{\mathrm{p}}}t) + \frac{1}{\sqrt{k_{\mathrm{p}}}}\dot{x}_{\mathrm{ave}}(0)\sin(\sqrt{k_{\mathrm{p}}}t)\Big)\mathbb{1}_n;
$$

(iii) for the position-averaging flow with absolute velocity damping ($k_p = \gamma_d = 0$, $\gamma_p = 1$, $k_d > 0$), asymptotic consensus on a weighted average value is achieved, that is, as $t \to \infty$

$$x(t) \to \Big(x_{\mathrm{ave}}(0) + \dot{x}_{\mathrm{ave}}(0)/k_d\Big)\mathbb{1}_n.$$

Proof. First, we show that, in the similarity transformation (7.11), if L is symmetric, then $c = \mathbb{0}_{n-1}$. To do this, we observe $\mathbb{e}_1^\mathsf{T} T = (1/n)\mathbb{1}_n^\mathsf{T}$ and compute

$$\mathbb{e}_1^\mathsf{T} TLT^{-1} = \mathbb{1}_n^\mathsf{T} LT^{-1} = \mathbb{0}_n^\mathsf{T},$$

where the last equality follows from $L^\mathsf{T}\mathbb{1}_n = \mathbb{0}_n$. This implies that the first row of TLT^{-1} is $\mathbb{0}_n$ and, in turn, that equation (7.13) are correct. Second, for the index range $i \in \{2,\dots,n\}$, in all three cases the second-order polynomial (7.10) has strictly positive coefficients, which implies that the $2(n-1)$ eigenvalues $\eta_{i,\pm}$, $i \in \{2,\dots,n\}$, of the second-order Laplacian matrix $\mathcal{L}$ have strictly negative real part. Therefore, by Theorem 7.6, the second-order Laplacian flow achieves asymptotic second-order consensus and, more specifically, $x_i(t) - (x_{\mathrm{ave}}(t))_i \to 0$ and $\dot{x}_i(t) - (\dot{x}_{\mathrm{ave}}(t))_i \to 0$ for all $i \in \{1,\dots,n\}$. Third and finally , the specific values for $x_{\mathrm{ave}}(t)$ follow from explicitly solving the state average dynamics (7.13). We leave the details to the reader. ∎

The three scenarios discussed in Theorem 7.7 are illustrated in Figure 7.3. Case (i) with relative position and velocity coupling, $\gamma_d = 1$, $\gamma_p > 0$, leads to consensus on a ramp, which is relevant in car platooning problems. Case (ii) with relative velocity and absolute position feedback, $\gamma_d = 1$, $k_p > 0$, leads to a consensus on harmonic oscillations, which can be found in the synchronization of electronic oscillators, where the states correspond to voltages and currents of resistively resonant (parallel LC) circuits. Finally, case (iii) with relative position and absolute velocity feedback, $\gamma_p = 1$, $k_d > 0$, leads to a consensus in positions, and it can be found in robotic consensus problems or in power network swing dynamics.

7.5 Appendix: Design of weight-balanced digraphs

Problem: Given a directed graph G that is strongly connected, but not weight-balanced, how do we choose the weights in order to obtain a weight-balanced digraph and a Laplacian satisfying $\mathbb{1}_n^\mathsf{T} L = \mathbb{0}_n$? (Note that an undirected graph is automatically weight-balanced.)

Answer: As usual, let $w > 0$ be the left eigenvector of L with eigenvalue 0 satisfying $w_1 + \cdots + w_n = 1$. In other words, w is a vector of convex combination coefficients, and the Laplacian L satisfies

$$L\mathbb{1}_n = \mathbb{0}_n, \quad \text{and} \quad w^\mathsf{T} L = \mathbb{0}_n^\mathsf{T}.$$

Define now a new matrix:

$$L_{\mathrm{rescaled}} = \mathrm{diag}(w)L.$$

It is immediate to see that

$$L_{\mathrm{rescaled}}\mathbb{1}_n = \mathrm{diag}(w)L\mathbb{1}_n = \mathbb{0}_n, \qquad \mathbb{1}_n^\mathsf{T} L_{\mathrm{rescaled}} = \mathbb{1}_n^\mathsf{T} \mathrm{diag}(w)L = w^\mathsf{T} L = \mathbb{0}_n^\mathsf{T}.$$

Note that:

(a) Case (i): asymptotic consensus on a ramp signal

(b) Case (ii): asymptotic synchronization on an harmonic signal

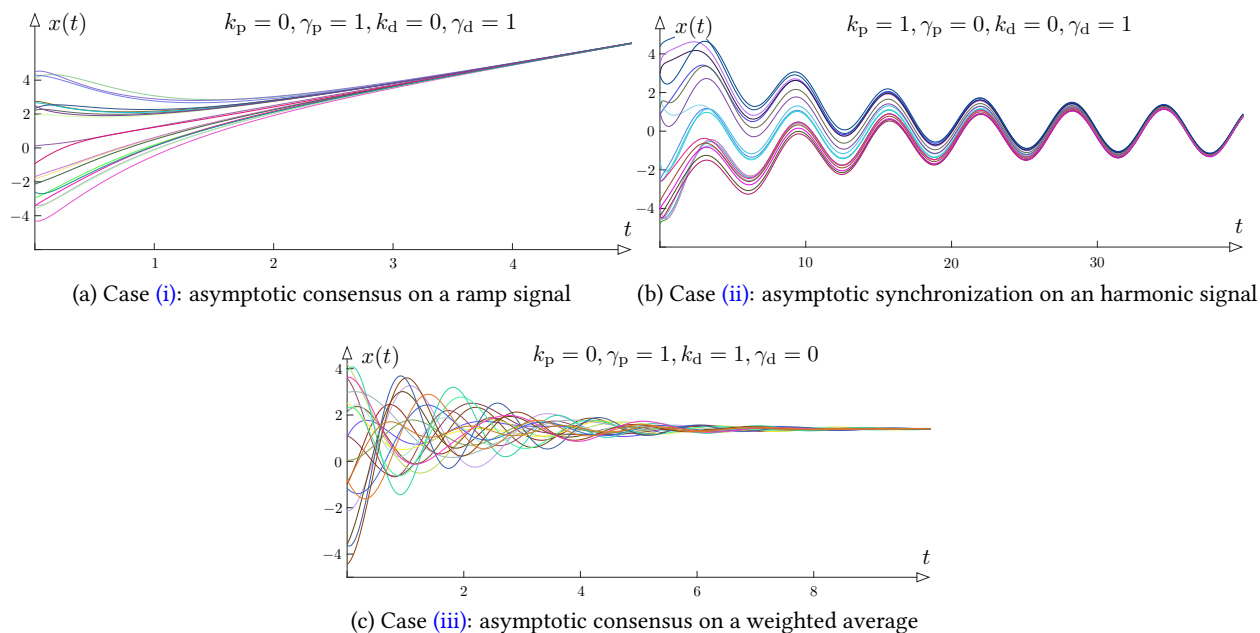(c) Case (iii): asymptotic consensus on a weighted average

Figure 7.3: Representative trajectories of the second-order Laplacian flow (7.9) for a randomly-generated undirected graph with $n = 20$ nodes, random initial conditions, and the three choices of gains discussed in Theorem 7.7.

- L_{rescaled} is again a Laplacian matrix because (i) its row-sums are zero, (ii) its diagonal entries are positive, and (iii) its non-diagonal entries are non-positive;

- L_{rescaled} is the Laplacian matrix for a new digraph G_{rescaled} with the same nodes and directed edges as G, but whose weights are rescaled as follows: $a_{ij} \mapsto w_i a_{ij}$. In other words, the weight of each out-edge of node i is rescaled by w_i.

7.6 Appendix: Distributed optimization using the Laplacian flow

In the following, we present a computational application of the Laplacian flow in distributed optimization. The materials in this section are inspired by (Wang and Elia, 2010; Gharesifard and Cortes, 2014; Droge et al., 2014; Cherukuri et al., 2017), and we present them here in a self-contained way. As only preliminaries notions, we introduce the following two definitions: A function $f : \mathbb{R}^n \to \mathbb{R}$ is said be *convex* if $f(\alpha x + \beta y) \le \alpha f(x) + \beta f(y)$ for all x and y in $\mathbb{R}^n$ and for all convex combination coefficients α and β, i.e., coefficients satisfying $\alpha, \beta \ge 0$ and $\alpha + \beta = 1$. A function is said to be *strictly convex* if the previous inequality holds strictly.

Consider a network of n processors that can perform local computation and communicate with another. The communication architecture is modeled by an undirected, connected, and weighted graph with n nodes and symmetric Laplacian $L = L^\mathsf{T} \in \mathbb{R}^{n \times n}$. The objective of the processor network is to solve the

optimization problem

$$\text{minimize } _{x\in\mathbb{R}} \ f(x) = \sum_{i=1}^{n} f_i(x), \tag{7.14}$$

where $f_i : \mathbb{R} \to \mathbb{R}$ is a strictly convex and twice continuously differentiable cost function known only to processor $i \in \{1, \ldots, n\}$. In a centralized setup, the decision variable x is globally available and the minimizers $x^* \in \mathbb{R}$ of the optimization problem (7.14) can be found by solving for the *critical points* of $f(x)$

$$\mathbb{0}_n = \frac{\partial}{\partial x} f(x) = \sum_{i=1}^{n} \frac{\partial}{\partial x} f_i(x).$$

A centralized continuous-time algorithm converging to the set of critical points is the *negative gradient flow*

$$\dot{x} = -\frac{\partial}{\partial x} f(x) \,.$$

To find a distributed approach to solving the optimization problem (7.14), we associate a local estimate $y_i \in \mathbb{R}$ of the global variable $x \in \mathbb{R}$ to every processor and solve the *equivalent problem*

$$\text{minimize } _{y\in\mathbb{R}^n} \ \tilde{f}(y) = \sum_{i=1}^{n} f_i(y_i) + \frac{1}{2} y^\mathsf{T} L y \qquad \text{subject to } Ly = \mathbb{0}_n, \tag{7.15}$$

where the consistency constraint $Ly = \mathbb{0}_n$ assures that $y_i = y_j$ for all $i, j \in \{1, \ldots, n\}$, that is, the local estimates of all processors coincide. We also augmented the cost function with the term $y^\mathsf{T} L y$, which clearly has no effect on the minimizers of (7.15) (due to the consistency constraint), but it provides supplementary damping and favorable convergence properties for our algorithm. The minimizers of the optimization problems (7.14) and (7.15) are then related by $y^* = x^* \mathbb{1}_n$.

   Without any further motivation, consider the function $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ given by

$$\mathcal{L}(y, z) = f(y) + \frac{1}{2} y^\mathsf{T} L y + z^\mathsf{T} L y.$$

In the literature on convex optimization this function is known as (augmented) Lagrangian function and $z \in \mathbb{R}^n$ is referred to as Lagrange multiplier. What is important for us is that the augmented Lagrangian function is strictly convex in y and linear (and hence[1] concave) in z. Hence, the augmented Lagrangian function admits a set of *saddle points* $(y^*, z^*) \in \mathbb{R}^n \times \mathbb{R}^n$, that is points satisfying

$$\mathcal{L}(y^*, z) \leq \mathcal{L}(y^*, z^*) \leq \mathcal{L}(y, z^*) \qquad \text{for all } (y, z) \in \mathbb{R}^n \times \mathbb{R}^n \,.$$

Since $\mathcal{L}(y, z)$ is differentiable in y and z, the saddle points can be obtained as solutions to the equations

$$\mathbb{0}_n = \frac{\partial}{\partial y} \mathcal{L}(y, z) = \frac{\partial}{\partial y} f(y) + Ly + Lz,$$

$$\mathbb{0}_n = \frac{\partial}{\partial z} \mathcal{L}(y, z) = Ly.$$

Our motivation for introducing the Lagrangian is the following lemma.

---

[1] A function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be *concave* (resp. *strictly concave*) if $-f(x)$ is a convex (resp. strictly convex) function.

Lemma 7.8 (Properties of saddle points). *Let $L = L^{\mathsf{T}} \in \mathbb{R}^{n \times n}$ be a symmetric Laplacian associated to an undirected, connected, and weighted graph, and consider the Lagrangian function $\mathcal{L}$, where each f_i is strictly convex and twice continuously differentiable for all $i \in \{1, \ldots, n\}$. Then*

(i) *if $(y^*, z^*) \in \mathbb{R}^n \times \mathbb{R}^n$ is a saddle point of $\mathcal{L}$, then so is $(y^*, z^* + \alpha \mathbb{1}_n)$ for any $\alpha \in \mathbb{R}$;*

(ii) *if $(y^*, z^*) \in \mathbb{R}^n \times \mathbb{R}^n$ is a saddle point of $\mathcal{L}$, then $y^* = x^* \mathbb{1}_n$ where $x^* \in \mathbb{R}$ is a solution of the original optimization problem (7.14); and*

(iii) *if $x^* \in \mathbb{R}$ is a solution of the original optimization problem (7.14), then there are $z^* \in \mathbb{R}^n$ and $y^* = x^* \mathbb{1}_n$ satisfying $Lz^* + \frac{\partial}{\partial y} \tilde{f}(y^*) = \mathbb{0}_n$ so that (y^*, z^*) is a saddle point of $\mathcal{L}$.*

We leave the proof to the reader in Exercise E7.15. Since the Lagrangian function is convex in y and concave in z, we can compute its saddle points by following the so-called *saddle-point dynamics*, consisting of a positive and negative gradient:

$$\dot{y} = -\frac{\partial}{\partial y}\mathcal{L}(y,z) = -\frac{\partial}{\partial y}f(y) - Ly - Lz, \tag{7.16a}$$

$$\dot{z} = \frac{\partial}{\partial z}\mathcal{L}(y,z) = Ly. \tag{7.16b}$$

For processor $i \in \{1, \ldots, n\}$, the saddle-point dynamics (7.16) read component-wise as

$$\dot{y}_i = -\frac{\partial}{\partial y_i}f_i(y_i) - \sum_{j=1}^{n} a_{ij}(y_i - y_j) - \sum_{j=1}^{n} a_{ij}(z_i - z_j),$$

$$\dot{z}_i = \frac{\partial}{\partial z_i}\mathcal{L}(y,z) = \sum_{j=1}^{n} a_{ij}(y_i - y_j).$$

Hence, the saddle-point dynamics can be implemented in a distributed processor network using only local knowledge of $f_i(y_i)$, local computation, nearest-neighbor communication and—of course—after discretizing the continuous-time dynamics; see Exercise E7.18. As shown in (Wang and Elia, 2010; Gharesifard and Cortes, 2014; Droge et al., 2014; Cherukuri et al., 2017), this distributed optimization setup is very versatile and robust and extends to directed graphs and non-differentiable convex objective functions. We will later establish using a powerful tool termed Krasovskiĭ-LaSalle Invariance Principle to show that the saddle-point dynamics (7.16) always converge to the set of saddle points; see Exercise E14.4.

For now we restrict our analysis to the case of quadratic cost functions $f_i(x) = P_i(x - \underline{x}_i)^2$, where $P_i > 0$ and $\underline{x}_i \in \mathbb{R}$. Thus, the cost function reads, up to a constant scalar, as

$$f(x) = \sum_{i=1}^{n} (x - \underline{x}_i)^{\mathsf{T}} P_i (x - \underline{x}_i) = \sum_{i=1}^{n} (x - x^*)^{\mathsf{T}} P_i (x - x^*) + \mathcal{O}(1),$$

where x^* is the weighted average $x^* = \left(\sum_{i=1}^{n} P_i\right)^{-1} \sum_{i=1}^{n} P_i \underline{x}_i$, which is the global minimizer of f (as obtained by $\partial f(x)/\partial x = \mathbb{0}_n$); see Exercise E7.17. In this case, the saddle-point dynamics (7.16) reduce to the linear system

$$\begin{bmatrix} \dot{\tilde{y}} \\ \dot{z} \end{bmatrix} = \underbrace{\begin{bmatrix} -P - L & -L \\ L & \mathbb{0}_{n \times n} \end{bmatrix}}_{=\mathcal{A}} \begin{bmatrix} \tilde{y} \\ z \end{bmatrix}, \tag{7.17}$$

where $\tilde{y} = y - x^*\mathbb{1}_n$ and $P = \mathrm{diag}(\{P_i\}_{i\in\{1,\dots,n\}})$. In what follows, we will establish the convergence of the dynamics (7.17) to the set of saddle points. First, observe that 0 is an eigenvalue of $\mathcal{A}$ with multiplicity 1 and the corresponding eigenvector, given by $\begin{bmatrix} \mathbb{0}_n^\mathsf{T} & \mathbb{1}_n^\mathsf{T} \end{bmatrix}^\mathsf{T}$ corresponds to the set of saddle points:

$$\begin{bmatrix} \mathbb{0}_n \\ \mathbb{0}_n \end{bmatrix} = \begin{bmatrix} -P - L & -L \\ L & \mathbb{0}_{n\times n} \end{bmatrix} \begin{bmatrix} \tilde{y} \\ z \end{bmatrix} \implies (P+L)\tilde{y} + Lz = \mathbb{0}_n \text{ and } L\tilde{y} = \mathbb{0}_n \implies \tilde{y} \in \mathrm{span}(\mathbb{1}_n)$$

$$\implies \tilde{y}^\mathsf{T} P \tilde{y} = \mathbb{0}_n \text{ obtained by multiplying } (P+L)\tilde{y} + Lz = \mathbb{0}_n \text{ by } \tilde{y}^\mathsf{T}$$

$$\implies \tilde{y} = \mathbb{0}_n \text{ and } z = \mathbb{1}_n .$$

Next, note that for any $z_1, z_2 \in \mathbb{R}^n$,

$$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix}^\mathsf{T} \begin{bmatrix} -P - L & -L \\ L & \mathbb{0}_{n\times n} \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} z_1^\mathsf{T} & z_2^\mathsf{T} \end{bmatrix} \begin{bmatrix} (-P-L)z_1 - Lz_2 \\ Lz_1 \end{bmatrix}$$

$$= z_1^\mathsf{T}(-P-L)z_1 - z_1^\mathsf{T} L z_2 + z_2^\mathsf{T} L z_1 = -z_1^\mathsf{T}(P+L)z_1 \le 0,$$

because L is symmetric and both P and L are positive semidefinite. This inequality implies that $\mathcal{A}$ is negative semidefinite. Since there is a unique zero eigenvalue associated with the set of saddle points, it remains to show that the matrix $\mathcal{A}$ has no purely imaginary eigenvalues. This is established in the following lemma whose proof is left to the reader in Exercise E7.16:

Lemma 7.9 (Absence of imaginary eigenvalues in saddle matrices (Benzi et al., 2005)). *Given a negative semidefinite matrix $\mathcal{B} \in \mathbb{R}^{n\times n}$ and a not necessarily square matrix $\mathcal{C} \in \mathbb{R}^{n\times m}$, define the* saddle *matrix $\mathcal{A} \in \mathbb{R}^{(n+m)\times(n+m)}$ by*

$$\mathcal{A} = \begin{bmatrix} \mathcal{B} & \mathcal{C} \\ -\mathcal{C}^\mathsf{T} & \mathbb{0}_{m\times m} \end{bmatrix}.$$

If $\mathrm{kernel}(\mathcal{B}) \cap \mathrm{image}(\mathcal{C}) = \{\mathbb{0}_n\}$, then the saddle matrix $\mathcal{A}$ has no eigenvalues on the imaginary axis except for 0.

It follows that the saddle point dynamics (7.17) converge to the set of saddle points $\begin{bmatrix} \tilde{y}^\mathsf{T} & z^\mathsf{T} \end{bmatrix}^\mathsf{T} \in \mathrm{span}\left(\begin{bmatrix} \mathbb{0}_n^\mathsf{T} & \mathbb{1}_n^\mathsf{T} \end{bmatrix}^\mathsf{T}\right)$. Since $\mathbb{1}_n^\mathsf{T}\dot{z} = 0$, it follows that $\mathrm{average}(z(t)) = \mathrm{average}(z_0)$, we can further conclude that the dynamics converge to a unique saddle point satisfying $\lim_{t\to\infty} y(t) = x^*\mathbb{1}_n$ and $\lim_{t\to\infty} z(t) = z_0\mathbb{1}_n$.

7.7 Historical notes and further reading

Section 7.1.1 "Example #1: Continuous-time opinion dynamics" presents the continuous-time averaging model by (Abelson, 1964) and its relationship with the discrete-time averaging model by (French, 1956; Harary, 1959; DeGroot, 1974). Abelson's work is one of the earliest on what we now call the Laplacian flow.

Regarding Example #2: "Flocking behavior for a group of animals" in Section 7.1.2, a classic early reference on this topic is (Reynolds, 1987). In that model, flocking behavior is controlled by three simple rules: Separation - avoid crowding neighbors (short range repulsion) Alignment - steer towards average heading of neighbors, and Cohesion - steer towards average position of neighbors (long range attraction).

The RC circuit example in Section 7.1.3 is taken from (Mesbahi and Egerstedt, 2010; Ren et al., 2007).

An early reference to Theorem 7.4 is the work by Abelson (1964) in mathematical sociology; more recent references with rigorous proofs in the control literature include (Lin et al., 2005; Ren and Beard, 2005).

Second-order Laplacian flows are widely studied. Early references are the works by Chow (1982) and Chow and Kokotović (1985) on slow coherency and area aggregation of power networks, modelled as first and second-order Laplacian flows; see also (Avramovic et al., 1980; Chow et al., 1984; Saksena et al., 1984) among others.

In the consensus literature, an early reference to second-order Laplacian flows is (Ren and Atkins, 2005). Relevant references include (Ren, 2008a,b; Zhu et al., 2009; Zhang and Tian, 2009; Yu et al., 2010). A proof of Theorem 7.6 based on the Jordan normal form is given in (Ren and Atkins, 2005; Ren, 2008b). We refer to (Zhu et al., 2009) for convergence results for general digraphs and gains with arbitrary signs.

For historical and additional results on diffusively-coupled identical linear systems (as described in Exercise E7.21, we refer to (Wu and Chua, 1995; Scardovi and Sepulchre, 2009; Li et al., 2010; Yang et al., 2011; Jafarpour et al., 2017).

A reference for the construction in Section 7.5 is (Ren et al., 2007).

7.8 Exercises

E7.1 **Properties of the matrix exponential.** Recall the definition of $e^A = \sum_{k=0}^{\infty} \frac{1}{k!} A^k$ for any square matrix A. Complete the following tasks:

    (i) show that $\sum_{k=0}^{\infty} \frac{1}{k!} A^k$ converges absolutely for all square matrices A,

        Hint: Recall that a matrix series $\sum_{k=1}^{\infty} A_k$ is said to converge absolutely if $\sum_{k=1}^{\infty} \|A_k\|$ converges, where $\| \cdot \|$ is a matrix norm. Introduce a sub-multiplicative matrix norm $\| \cdot \|$ and show $\|e^A\| \le e^{\|A\|}$.

    (ii) show that, if $A = \operatorname{diag}(a_1, \ldots, a_n)$, then $e^A = \operatorname{diag}(e^{a_1}, \ldots e^{a_n})$,

    (iii) show that $AB = BA$ implies $e^{AB} = e^A\, e^B$,

    (iv) show that $e^{TAT^{-1}} = T\, e^A\, T^{-1}$ for any invertible T, and

    (v) compute the matrix exponential of e^{tJ} where J is a Jordan block of arbitrary size and $t \in \mathbb{R}$.

E7.2 **Continuous-time affine systems.** Given $A \in \mathbb{R}^{n \times n}$ and $b \in \mathbb{R}^n$, consider the continuous-time affine systems

$$\dot{x}(t) = Ax(t) + b.$$

Assume A is Hurwitz and, similarly to Exercise E2.10, show that

    (i) the matrix A is invertible,

    (ii) the only equilibrium point of the system is $-A^{-1}b$, and

    (iii) $\lim_{t \to \infty} x(t) = -A^{-1}b$ for all initial conditions $x(0) \in \mathbb{R}^n$.

E7.3 **Consensus for Laplacian matrices: missing proofs.** Complete the proof of Theorem 7.4, that is, prove statements (iv) and (v).

E7.4 **Laplacian average consensus in directed networks.** Consider the directed network in figure below with arbitrary positive weights and its associated Laplacian flow $\dot{x}(t) = -L(x(t))$.
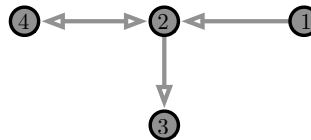
    (i) Can the network reach consensus, that is, as $t \to \infty$ does $x(t)$ converge to a limiting point in $\operatorname{span}\{\mathbb{1}_n\}$?

    (ii) Does $x(t)$ achieve average consensus, that is, $\lim_{t \to \infty} x(t) = \operatorname{average}(x_0)\mathbb{1}_n$?

    (iii) Will your answers change if you smartly add one directed edge and adapt the weights?

E7.5 **Convergence of discrete-time and continuous-time averaging.** Consider the following two weighted digraphs and their associated non-negative adjacency matrices A and Laplacian matrices L of appropriate dimensions. Consider the associated discrete-time iterations $x(t+1) = Ax(t)$ and continuous-time Laplacian flows $\dot{x}(t) = -Lx(t)$. For each of these two digraphs, argue about whether the discrete and/or continuous-time systems converge as $t \to \infty$. If they converge, what do they converge to? Please justify your answers.

E7.6 **Euler discretization of the Laplacian.** Given a weighted digraph G with Laplacian matrix L and maximum out-degree $d_{\max} = \max\{d_{\text{out}}(1), \ldots, d_{\text{out}}(n)\}$. Show that:

    (i) if $\varepsilon < 1/d_{\max}$, then the matrix $I_n - \varepsilon L$ is row-stochastic,

    (ii) if $\varepsilon < 1/d_{\max}$ and G is weight-balanced, then the matrix $I_n - \varepsilon L$ is doubly-stochastic, and

    (iii) if $\varepsilon < 1/d_{\max}$ and G is strongly connected, then $I_n - \varepsilon L$ is primitive.

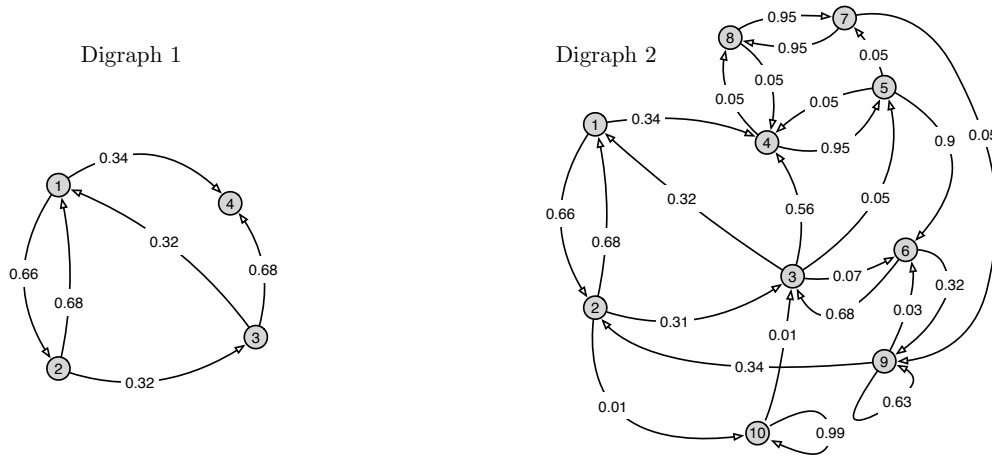Given these results, note that (no additional assignment in what follows)

Figure E7.1: Two example weighted digraphs

- $I_n - \varepsilon L$ is the one-step Euler discretization of the continuous-time Laplacian flow and is a discrete-time consensus algorithm; and
- $I_n - \varepsilon L$ is a possible choice of weights for an undirected unweighted graph (which is therefore also weight-balanced) in the design of a doubly-stochastic matrix (as we did in the discussion about Metropolis-Hastings).

E7.7 **Doubly-stochastic matrices on strongly-connected digraphs.** Given a strongly-connected unweighted digraph G, design weights along the edges of G (and possibly add self-loops) so that the weighted adjacency matrix is doubly-stochastic.

E7.8 **Constants of motion.** In the study of mechanics, energy and momentum are two constants of motion, that is, these quantities are constant along each evolution of the mechanical system. Show that

   (i) If A is a row stochastic matrix with $w^\mathsf{T} A = w^\mathsf{T}$, then $w^\mathsf{T} x(k) = w^\mathsf{T} x(0)$ for all times $k \in \mathbb{Z}_{\geq 0}$ where $x(k+1) = Ax(k)$.

   (ii) If L is a Laplacian matrix with with $w^\mathsf{T} L = \mathbb{0}_n^\mathsf{T}$, then $w^\mathsf{T} x(t) = w^\mathsf{T} x(0)$ for all times $t \in \mathbb{R}_{\geq 0}$ where $\dot{x}(t) = -Lx(t)$.

E7.9 **Weight-balanced digraphs with a globally reachable node.** Given a weighted directed graph G, show that if G is weight-balanced and has a globally reachable node, then G is strongly connected.

E7.10 **The Lyapunov inequality for the Laplacian matrix of a strongly-connected digraph.** Let L be the Laplacian matrix of a strongly-connected weighted digraph. Find a positive-definite matrix P such that

   (i) $PL + L^\mathsf{T} P$ is positive semidefinite (this is the so-called *Lyapunov inequality*), and

   (ii) $(PL + L^\mathsf{T} P)\mathbb{1}_n = \mathbb{0}_n$.

E7.11 **Eigenvectors of the second-order Laplacian matrix.** Consider a Laplacian matrix L, scalar coefficients $k_\mathrm{p}, k_\mathrm{d}, \gamma_\mathrm{p}, \gamma_\mathrm{d} \in \mathbb{R}$ and the induced second-order Laplacian matrix $\mathcal{L}$. Let $v_{\mathrm{l},i}$ and $v_{\mathrm{r},i}$ be the left and right eigenvectors of L corresponding to the eigenvalue λ_i, show that

   (i) the right eigenvectors of $\mathcal{L}$ corresponding to the eigenvalues $\eta_{i,\pm}$ are

$$\begin{bmatrix} v_{\mathrm{r},i} \\ \eta_{i,\pm} v_{\mathrm{r},i} \end{bmatrix},$$

(ii) for $k_{\mathrm{p}} > 0$, the left eigenvectors of $\mathcal{L}$ corresponding to the eigenvalues $\eta_{i,\pm}$ are

$$\begin{bmatrix} v_{l,i} \\ \dfrac{-\eta_{i,\pm}}{k_{\mathrm{p}} + \gamma_{\mathrm{p}}\lambda_i} v_{l,i} \end{bmatrix}.$$

E7.12 Laplacian oscillators. Given the Laplacian matrix $L = L^{\mathsf{T}} \in \mathbb{R}^{n \times n}$ of an undirected, weighted, and connected graph with edge weights a_{ij}, $i, j \in \{1, \dots, n\}$, define the *Laplacian oscillator flow* by

$$\ddot{x}(t) + Lx(t) = \mathbb{0}_n. \tag{E7.1}$$

This flow is written as first-order differential equation as

$$\begin{bmatrix} \dot{x}(t) \\ \dot{z}(t) \end{bmatrix} = \begin{bmatrix} \mathbb{0}_{n \times n} & I_n \\ -L & \mathbb{0}_{n \times n} \end{bmatrix} \begin{bmatrix} x(t) \\ z(t) \end{bmatrix} =: \mathcal{L} \begin{bmatrix} x(t) \\ z(t) \end{bmatrix}.$$

(i) Write the second-order Laplacian flow in components.

(ii) Write the characteristic polynomial of the matrix $\mathcal{L}$ using only the determinant of an $n \times n$ matrix.

(iii) Given the eigenvalues $\lambda_1 = 0, \lambda_2, \dots, \lambda_n$ of L, show that the eigenvalues $\eta_1, \dots, \eta_{2n}$ of $\mathcal{A}$ satisfy

$$\eta_1 = \eta_2 = 0, \qquad \eta_{2i,2i-1} = \pm\sqrt{\lambda_i}\mathrm{i}, \text{ for } i \in \{2, \dots, n\},$$

where i is the imaginary unit.

(iv) Show that the solution is the superposition of a ramp signal and of $n - 1$ harmonics, that is,

$$x(t) = \big(\operatorname{average}(x(0)) + \operatorname{average}(\dot{x}(0))t\big)\mathbb{1}_n + \sum_{i=2}^{n} a_i \sin(\sqrt{\lambda_i}t + \phi_i)v_i,$$

where $\{\mathbb{1}_n/\sqrt{n}, v_2, \dots, v_n\}$ are the orthonormal eigenvectors of L and where the amplitudes a_i and phases ϕ_i are determined by the initial conditions $\big(x(0), \dot{x}(0)\big)$.

E7.13 Delayed Laplacian flow. Define the *delayed Laplacian flow dynamics* over a connected, weighted, and graph G by:

$$\dot{x}_i(t) = \sum_{j \in \mathcal{N}} a_{ij}(x_j(t - \tau) - x_i(t - \tau)), \qquad i \in \{1, \dots, n\},$$

where $a_{ij} > 0$ is the weight on the edge $\{i, j\} \in E$, and $\tau > 0$ is a positive scalar delay term. The Laplace domain representation of the system is $X(s) = G(s)x(0)$ where $G(s)$ is associated transfer function

$$G(s) = (sI_n + e^{-s\tau}L)^{-1},$$

and $L = L^{\mathsf{T}} \in \mathbb{R}^{n \times n}$ is the network Laplacian matrix. Show that the transfer function $G(s)$ admits poles on the imaginary axis if the following *resonance* condition is true for an eigenvalue λ_i, $i \in \{1, \dots, n\}$, of the Laplacian matrix:

$$\tau = \frac{\pi}{2\lambda_i}.$$

E7.14 Robotic coordination and geometric optimization on the real line. Consider $n \geq 3$ robots with dynamics $\dot{p}_i = u_i$, where $i \in \{1, \dots, n\}$ is an index labeling each robot, $p_i \in \mathbb{R}$ is the position of robot i, and $u_i \in \mathbb{R}$ is a steering control input. For simplicity, assume that the robots are indexed according to their initial position: $p_1(0) \leq p_2(0) \leq \cdots \leq p_n(0)$. We consider the following distributed control laws to achieve some geometric configuration:

(i) *Move towards the centroid of your neighbors:* Each robot $i \in \{2, \ldots, n-1\}$ (having two neighbors) moves to the centroid of the local subset $\{p_{i-1}, p_i, p_{i+1}\}$:

$$\dot{p}_i = \frac{1}{3}(p_{i-1} + p_i + p_{i+1}) - p_i, \quad i \in \{2, \ldots, n-1\}. \tag{E7.2}$$

The robots $\{1, n\}$ (each having one neighbor) move to the centroid of the local subsets $\{p_1, p_2\}$ and $\{p_{n-1}, p_n\}$, respectively:

$$\dot{p}_1 = \frac{1}{2}(p_1 + p_2) - p_1 \quad \text{and} \quad \dot{p}_n = \frac{1}{2}(p_{n-1} + p_n) - p_n. \tag{E7.3}$$

Show that, by using the coordination laws (E7.2) and (E7.3), the robots asymptotically rendezvous.

(ii) *Move towards the centroid of your neighbors or walls:* Consider two walls at the positions $p_0 \leq p_1$ and $p_{n+1} \geq p_n$ so that all robots are contained between the walls. The walls are stationary, that is, $\dot{p}_0 = 0$ and $\dot{p}_{n+1} = 0$. Again, the robots $i \in \{2, \ldots, n-1\}$ (each having two neighbors) move to the centroid of the local subset $\{p_{i-1}, p_i, p_{i+1}\}$. The robots $\{1, n\}$ (each having one robotic neighbor and one neighboring wall) move to the centroid of the local subsets $\{p_0, p_1, p_2\}$ and $\{p_{n-1}, p_n, p_{n+1}\}$, respectively. Hence, the closed-loop robot dynamics are

$$\dot{p}_i = \frac{1}{3}(p_{i-1} + p_i + p_{i+1}) - p_i, \quad i \in \{1, \ldots, n\}. \tag{E7.4}$$

Show that, by using coordination law (E7.4), the robots become uniformly spaced on the interval $[p_0, p_{n+1}]$.

(iii) *Move away from the centroid of your neighbors or walls:* Again consider two stationary walls at $p_0 \leq p_1$ and $p_{n+1} \geq p_n$ containing the positions of all robots. We partition the interval $[p_0, p_{n+1}]$ into regions of interest, whereby each robot is assigned the territory containing all points closer to itself than to other robots. In other words, robot $i \in \{2, \ldots, n-1\}$ (having two neighbors) is assigned the region $\mathcal{V}_i = [(p_i + p_{i-1})/2, (p_{i+1} + p_i)/2]$, robot 1 is assigned the region $\mathcal{V}_1 = [p_0, (p_1 + p_2)/2]$, and robot n is assigned the region $\mathcal{V}_n = [(p_{n-1} + p_n)/2, p_{n+1}]$. We aim to design a distributed algorithm such that the robots are assigned asymptotically equal-sized regions. (This territory partition is called a *Voronoi partition*; see (Martínez et al., 2007) for further detail.) We consider the following simple coordination law, where each robot i heads for the midpoint $c_i(\mathcal{V}_i(p))$ of its partition $\mathcal{V}_i$:

$$\dot{p}_i = c_i(\mathcal{V}_i(p)) - p_i. \tag{E7.5}$$

Show that, by using the coordination law (E7.5), the robots' assigned regions asymptotically become equally large.

E7.15 **Properties of saddle points.** Prove Lemma 7.8.

E7.16 **Absence of imaginary eigenvalues in saddle matrices.** Prove Lemma 7.9.

E7.17 **Centralized formulation of sum-of-squares cost.** Consider a distributed optimization problem with n agents, where the cost function $f_i(x)$ of each agent $i \in \{1, \ldots, n\}$ is defined by $f_i(x) = P_i(x - x_i^*)^2$, where $P_i > 0$ and $x_i^* \in \mathbb{R}$. Consider the joint sum-of-squares cost function

$$f_{\text{sos}}(x) = \sum_{i=1}^{n} P_i(x - x_i^*)^2.$$

(i) Calculate the global minimizer x^* of $f_{\text{sos}}(x)$, and

(ii) show that the sum-of-squares cost $f_{\text{sos}}(x)$ is, up to a constant term, equal to the centralized cost function

$$f_{\text{centralized}}(x) = \left(\sum_{i=1}^{n} P_i\right)(x - x^*)^2.$$

E7.18 **Discrete saddle-point algorithm for distributed optimization.** Consider the centralized optimization problem

$$z^\star := \operatorname{argmin}_{z \in \mathbb{R}} \frac{1}{2} \sum_{i=1}^{n} p_i(z - r_i)^2, \tag{E7.6}$$

where $p_i > 0$ and $r_i \in \mathbb{R}$ are fixed scalar quantities for each $i \in \{1, \ldots, n\}$. Our aim is to solve this optimization problem in a distributed fashion, that is, distributing the computation among a group of n agents. Each agent i has access only to p_i and r_i and can communicate with the other agents via a network defined by the Laplacian matrix L. We assume that this network is undirected and connected.

(i) Show that solving the optimization problem (E7.6) is equivalent to solving

$$x^\star := \operatorname{argmin}_{x \in \mathbb{R}^n} \ \frac{1}{2} \sum_{i=1}^{n} p_i(x_i - r_i)^2, \tag{E7.7}$$
$$\text{subject to} \quad Lx = \mathbb{0}_n,$$

where $x = [x_1, \ldots, x_n]^\mathsf{T}$. In other words, show that $x^\star = z^\star \mathbb{1}_n$.

(ii) Write the KKT conditions of the optimization problem (E7.7), using the notation $P := \operatorname{diag}\{p_1, \ldots, p_n\} \in \mathbb{R}^{n \times n}$, $r = [r_1, \ldots, r_n]^\mathsf{T}$. Let $(\bar{x}, \bar{\lambda})$ be a solution of such KKT system. Show that a generic pair $(\tilde{x}, \tilde{\lambda})$ is a solution of the KKT system if and only if $\bar{x} = \tilde{x} = x^\star$ and $\tilde{\lambda} = \bar{\lambda} + \alpha \mathbb{1}_n$, for some $\alpha \in \mathbb{R}$.

Hint: *The solution requires knowledge of the Karush-Kuhn-Tucker (KKT) conditions for optimality; this is a classic topic in nonlinear constrained optimization discussed in numerous textbooks, e.g., in (Luenberger and Ye, 2008).*

(iii) Recall the definition of the saddle-point dynamics (7.16) and consider the discrete-time distributed saddle point algorithm

$$x_i(k+1) = x_i(k) - \tau\big(p_i(x_i(k) - r_i) + \sum\nolimits_{j \in \mathcal{N}^{\text{in}}(i)} L_{ji}\lambda_j(k)\big), \tag{E7.8a}$$

$$\lambda_i(k+1) = \lambda_i(k) + \tau\big(\sum\nolimits_{j \in \mathcal{N}^{\text{out}}(i)} L_{ij}x_j(k)\big), \tag{E7.8b}$$

where $\tau > 0$ is a sufficiently small step size. Show that, if the algorithm (E7.8) converges, then it converges to a solution of the optimization problem (E7.7).

(iv) Define the error vector by

$$e(k) := \begin{bmatrix} x(k) - \bar{x} \\ \lambda(k) - \bar{\lambda} \end{bmatrix}.$$

Find the error dynamics of the algorithm (E7.8), that is, the matrix G such that $e(k+1) = Ge(k)$.

(v) Show that, for $\tau > 0$ small enough, if μ is an eigenvalue of G, then either $\mu = 1$ or $|\mu| < 1$ and that $\begin{bmatrix} \mathbb{0}_n \\ \mathbb{1}_n \end{bmatrix}$ is, modulo rescaling, the only eigenvector relative to the eigenvalue $\mu = 1$. Use these results to study the convergence properties of the distributed algorithm (E7.8). Will $x(k) \to x^\star$ as $k \to \infty$?

Hint: *Use Lemma 7.8.*

E7.19 **Synchronization of inductor/capacitor storage circuits.** Consider a circuit composed of n identical resonant inductor/capacitor storage nodes (i.e., a parallel interconnection of a capacitor and an inductor) coupled through a connected and undirected graph whose edges are identical resistors; see Figure E7.2. The parameters ℓ, c, r take identical values on each inductor, capacitor and resistors, respectively.

(i) Write a state-space model of the resistively-coupled inductor/capacitor storage nodes in terms of the time constant $\tau = 1/rc$, the resonant frequency $\omega_0 = 1/\sqrt{\ell c}$, and the unweighted Laplacian matrix L of the resistive network.
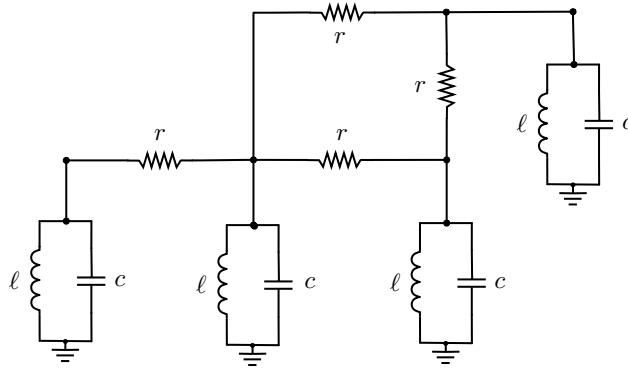
Figure E7.2: A circuit of identical inductor/capacitor storage nodes coupled through identical resistors.

   (ii) Characterize the asymptotic behavior of this system.

E7.20 **Formation control and affine Laplacian flow.** Consider a group of n vehicles moving in the plane. Each vehicle $i \in \{1, \dots, n\}$ is described by its kinematics $\dot{x}_i = u_i$, where $x_i \in \mathbb{C}$ is the vehicle's position in the complex plane and $u_i \in \mathbb{C}$ is a steering command. The vehicle initial position in the complex plane is a square formation: $x(0) = \begin{bmatrix} 1 & \mathrm{i} & -1 & -\mathrm{i} \end{bmatrix}^\top$, where i is the imaginary unit. We aim to move the vehicles to the final configuration:

$$\lim_{t \to \infty} x(t) = x_{\text{final}} = \begin{bmatrix} 0.5 + 0.5\mathrm{i} & -0.5 + 0.5\mathrm{i} & -0.5 - 0.5\mathrm{i} & 0.5 - 0.5\mathrm{i} \end{bmatrix}^\top. \tag{E7.9}$$

To achieve this goal, you will investigate a class of distributed control laws described by the *affine Laplacian flow*

$$\dot{x}(t) = -L(\alpha x(t) + \beta), \tag{E7.10}$$

where $\alpha > 0$ is a constant scalar gain, $\beta \in \mathbb{C}^n$ is a constant vector offset, and L is a Laplacian matrix of a strongly connected and weight-balanced digraph. Your tasks are the following:

   (i) Show that the affine Laplacian flow (E7.10) converges for any choice of $\alpha > 0$ and $\beta \in \mathbb{C}^n$.

  (ii) Characterize all the values for $\alpha > 0$ and $\beta \in \mathbb{C}^n$ such that the desired final configuration x_{final} is achieved by the affine Laplacian flow (E7.10).

E7.21 **Synchronization and stabilization of diffusively-coupled identical linear systems.** In this exercise we study a generalization and variation of the first and second-order Laplacian flows. We start by recalling that the *Kronecker product* of $A \in \mathbb{R}^{n \times m}$ and $B \in \mathbb{R}^{q \times r}$ is the $nq \times mr$ matrix $A \otimes B$ given by

$$A \otimes B = \begin{bmatrix} a_{11}B & \dots & a_{1m}B \\ \vdots & \ddots & \vdots \\ a_{n1}B & \ddots & a_{nm}B \end{bmatrix}. \tag{E7.11}$$

Horn and Johnson (1994) reviews many useful properties of the Kronecker product, including, for example

$$A \otimes B + A \otimes C = A \otimes (B + C), \tag{E7.12}$$

$$(A \otimes B)(C \otimes D) = (AC) \otimes (BD). \tag{E7.13}$$

Now, consider n identical linear single-input single-output dynamical systems

$$\dot{x}_i(t) = Ax_i(t) + Bu_i(t),$$
$$y_i(t) = Cx_i(t),$$

with $x_i : \mathbb{R}_{\geq 0} \to \mathbb{R}^k$, $A \in \mathbb{R}^{k \times k}$, $B \in \mathbb{R}^{k \times 1}$, and $C \in \mathbb{R}^{1 \times k}$. Given an undirected symmetric graph with Laplacian L with eigenvalues $0 = \lambda_1 \leq \cdots \leq \lambda_n$, consider the *diffusive coupling law*

$$\mathbf{u}(t) = -L\mathbf{y}(t),$$

where $\mathbf{x} = \begin{bmatrix} x_1^\mathsf{T}, \ldots, x_n^\mathsf{T} \end{bmatrix}^\mathsf{T} \in \mathbb{R}^{nk}$, $\mathbf{u} = \begin{bmatrix} u_1, \ldots, u_n \end{bmatrix}^\mathsf{T} \in \mathbb{R}^n$, and $\mathbf{y} = \begin{bmatrix} y_1, \ldots, y_n \end{bmatrix}^\mathsf{T} \in \mathbb{R}^n$. Show that

(i) the closed-loop system obeys

$$\dot{\mathbf{x}} = (I_n \otimes A - L \otimes BC)\mathbf{x}; \tag{E7.14}$$

(ii) if each matrix $A - \lambda_i BC$, $i \in \{2, \ldots, n\}$, is Hurwitz, then the linear systems become asymptotically synchronized in the sense that, for each $i, j \in \{1, \ldots, n\}$,

$$\lim_{t \to \infty} x_i(t) - x_j(t) = 0,$$

$$x_i(t) = \mathrm{e}^{At}\left(\frac{1}{n}\sum_{j=1}^{n} x_j(0)\right) + h_i(t), \qquad \text{where } \lim_{t \to \infty} h_i(t) = \mathbb{0}_k,$$

$$\lim_{t \to \infty} L\mathbf{y}(t) = \mathbb{0}_n;$$

(iii) the system (E7.14) is exponentially stable if and only each matrix $A - \lambda_i BC$, $i \in \{1, \ldots, n\}$, is Hurwitz;

(iv) given $k_\mathrm{p}, k_\mathrm{d} \geq 0$ and $\gamma_\mathrm{p}, \gamma_\mathrm{d} \geq 0$, assume $A = \begin{bmatrix} 0 & 1 \\ -k_\mathrm{p} & -k_\mathrm{d} \end{bmatrix}$, $B = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, and $C = \begin{bmatrix} \gamma_\mathrm{p} & \gamma_\mathrm{d} \end{bmatrix}$. Show that

    a) equation (E7.14) is the same as equation (7.9),
    b) if $\lambda_2 > 0$, $k_\mathrm{p} + \gamma_\mathrm{p} > 0$ and $k_\mathrm{d} + \gamma_\mathrm{d} > 0$, then each matrix $A - \lambda_i BC$, $i \in \{2, \ldots, n\}$, is Hurwitz.

Note: *The result in this exercise is a special case of (Xia and Scardovi, 2016, Theorem 1).*

The Incidence Matrix and its Applications

After studying adjacency and Laplacian matrices, in this chapter we introduce one final matrix associated with a graph: the incidence matrix. We study the properties of incidence matrices and their application to a class of estimation problems with relative measurements and to the study of cycles and cutset spaces. For simplicity we restrict our attention to undirected graphs.

8.1 The incidence matrix

Let G be an undirected unweighted graph with n nodes and m edges (and no self-loops, as by convention). Assign to each edge of G a unique identifier $e \in \{1, \ldots, m\}$ and an arbitrary direction. The (oriented) *incidence matrix* $B \in \mathbb{R}^{n \times m}$ of the graph G is defined component-wise by

$$B_{ie} = \begin{cases} +1, & \text{if node } i \text{ is the source node of edge } e, \\ -1, & \text{if node } i \text{ is the sink node of edge } e, \\ 0, & \text{otherwise.} \end{cases} \tag{8.1}$$

Here, we adopt the convention that an edge (i, j) has the source i and the sink j.

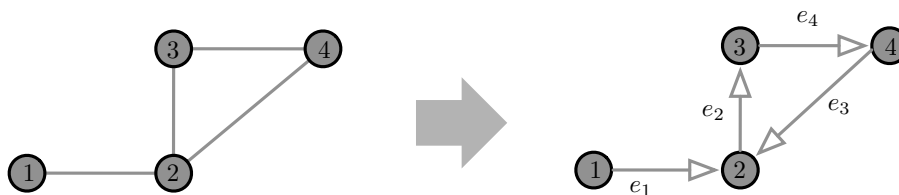It is useful to consider the following example graph, as depicted in figure.

Figure 8.1: How to number and orient the edges of a graph

As depicted on the right, we add an orientation to all edges, we order them and label them as follows:

$e_1 = (1, 2)$, $e_2 = (2, 3)$, $e_3 = (4, 2)$, and $e_4 = (3, 4)$. Accordingly, the incidence matrix is

$$B = \begin{bmatrix} +1 & 0 & 0 & 0 \\ -1 & +1 & -1 & 0 \\ 0 & -1 & 0 & +1 \\ 0 & 0 & +1 & -1 \end{bmatrix}.$$

Note: $\mathbb{1}_n^{\mathsf{T}} B = \mathbb{0}_m^{\mathsf{T}}$ since each column of B contains precisely one element equal to $+1$, one element equal to -1 and all other zeros.

Note: assume the edge $e \in \{1, \dots, m\}$ is oriented from i to j, then for any $x \in \mathbb{R}^n$,

$$(B^{\mathsf{T}} x)_e = x_i - x_j.$$

8.2   Properties of the incidence matrix

Given an undirected weighted graph G with edge set $\{1, \dots, m\}$ and adjacency matrix A, recall

$$L = D - A, \qquad \text{where } D \text{ is the degree matrix.}$$

Lemma 8.1 (From the incidence to the Laplacian matrix). *Let G be an undirected graph with n nodes, m edges, and incidence matrix B. If $\mathrm{diag}(\{a_e\}_{e \in \{1,\dots,m\}})$ is the diagonal matrix of edge weights, then*

$$L = B \, \mathrm{diag}(\{a_e\}_{e \in \{1,\dots,m\}}) B^{\mathsf{T}}.$$

Note: In the right-hand side, the matrix dimensions are $(n \times m) \times (m \times m) \times (m \times n) = n \times n$. Also note that, while the incidence matrix B depends upon the selected direction and numbering of each edge, the Laplacian matrix is independent of that.

Proof. Recall that, for matrices O, P and Q of appropriate dimensions, we have $(OPQ)_{ij} = \sum_{k,h} O_{ik} P_{kh} Q_{hj}$. Moreover, if the matrix P is diagonal, then $(OPQ)_{ij} = \sum_k O_{ik} P_{kk} Q_{kj}$.

For $i \neq j$, we compute

$$\begin{aligned} (B \, \mathrm{diag}(\{a_e\}_{e \in \{1,\dots,m\}}) B^{\mathsf{T}})_{ij} &= \sum_{e=1}^{m} B_{ie} a_e (B^{\mathsf{T}})_{ej} \\ &= \sum_{e=1}^{m} B_{ie} B_{je} a_e \qquad (e\text{-th term} = 0 \text{ unless } e \text{ is oriented } \{i, j\}) \\ &= (+1) \cdot (-1) \cdot a_{ij} = \ell_{ij}, \end{aligned}$$

where $L = \{\ell_{ij}\}_{i,j \in \{1,\dots,n\}}$, and along the diagonal of B we compute

$$(B \, \mathrm{diag}(\{a_e\}_{e \in \{1,\dots,m\}}) B^{\mathsf{T}})_{ii} = \sum_{e=1}^{m} B_{ie}^2 a_e = \sum_{e=1,\; e=(i,*) \text{ or } e=(*,i)}^{m} a_e = \sum_{j=1, j \neq i}^{n} a_{ij},$$

where, in the last equality, we counted each edge precisely once and we noted that self-loops are not allowed. ∎

Lemma 8.2 (Rank of the incidence matrix). *Let G be an undirected graph with n nodes, m edges, and incidence matrix B. Let d be the number of connected components of G. Then*

$$\mathrm{rank}(B) = n - d.$$

Proof. We prove this result for a connected graph with $d = 1$, but the proof strategy easily extends to $d > 1$. Recall that the rank of the Laplacian matrix L equals $n - d = n - 1$. Since the Laplacian matrix can be factorized as $L = B \, \mathrm{diag}(\{a_e\}_{e \in \{1,\dots,m\}}) B^{\mathsf{T}}$, where $\mathrm{diag}(\{a_e\}_{e \in \{1,\dots,m\}})$ has full rank m (and $m \geq n - 1$ due to connectivity), we have that necessarily $\mathrm{rank}(B) \geq n - 1$. On the other hand $\mathrm{rank}(B) \leq n - 1$ since $B^{\mathsf{T}} \mathbb{1}_n = \mathbb{0}_n$. It follows that B has rank $n - 1$. ∎

The factorization of the Laplacian matrix as $L = B \, \mathrm{diag}(\{a_e\}_{e \in \{1,\dots,m\}}) B^{\mathsf{T}}$ plays an important role of relative sensing networks. For example, we can decompose, the Laplacian flow $\dot{x} = -Lx$ into

open-loop plant:	$\dot{x}_i = u_i$,	$i \in \{1, \dots, n\}$,	or	$\dot{x} = u$,
measurements:	$y_{ij} = x_i - x_j$,	$\{i, j\} \in E$,	or	$y = B^{\mathsf{T}} x$,
control gains:	$z_{ij} = a_{ij} y_{ij}$,	$\{i, j\} \in E$,	or	$z = \mathrm{diag}(\{a_e\}_{e \in \{1,\dots,m\}}) y$,
control inputs:	$u_i = -\sum\limits_{\{i,j\} \in E} z_{ij}$,	$i \in \{1, \dots, n\}$,	or	$u = -Bz$.

Indeed, this control structure, illustrated as a block-diagram in Figure 8.2, is required to implement flocking-type behavior as in Example 7.1.2. The control structure in Figure 8.2 has emerged as a canonical control structure in many relative sensing and flow network problems also for more complicated open-loop dynamics and possibly nonlinear control gains; e.g., see (Bai et al., 2011).
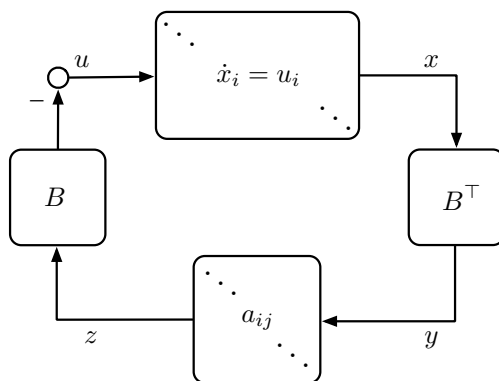

Figure 8.2: Illustration of the canonical control structure for a relative sensing network.

8.3 Cuts and cycles

Given an undirected unweighted graph with n nodes and m edges, its oriented incidence matrix naturally defines two useful vector subspaces of $\mathbb{R}^m$. With the customary convention to refer to $\mathbb{R}^m$ as the *edge space*,

the incidence matrix induces a direct sum decomposition of the edge space based on the concepts of cycles and graph cuts. We illustrate these concepts in what follows.

We start with some simple preliminary definitions. A *cut* χ of the graph G is a strict non-empty subset of the nodes $\{1, \ldots, n\}$. A cut and its complement χ^{c} define a *partition* $\{\chi, \chi^{\mathsf{c}}\}$ of $\{1, \ldots, n\}$ in the sense that $\chi \neq \emptyset$, $\chi^{\mathsf{c}} \neq \emptyset$, $\chi \cap \chi^{\mathsf{c}} = \emptyset$, and $\{1, \ldots, n\} = \chi \cup \chi^{\mathsf{c}}$. Given a cut χ, the set of edges that have one endpoint in each subset of the partition is called the *cutset* of χ.

We are now ready to introduce the main concepts of this section.

Definition 8.3 (Cutset orientation vectors and cutset space). *Let G be an undirected graph with n nodes, m edges, and with an arbitrary enumeration and orientation of its.*

(i) *Given a cut $\chi \subset \{1, \ldots, n\}$ of G, the* cutset orientation vector $v_\chi \in \{-1, 0, +1\}^m$ *of χ is defined component-wise, for each edge $e \in \{1, \ldots, m\}$,*

$$(v_\chi)_e = \begin{cases} +1, & \textit{if } e \text{ has its source in } \chi \text{ and sink in } \chi^{\mathsf{c}}, \\ -1, & \textit{if } e \text{ has its source in } \chi^{\mathsf{c}} \text{ and sink in } \chi, \\ 0, & \textit{otherwise.} \end{cases}$$

Here the source (resp. sink) of a directed edge (i, j) is the node i (resp j).

(ii) *The* cutset space *of G is subspace of $\mathbb{R}^m$ spanned by the cutset orientation vectors corresponding to all cuts of the nodes of G, that is, $\operatorname{span}\{v_\chi \in \mathbb{R}^m \mid \chi \text{ is a cut of } G\}$.*

Definition 8.4 (Signed path vectors and cycle space). *Let G be an undirected graph with n nodes, m edges, and with an arbitrary enumeration and orientation of its edges. Let γ be a simple undirected path in G.*

(i) *The* signed path vector $w_\gamma \in \{-1, 0, +1\}^m$ *of γ is defined component-wise, for each edge $e \in \{1, \ldots, m\}$,*

$$(w_\gamma)_e = \begin{cases} +1, & \textit{if } e \text{ is traversed positively by } \gamma, \\ -1, & \textit{if } e \text{ is traversed negatively by } \gamma, \\ 0, & \textit{otherwise.} \end{cases}$$

(ii) *The* cycle space *of G is the subspace of $\mathbb{R}^m$ spanned by the signed path vectors corresponding to all simple undirected cycles in G, that is, $\operatorname{span}\{w_\gamma \in \mathbb{R}^m \mid \gamma \text{ is a simple cycle in } G\}$.*

Next, we illustrate the notions of cutset orientation vector and signed path vector (for a path that is a simple cycle) for a low-dimensional example. Figure 8.3 shows three possible cuts and the only possible cycle in a cycle digraph with 3 nodes. It is a simple exercise to write

$$v_{\{1\}} = \begin{bmatrix} +1 \\ +1 \\ 0 \end{bmatrix}, v_{\{2\}} = \begin{bmatrix} -1 \\ 0 \\ +1 \end{bmatrix}, v_{\{3\}} = \begin{bmatrix} 0 \\ -1 \\ -1 \end{bmatrix}, \qquad w_\gamma = \begin{bmatrix} +1 \\ -1 \\ +1 \end{bmatrix},$$

and to verify that $v_{\{1\}} + v_{\{2\}} + v_{\{3\}} = \mathbb{0}_3$, that $\operatorname{span}\{v_{\{1\}}, v_{\{2\}}, v_{\{3\}}\} \perp \operatorname{span}\{w_\gamma\}$, and that $\{v_{\{1\}}, v_{\{2\}}, v_{\{3\}}\}$ are the rows of

$$B = \begin{bmatrix} +1 & +1 & 0 \\ -1 & 0 & +1 \\ 0 & -1 & -1 \end{bmatrix}.$$
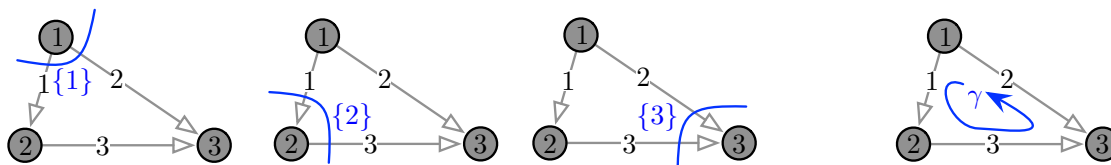
Figure 8.3: Three possible cuts and the only simple cycle in a cycle digraph.

With these conventions we are now in a position to state the main result of this section.

Theorem 8.5 (Cycle and cutset spaces). *Let G be an undirected graph with n nodes, m edges, and incidence matrix B. The following statements hold:*

(i) *the cycle space is $\mathrm{kernel}(B)$ and has dimension $m - n + 1$,*

(ii) *the cutset space is $\mathrm{image}(B^\mathsf{T})$ and has dimension $n - 1$, and*

(iii) *$\mathrm{kernel}(B) \perp \mathrm{image}(B^\mathsf{T})$ and $\mathrm{kernel}(B) \oplus \mathrm{image}(B^\mathsf{T}) = \mathbb{R}^m$.*

Statement (iii) is known as a statement in the fundamental theorem of linear algebra (Strang, 1993).

Proof of Theorem 8.5. The proof of statement (i) is given in Exercise E8.2.

Statement (ii) is proved as follows. For a cut χ, let $\mathbb{e}_\chi \in \{0,1\}^n$ be the *cut indicator vector* defined by $(\mathbb{e}_\chi)_i = 1$ if $i \in \chi$ and zero otherwise. Then, using the definitions, the cutset orientation vector for the cut χ is

$$v_\chi = B^\mathsf{T} \mathbb{e}_\chi.$$

This equality implies that $v_\chi \in \mathrm{image}(B^\mathsf{T})$ for all χ. Next, note that there are $n - 1$ independent cutset orientation vectors corresponding to the cuts $\{\{i\} \mid i \in \{1, \dots, n-1\}\}$. Hence these $n-1$ vectors are a basis of $\mathrm{image}(B^\mathsf{T})$ and the statement is established.

Finally, statement (iii) is proved in two steps. First, for any subspace $V \subset \mathbb{R}^m$, we have the direct sum decomposition of orthogonal subspaces $V \oplus V^\perp = \mathbb{R}^m$. Second, for any matrix B,

$$w \in \mathrm{kernel}(B) \iff \forall v \in \mathbb{R}^m \quad (Bw)^\mathsf{T} v = 0 \iff \forall v \in \mathbb{R}^m \quad w^\mathsf{T}(B^\mathsf{T} v) = 0 \iff w \in (\mathrm{image}(B^\mathsf{T}))^\perp.$$

Hence, we know $\mathrm{kernel}(B) = (\mathrm{image}(B^\mathsf{T}))^\perp$ and the statement follows. ∎

From the proof of the previous theorem and a bit more work, one can state the following result.

Lemma 8.6 (Bases for the cutset space and the cycle space). *Let G be a connected unweighted undirected graph with nodes $\{1, \dots, n\}$ and m edges.*

(i) *For each node $i \in \{1, \dots, n-1\}$, let $v_{\{i\}} \in \{-1, 0, +1\}^m$ denote the cutset orientation vector for the cut $\{i\}$, that is, let $v_{\{i\}}$ be the transpose of the i-th row of B. Then $\{v_{\{1\}}, \dots, v_{\{n-1\}}\}$ is a basis of the cutset space $\mathrm{image}(B^\mathsf{T})$.*

(ii) *Given a spanning tree T of G, for each edge $e \in V \setminus T$, define the fundamental cycle associated to T and e, denoted by $\gamma_{T,e}$, to be the cycle consisting of e and the path on T connecting the endpoints of e. Let $w_{T,e}$ be the associated signed path vector. Then*

    a) the fundamental cycle of each edge $e \in V \setminus T$ exists unique and is simple, and

    b) the set of signed path vectors $\{w_{T,e} \mid e \in V \setminus T\}$ is a basis of the cycle space $\mathrm{kernel}(B)$.

We illustrate this lemma with the graph and the incidence matrix given in Figure 8.4.

$$B = \begin{bmatrix} +1 & +1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & +1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & +1 & +1 & +1 & 0 \\ 0 & 0 & -1 & -1 & 0 & 0 & +1 \\ 0 & 0 & 0 & 0 & -1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 \end{bmatrix}.$$
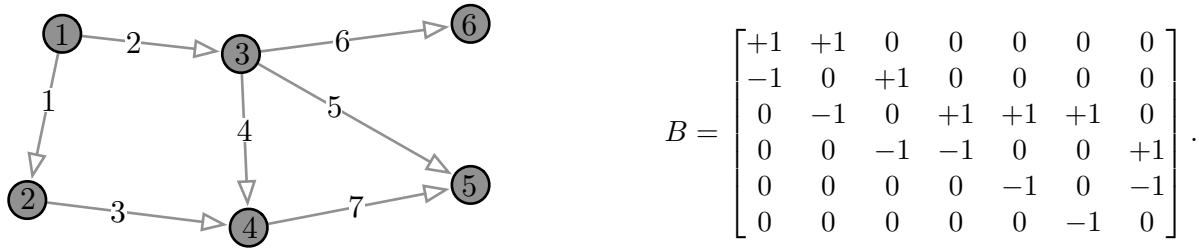
Figure 8.4: An undirected graph with arbitrary edge orientation and its associated incidence matrix $B \in \mathbb{R}^{6 \times 7}$.

Regarding a basis for the cutset space $\mathrm{image}(B^{\mathsf{T}})$, it is immediate to state that (the transpose of) any 5 of the 6 rows of B form a basis of $\mathrm{image}(B^{\mathsf{T}})$. Indeed, since $\mathrm{rank}(B) = n - 1$, any $n - 1$ columns of the matrix B^{T} form a basis for the cutset space. Figure 8.5 illustrates the 5 cuts and a corresponding basis for the cutset space.
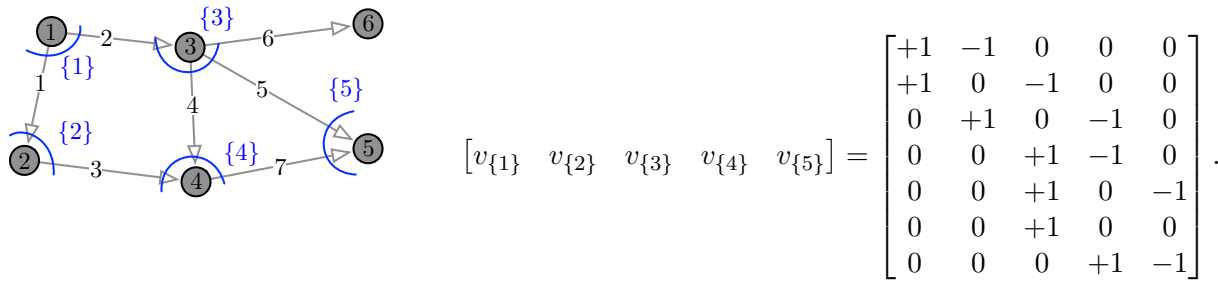

$$\begin{bmatrix} v_{\{1\}} & v_{\{2\}} & v_{\{3\}} & v_{\{4\}} & v_{\{5\}} \end{bmatrix} = \begin{bmatrix} +1 & -1 & 0 & 0 & 0 \\ +1 & 0 & -1 & 0 & 0 \\ 0 & +1 & 0 & -1 & 0 \\ 0 & 0 & +1 & -1 & 0 \\ 0 & 0 & +1 & 0 & -1 \\ 0 & 0 & +1 & 0 & 0 \\ 0 & 0 & 0 & +1 & -1 \end{bmatrix}.$$

Figure 8.5: Five cuts, corresponding to the first 5 nodes, and their cutset orientation vectors generating $\mathrm{image}(B^{\mathsf{T}})$.

In the proof of Theorem 8.5, we also stated that, for a cut χ, $\mathbb{e}_\chi \in \{0,1\}^n$ is the *cut indicator vector* defined by $(\mathbb{e}_\chi)_i = 1$ if $i \in \chi$ and zero otherwise, and that the cutset orientation vector for χ is given by

$$v_\chi = B^{\mathsf{T}} \mathbb{e}_\chi. \tag{8.2}$$

Indeed, one can show the following statement for the example in Figure 8.5: the cut separating nodes $\{1, 2, 3\}$ from $\{4, 5, 6\}$ has cut indicator vector $\begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \end{bmatrix}^{\mathsf{T}}$ and cutset vector $v_{\{1\}} + v_{\{2\}} + v_{\{3\}}$ is equal to the sum of the first three columns of B^{T}.

Next, regarding a basis for the cutset space $\mathrm{image}(B^{\mathsf{T}})$, the spanning tree T composed of the edges $\{1, 2, 4, 5, 6\}$ and the two fundamental cycles associated to edges 3 and 7 are illustrated in Figure 8.6.
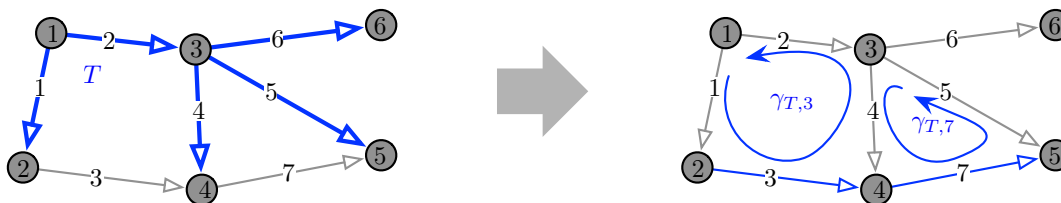
Figure 8.6: A graph with 6 nodes, 7 edges, and hence 2 independent cycles. A spanning tree T and the two resulting fundamental cycles.

The corresponding signed path vectors are

$$w_{T,3} = \begin{bmatrix} +1 \\ -1 \\ +1 \\ -1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, w_{T,7} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ +1 \\ -1 \\ 0 \\ +1 \end{bmatrix}, \quad \text{and} \quad \text{kernel}(B) = \text{span}\{w_{T,3}, w_{T,7}\}.$$

Note that the cycle traversing the edges $(1, 3, 7, 5, 2)$ in counter-clockwise orientation has a signed path vector given by the linear combination $w_{T,3} + w_{T,7}$.

Example 8.7 (Kirchhoff's and Ohm's laws revisited). In the following, we revisit the electrical resistor network from Section 6.1.2, and re-derive its governing equations via the incidence matrix; we refer to (Dörfler et al., 2017) for a more detailed treatment. Recall that with each node $i \in \{1, \dots, n\}$ of the network, we associate an external current injection $c_{\text{injected at } i}$. With each edge $\{i, j\} \in E$ we associate a positive conductance (i.e., the inverse of the resistance) $a_{ij} > 0$ and (after introducing an arbitrary direction for each edge) a current flow $c_{i \to j}$ and a voltage drop u_{ij}.

Kirchhoff's voltage law states that the sum of all voltage drops around each cycle must be zero. In other words, for each cycle in the network, the corresponding signed path vector $w \in \{-1, 0, 1\}^m$ satisfies $w^\mathsf{T} u = 0$. Equivalently, by Theorem 8.5, there exists a vector $v \in \mathbb{R}^n$ such that $u = B^\mathsf{T} v$, where $B \in \mathbb{R}^{n \times m}$ is the incidence matrix of (oriented) network. In Chapter 6 we referred to v as the vector of nodal voltages or potentials.

Kirchhoff's current law states that the sum of all current injections at every node must be zero. In other words, for each node $i \in \{1, \dots, n\}$ in the network, we have that $c_{\text{injected at } i} = \sum_{j=1}^{n} c_{i \to j}$. Consider now the cut isolating node i and its corresponding cutset orientation vector given by the ith column b_i^T of B^T; see Figure 8.5. Then, we have that $c_{\text{injected at } i} = \sum_{j=1}^{n} c_{i \to j} = b_i^\mathsf{T} c$. Equivalently, we have that $c_{\text{injected}} = Bc$.

Finally, *Ohm's law* states that the current $c_{j \to i}$ and the voltage drop u_{ij} over a resistor with resistance $1/a_{ij}$ are related as $c_{j \to i} = a_{ij} u_{ij}$. By combining Kirchhoff's and Ohm's laws, we arrive at

$$c_{\text{injected}} = Bc = B \operatorname{diag}(a_{ij}) u = B \operatorname{diag}(a_{ij}) B^\mathsf{T} v = Lv,$$

where we used Lemma 8.1 to recover the conductance matrix L. •

8.4 Appendix: Distributed estimation from relative measurements

In Chapter 1 we considered estimation problems for wireless sensor networks in which each node measures a scalar "absolute" quantity (expressing some environmental variable such as temperature, vibrations, etc). In this section, we consider a second class of examples in which meaurements are "relative," i.e., pairs of nodes measure the difference between their corresponding variables. Estimation problems involving relative measurements are numerous. For example, imagine a group of robots (or sensors) where no robot can sense its position in an absolute reference frame, but a robot can measure other robot's relative positions by means of on-board sensors. Similar problems arise in study of clock synchronization in networks of processors.

8.4.1 Problem statement

The optimal estimation based on relative measurement problem is stated as follows. As illustrated in Figure 8.7, we are given an undirected graph $G = (\{1, \ldots, n\}, E)$ with the following properties. First, each node $i \in \{1, \ldots, n\}$ of the network is associated with an unknown scalar quantity x_i (the x-coordinate of node i in figure). Second, the m undirected edges are given an orientation and, for each edge $e = (i, j)$,
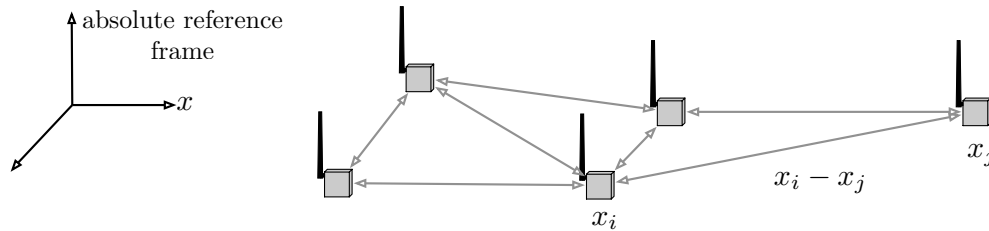

Figure 8.7: A wireless sensor network in which sensors can measure each other's relative distance and bearing. We assume that, for each link between node i and node j, the relative distance along the x-axis $x_i - x_j$ is available, where x_i is the x-coordinate of node i.

$e \in E$, the following scalar measurements are available:

$$y_{(i,j)} = x_i - x_j + v_{(i,j)} = (B^\mathsf{T} x)_e + v_{(i,j)},$$

where B is the graph incidence matrix and the measurement noises $v_{(i,j)}$, $(i, j) \in E$, are independent jointly-Gaussian variables with zero-mean $\mathbb{E}[v_{(i,j)}] = 0$ and variance $\mathbb{E}[v_{(i,j)}^2] = \sigma_{(i,j)}^2 > 0$. The joint matrix covariance is the diagonal matrix $\Sigma = \operatorname{diag}(\{\sigma_{(i,j)}^2\}_{(i,j)\in E}) \in \mathbb{R}^{m \times m}$. (For later use, it is convenient to define also $y_{(j,i)} = -y_{(j,i)} = x_j - x_i - v_{(i,j)}$.)

The optimal estimate $\widehat{x}^*$ of the unknown vector $x \in \mathbb{R}^n$ via the relative measurements $y \in \mathbb{R}^m$ is the solution to

$$\min_{\widehat{x}} \|B^\mathsf{T}\widehat{x} - y\|_{\Sigma^{-1}}^2.$$

Since no absolute information is available about x, we add the additional constraint that the optimal estimate should have zero mean and summarize this discussion as follows.

Definition 8.8 (Optimal estimation based on relative measurements). *Given an incidence matrix B, a set of relative measurements y with covariance Σ, find $\widehat{x}$ satisfying*

$$\min_{\widehat{x} \perp \mathbb{1}_n} \|B^\mathsf{T}\widehat{x} - y\|_{\Sigma^{-1}}^2. \tag{8.3}$$

8.4.2 Optimal estimation via centralized computation

From the theory of least square estimation, the optimal solution to problem 8.3 is obtained as by differentiating the quadratic cost function with respect to the unknown variable $\widehat{x}$ and setting the derivative to zero. Specifically:

$$0 = \frac{\partial}{\partial \widehat{x}} \|B^\mathsf{T}\widehat{x} - y\|_{\Sigma^{-1}}^2 = 2B\Sigma^{-1}B^\mathsf{T}\widehat{x}^* - 2B\Sigma^{-1}y.$$

The optimal solution is therefore obtained as the unique vector $\widehat{x}^* \in \mathbb{R}^n$ satisfying

$$\begin{aligned} B\Sigma^{-1}B^\mathsf{T}\widehat{x}^* = B\Sigma^{-1}y \quad &\Longleftrightarrow \quad L\widehat{x}^* = B\Sigma^{-1}y, \\ &\qquad\qquad \mathbb{1}_n^\mathsf{T}\widehat{x}^* = 0, \end{aligned} \tag{8.4}$$

where the Laplacian matrix L is defined by $L = B\Sigma^{-1}B^\mathsf{T}$. This matrix is the Laplacian for the weighted graph whose weights are the noise covariances associated to each relative measurement edge.

Before proceeding we review the definition and properties of the pseudoinverse Laplacian matrix given in Exercise E6.8. Recall that the *Moore-Penrose pseudoinverse* of an $n \times m$ matrix M is the unique $m \times n$ matrix $M^\dagger$ with the following properties:

(i) $MM^\dagger M = M$,

(ii) $M^\dagger M M^\dagger = M^\dagger$, and

(iii) $MM^\dagger$ is symmetric and $M^\dagger M$ is symmetric.

For our Laplacian matrix L, let $U \in \mathbb{R}^{n \times n}$ be an orthonormal matrix of eigenvectors of L. It is known that

$$L = U \begin{bmatrix} 0 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix} U^\mathsf{T} \implies L^\dagger = U \begin{bmatrix} 0 & 0 & \dots & 0 \\ 0 & 1/\lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1/\lambda_n \end{bmatrix} U^\mathsf{T}.$$

Moreover, it is known that $LL^\dagger = L^\dagger L = I_n - \frac{1}{n}\mathbb{1}_n\mathbb{1}_n^\mathsf{T}$ and $L^\dagger \mathbb{1}_n = \mathbb{0}_n$.

Lemma 8.9 (Unique optimal estimate). *If the undirected graph G is connected, then*

(i) there exists a unique solution to equations (8.4) solving the optimization problem in equation (8.3); and

(ii) this unique solution is given by

$$\widehat{x}^* = L^\dagger B\Sigma^{-1}y.$$

Proof. We claim there exists a unique solution to equation (8.4) and prove it as follows. Since G is connected, the rank of L is $n - 1$. Moreover, since L is symmetric and since $L\mathbb{1}_n = \mathbb{0}_n$, the image of L is the $(n - 1)$-dimensional vector subspace orthogonal to the subspace spanned by the vector $\mathbb{1}_n$. The vector $B\Sigma^{-1}y$ belongs to the image of L because the column-sums of B are zero, that is, $\mathbb{1}_n^{\mathsf{T}}B = \mathbb{0}_n^{\mathsf{T}}$, so that $\mathbb{1}_n^{\mathsf{T}}B\Sigma^{-1}y = \mathbb{0}_n^{\mathsf{T}}$. Finally, the requirement that $\mathbb{1}_n^{\mathsf{T}}\widehat{x}^* = 0$ ensures $\widehat{x}^*$ is perpendicular to the kernel of L.

The expression $\widehat{x}^* = L^{\dagger}B\Sigma^{-1}y$ follows from left-multiplying left and right hand side of equation (8.4) by the pseudoinverse Laplacian matrix $L^{\dagger}$ and using the property $L^{\dagger}L = I_n - \frac{1}{n}\mathbb{1}_n\mathbb{1}_n^{\mathsf{T}}$. One can also verify that $\mathbb{1}_n^{\mathsf{T}}L^{\dagger}B\Sigma^{-1}y = 0$, because $L^{\dagger}\mathbb{1}_n = \mathbb{0}_n$. ∎

8.4.3   Optimal estimation via decentralized computation

To compute $\widehat{x}^*$ in a distributed way, we propose the following distributed algorithm. Pick a small $\alpha > 0$ and let each node implement the *affine averaging algorithm*:

$$\widehat{x}_i(k + 1) = \widehat{x}_i(k) - \alpha \sum_{j \in \mathcal{N}(i)} \frac{1}{\sigma_{(i,j)}^2}\Big(\widehat{x}_i(k) - \widehat{x}_j(k) - y_{(i,j)}\Big),$$

$$\widehat{x}_i(0) = 0. \tag{8.5}$$

There are two interpretations of this algorithm. First, note that the estimate at node i is adjusted at each iteration as a function of edge errors: each edge error (difference between estimated and measured edge difference) contributes to a weighted small correction in the node value. Second, note that the affine Laplacian flow

$$\dot{\widehat{x}} = -L\widehat{x} + B\Sigma^{-1}y \tag{8.6}$$

results in a steady-state satisfying $L\widehat{x} = B\Sigma^{-1}y$, which readily delivers the optimal estimate $\widehat{x}^* = L^{\dagger}B\Sigma^{-1}y$ for appropriately chosen initial conditions. The algorithm (8.5) results from an Euler discretization of the affine Laplacian flow (8.6) with step size α.

Lemma 8.10. *Given a graph G describing a relative measurement problem for the unknown variables $x \in \mathbb{R}^n$, with measurements $y \in \mathbb{R}^m$, and measurement covariance matrix $\Sigma = \mathrm{diag}(\{\sigma_{(i,j)}^2\}_{(i,j)\in E}) \in \mathbb{R}^{m \times m}$. The following statements hold:*

(i) the affine averaging algorithm can be written as

$$\widehat{x}(k + 1) = (I_n - \alpha L)\widehat{x}(k) + \alpha B\Sigma^{-1}y,$$

$$\widehat{x}(0) = \mathbb{0}_n. \tag{8.7}$$

(ii) if G is connected and if $\alpha < 1/d_{\max}$ where $d_{\max}$ is the maximum weighted out-degree of G, then the solution $k \mapsto \widehat{x}(k)$ of the affine averaging algorithm (8.5) converges to the unique solution $\widehat{x}^*$ of the optimization problem 8.3.

Proof. To show fact (i), note that the algorithm can be written in vector form as

$$\widehat{x}(k + 1) = \widehat{x}(k) - \alpha B\Sigma^{-1}(B^{\mathsf{T}}\widehat{x}(k) - y),$$

and, using $L = B\Sigma^{-1}B^{\mathsf{T}}$, as equation (8.7).

To show fact (ii), define the error signal $\eta(k) = \widehat{x}^* - \widehat{x}(k)$. Note that $\eta(0) = \widehat{x}^*$ and that $\mathrm{average}(\eta(0)) = 0$ because $\mathbb{1}_n^\mathsf{T}\widehat{x}^* = 0$. Compute

$$\begin{aligned}
\eta(k+1) &= (I_n - \alpha L + \alpha L)\widehat{x}^* - (I_n - \alpha L)\widehat{x}(k) - \alpha B\Sigma^{-1}y \\
&= (I_n - \alpha L)\eta(k) + \alpha(L\widehat{x}^* - B\Sigma^{-1}y) \\
&= (I_n - \alpha L)\eta(k).
\end{aligned}$$

Now, according to Exercise E7.6, α is sufficiently small so that $I_n - \alpha L$ is non-negative. Moreover, $(I_n - \alpha L)$ is doubly-stochastic and symmetric, and its corresponding undirected graph is connected and aperiodic. Therefore, Theorem 5.1 implies that $\eta(k) \to \mathrm{average}(\eta(0))\mathbb{1}_n = \mathbb{0}_n$. ∎

8.5 Historical notes and further reading

Standard references on incidence matrices include texts on algebraic graph theory such as (Biggs, 1994; Foulds, 1995; Godsil and Royle, 2001). An extensive discussion about algebraic potential theory on graphs is given by Biggs (1997).

The algorithm in Section 8.4.3 is taken from (Bolognani et al., 2010). For the notion of edge Laplacian and its properties, we refer to (Zelazo, 2009; Zelazo and Mesbahi, 2011; Zelazo et al., 2013). Additional references on distributed estimation for relative sensing networks include (Barooah and Hespanha, 2007, 2008; Bolognani et al., 2010; Piovan et al., 2013).

A recent survey on cycle bases, their rich structure, and related algorithms is given by Kavitha et al. (2009).

8.6   Exercises

E8.1   **Incidence matrix, cutset and cycle spaces for basic graphs.** Recall Example 4.1, and consider the following unweighted undirected graphs with node set $\{1, \ldots, 4\}$:

(i) the path graph P_4;

(ii) the cycle graph C_4;

(iii) the star graph S_4; and

(iv) the complete graph K_n.

For each graph, select an arbitrary orientation of the edges, compute the incidence matrix, compute a basis for the cutset space, and compute a basis for the cycle space.

E8.2   **Incidence matrix and signed path vectors.** Given an undirected graph G, consider an arbitrary orientation of its edges, its incidence matrix $B \in \mathbb{R}^{n \times m}$, and a simple path γ with distinct initial and final nodes described by a signed path vector $w^\gamma \in \mathbb{R}^m$.

(i) Show that the vector $y = Bw^\gamma \in \mathbb{R}^n$ has components

$$y_i = \begin{cases} +1, & \text{if node } i \text{ is the initial node of } \gamma, \\ -1, & \text{if node } i \text{ is the final node of } \gamma, \\ 0, & \text{otherwise.} \end{cases}$$

(ii) Prove statement (i) in Theorem 8.5.

E8.3   **Continuous distributed estimation from relative measurements.** Consider the continuous distributed estimation algorithm given by the affine Laplacian flow (8.6). Show that for an undirected and connected graph G and appropriately initial conditions $\hat{x}(0) = \mathbb{0}_n$, the affine Laplacian flow (8.6) converges to the unique solution $\hat{x}^*$ of the estimation problem given in Lemma 8.9.

E8.4   **The edge Laplacian matrix (Zelazo and Mesbahi, 2011).** For an unweighted undirected graph with n nodes and m edges, introduce an arbitrary orientation for the edges. Recall the notions of incidence matrix $B \in \mathbb{R}^{n \times m}$ and Laplacian matrix $L = BB^\mathsf{T} \in \mathbb{R}^{n \times n}$ and define the *edge Laplacian matrix* by

$$L_{\text{edge}} = B^\mathsf{T} B \in \mathbb{R}^{m \times m}.$$

(Note that, in general, the edge Laplacian matrix is not a Laplacian matrix.) Select an edge orientation and compute B, L and L_{edge} for

(i) a line graph with three nodes, and

(ii) for the graph with four nodes in Figure 8.1.

Show that, for an arbitrary undirected graph,

(iii) $\text{kernel}(L_{\text{edge}}) = \text{kernel}(B)$;

(iv) $\text{rank}(L) = \text{rank}(L_{\text{edge}})$;

(v) for an acyclic graph L_{edge} is nonsingular; and

(vi) the non-zero eigenvalues of L_{edge} are equal to the non-zero eigenvalues of L.

E8.5   **Evolution of the local disagreement error (Zelazo and Mesbahi, 2011).** Consider the Laplacian flow $\dot{x} = -Lx$, defined over an undirected and connected graph with n nodes and m edges. Beside the *absolute disagreement error* $\delta(t) = x(t) - \text{average}(x(t))\mathbb{1}_n \in \mathbb{R}^n$ considered thus far, we can also analyze the *relative disagreement error* $e_{ij}(t) = x_i(t) - x_j(t)$, for $\{i, j\} \in E$.

(i) Write a differential equation for the relative disagreement errors $t \mapsto e(t) \in \mathbb{R}^m$.

(ii) Based on Exercise E8.4, show that the relative disagreement errors converge to zero with exponential convergence rate given by the algebraic connectivity $\lambda_2(L)$.

E8.6 **Averaging with distributed integral control.** Consider a Laplacian flow implemented as a relative sensing network over a connected and undirected graph with n nodes, m edges, incidence matrix $B \in \mathbb{R}^{n \times m}$ and weights $a_{ij} > 0$ for $i, j \in \{1, \ldots, n\}$, and subject to a constant disturbance term $\eta \in \mathbb{R}^m$, as shown in Figure E8.1.
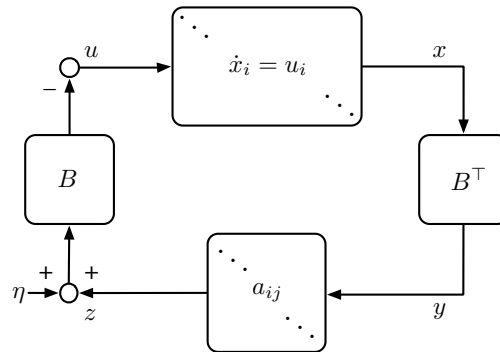
Figure E8.1: A relative sensing network with a constant disturbance input $\eta \in \mathbb{R}^m$.

(i) Derive the dynamic closed-loop equations describing the model in Figure E8.1.

(ii) Show that the state $x(t)$ converges asymptotically to some constant vector $x^* \in \mathbb{R}^n$ depending on the value of the disturbance η and that x^* is not necessarily a consensus state.

Consider the system in Figure E8.1 with a distributed integral controller forcing convergence to consensus, as shown in Figure E8.2. Recall that $\frac{1}{s}$ is the the Laplace symbol for the integrator.
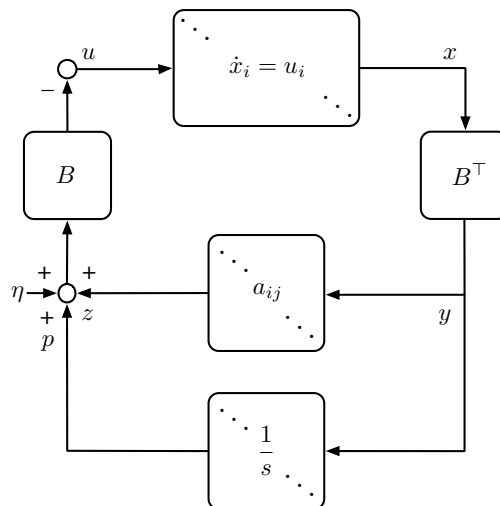

Figure E8.2: Relative sensing network with a disturbance $\eta \in \mathbb{R}^m$ and distributed integral action.

(iii) Derive the dynamic closed-loop equations describing the model in Figure E8.2.

(iv) Show that the distributed integral controller in Figure E8.2 asymptotically stabilizes the set of steady states (x^*, p^*), with $x^* \in \mathrm{span}\{\mathbb{1}_n\}$ corresponding to consensus.

Hint: To show stability, use Lemma 7.9.

E8.7 **Sensitivity of Laplacian eigenvalues.** Consider an unweighted undirected graph $G = (V, E)$ with incidence matrix $B \in \mathbb{R}^{n \times m}$, and Laplacian matrix $L = BB^\mathsf{T} \in \mathbb{R}^{n \times n}$. Define a undirected graph G' by adding one unweighted edge $e \notin E$ to G, that is, $G' = (V, E \cup e)$. Show that

$$\lambda_{\max}(L_G) \leq \lambda_{\max}(L_{G'}) \leq \lambda_{\max}(L_G) + 2.$$

Hint: You may want to take a detour via the edge Laplacian matrix $L_{\mathrm{edge}} = B^\mathsf{T} B \in \mathbb{R}^{m \times m}$ (see Exercise E8.4) and use the following fact (Horn and Johnson, 1985, Theorem 4.3.17): if A is a symmetric matrix with eigenvalues ordered as $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$, and B is a principal submatrix of A with eigenvalues ordered as $\mu_1 \leq \mu_2 \leq \ldots \leq \mu_{n-1}$, then the eigenvalues of A and B interlace, that is, $\lambda_1 \leq \mu_1 \leq \lambda_2 \leq \ldots \leq \mu_{n-1} \leq \lambda_n$.

E8.8 **The orthogonal projection onto the cutset space (Jafarpour and Bullo, 2017).** Recall the following well-known facts from linear algebra: a square matrix $P \in \mathbb{R}^{m \times m}$ is an *orthogonal projection* if $P = P^\mathsf{T}$ and $P^2 = P$; given a full-rank matrix $X \in \mathbb{R}^{m \times n}$, $n < m$, the matrix $P = X(X^\mathsf{T} X)^{-1} X^\mathsf{T}$ is the orthogonal projection onto the $\mathrm{image}(X)$. Prove that

(i) $\Pi_n = I_n - \mathbb{1}_n \mathbb{1}_n^\mathsf{T}/n$ is the orthogonal projection onto $\mathbb{1}^\perp$, and

(ii) if X is not full rank (i.e., it has a trivial kernel), the matrix $P = X(X^\mathsf{T} X)^\dagger X^\mathsf{T}$ is the orthogonal projection onto $\mathrm{image}(X)$, where $(X^\mathsf{T} X)^\dagger$ is the pseudoinverse of $X^\mathsf{T} X$.

Hint: Recall the defining properties of the pseudoinverse in Exercise E6.8.

Given an unweighted undirected graph with an oriented incidence matrix B, Laplacian matrix $L = BB^\mathsf{T}$, and pseudoinverse Laplacian matrix $L^\dagger$, recall that $\mathbb{R}^m = \mathrm{image}(B^\mathsf{T}) \oplus \mathrm{kernel}(B)$ is the orthogonal decomposition into cutset space and cycle space. Show that

(iii) $P = B^\mathsf{T} L^\dagger B$ is an orthogonal projection matrix, and

(iv) $P = B^\mathsf{T} L^\dagger B$ is the orthogonal projection onto the cutset space $\mathrm{image}(B^\mathsf{T})$.

Positive and Compartmental Systems

In this chapter we study various positive systems, that is, dynamical systems with state variables that are non-negative for all times. For simplicity we focus on continuous-time models, though a comparable theory exists for discrete-time systems. We are particularly interested in compartmental systems, that is, models of dynamical processes characterized by conservation laws (e.g., mass, fluid, energy) and by the flow of material between units known as compartments. Example compartmental systems are transportation networks, queueing networks, communication networks, epidemic propagation models in social contact networks, as well as ecological and biological networks. Linear compartmental and positive systems are described by so-called Metzler matrices; we define and study such matrices in this chapter.

9.1 Example systems

In this section we review some examples of compartmental systems.

Ecological and environmental systems   The flow of energy and nutrients (water, nitrates, phosphates, etc) in *ecosystems* is typically studied using compartmental modelling. For example, Figure 9.1 illustrates a widely-cited water flow model for a desert ecosystem (Noy-Meir, 1973). Other classic ecological network systems include models for dissolved oxygen in stream, nutrient flow in forest growth and biomass flow in fisheries (Walter and Contreras, 1999).
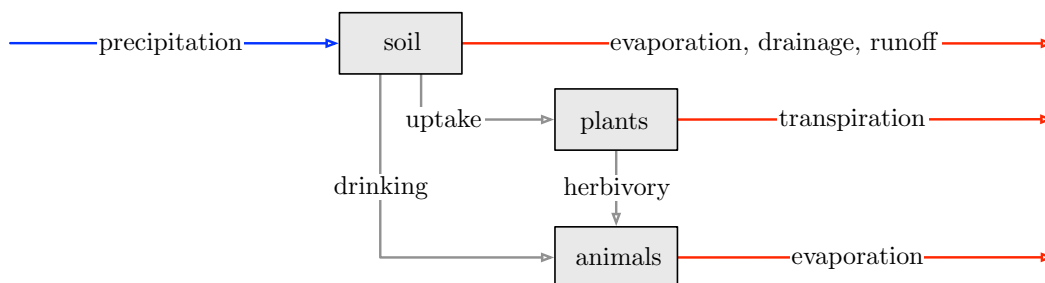

Figure 9.1: Water flow model for a desert ecosystem. The blue line denotes an inflow from the outside environment. The red lines denote outflows into the outside environment.

Epidemiology of infectious deseases   To study the *propagation of infectious deseases*, the population at risk is typically divided into compartments consisting of individdiuals who are susceptible (S), infected (I), and, possibly, recovered and no longer susceptible (R). As illustrated in Figure 9.2, the three basic epidemiological models are (Hethcote, 2000) called SI, SIS, SIR, depending upon how the desease spreads. A detailed discussion is postponed until Chapter 16.
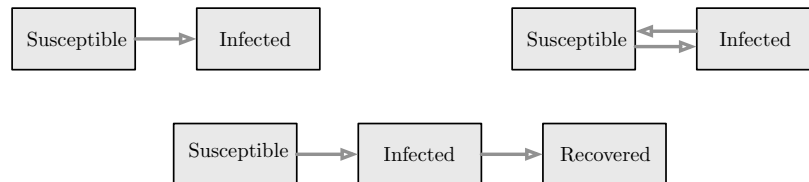

Figure 9.2: The three basic models SI, SIS and SIR for the propagation of an infectious desease

Drug and chemical kinetics in biomedical systems   Compartmental model are also widely adopted to characterize the kinetics of drugs and chemicals in biomedical systems. Here is a classic example (Charkes et al., 1978) from nuclear medicine: bone scintigraphy (also called bone scan) is a medical test in which the patient is injected with a small amount of radioactive material and then scanned with an appropriate radiation camera.
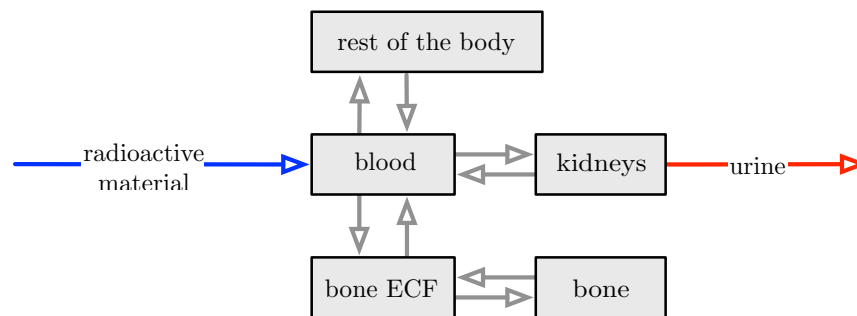

Figure 9.3: The kinetics of a radioactive isotope through the human body (ECF = extra-cellular fluid).

9.2   Positive systems and Metzler matrices

Motivated by the examples in the previous sections, we start our study by characterizing the class of positive systems.

Definition 9.1 (Positive systems). *A dynamical system $\dot{x}(t) = f(x(t), t)$, $x \in \mathbb{R}^n$, is* positive *if $x(0) \geq \mathbb{0}_n$ implies $x(t) \geq \mathbb{0}_n$ for all t.*

We are especially interested in linear and affine systems, described by

$$\dot{x}(t) = Ax(t), \qquad \text{and} \quad \dot{x}(t) = Ax(t) + b.$$

Note that the set of affine systems includes the set of linear systems (each linear system is affine with $b = \mathbb{0}_n$).

It is now convenient to introduce a second useful definition.

Definition 9.2 (Metzler matrix). *For a matrix $A \in \mathbb{R}^{n \times n}$, $n \geq 2$,*

   (i) *A is* Metzler *if all its off-diagonal elements are non-negative;*

  (ii) *if A is Metzler, its* associated digraph *is a weighted digraph defined as follows: $\{1, \dots, n\}$ are the nodes, there are no self-loops, (i, j), $i \neq j$ is an edge with weight a_{ij} if and only if $a_{ij} > 0$; and*

 (iii) *if A is Metzler, A is* irreducible *if its induced digraph is strongly connected.*

In other words, A is Metzler if and only if there exists a scalar $\alpha > 0$ such that $A + \alpha I_n$ is non-negative. For example, if G is a weighted digraph with Laplacian matrix L, then $-L$ is a Metzler matrix with zero row-sums.

Metzler matrices are sometimes also referred to as quasi-positive or essentially non-negative.

We are now ready to classify which affine systems are positive.

Theorem 9.3 (Positive affine systems and Metzler matrices). *For the affine system $\dot{x}(t) = Ax(t) + b$, the following statements are equivalent:*

  (i) *the system is positive, that is, $x(t) \geq \mathbb{0}_n$ for all t and all $x(0) \geq \mathbb{0}_n$,*

 (ii) *A is Metzler and $b \geq \mathbb{0}_n$.*

Proof. We start by showing that statement (i) implies statement (ii). If $x(0) = \mathbb{0}_n$, then $\dot{x}$ cannot have any negative components, hence $b \geq \mathbb{0}_n$. If any off-diagonal entry (i, j), $i \neq j$, of A is strictly negative, then consider an initial condition $x(0)$ with all zero entries except for $x(j) > b_i/|a_{ij}|$. It is easy to see that $\dot{x}_i(0) < 0$ which is a contradiction.

Next, we show that statement (ii) implies statement (i). It suffices to note that, anytime there exists i such that $x_i(t) = 0$, the conditions $x(t) \geq \mathbb{0}_n$, A Metzler and $b \geq \mathbb{0}_n$ together imply $\dot{x}_i(t) = \sum_{i \neq j} a_{ij}x_j(t) + b_i \geq 0$. ∎

This results motivates the importance of Metzler matrices. Therefore we now study their properties in two theorems. We start by writing a version of Perron-Frobenius Theorem 2.12 for non-negative matrices.

Theorem 9.4 (Perron-Frobenius Theorem for Metzler matrices). *If $A \in \mathbb{R}^{n \times n}$, $n \geq 2$, is Metzler, then*

  (i) *there exists a real eigenvalue λ such that $\lambda \geq \Re(\mu)$ for all other eigenvalues μ, and*

 (ii) *the right and left eigenvectors of λ can be selected non-negative.*

If additionally A is irreducible, then

 (iii) *there exists a real simple eigenvalue λ such that $\lambda \geq \Re(\mu)$ for all other eigenvalues μ, and*

 (iv) *the right and left eigenvectors of λ are unique and positive (up to rescaling).*

As for non-negative matrices, we refer to λ as to the dominant eigenvalue. We invite the reader to work out the details of the proof in Exercise E9.2. Next, we give necessary and sufficient conditions for the dominant eigenvalue of a Metzler matrix to be strictly negative.

Theorem 9.5 (Properties of Hurwitz Metzler matrices). *For a Metzler matrix A, the following statements are equivalent:*

   (i) *A is Hurwitz,*

  (ii) *A is invertible and $-A^{-1} \geq 0$, and*

 (iii) *for all $b \geq \mathbb{0}_n$, there exists $x^* \geq \mathbb{0}_n$ solving $Ax^* + b = \mathbb{0}_n$.*

Moreover, if A is Metzler, Hurwitz and irreducible, then $-A^{-1} > 0$.

Proof. We start by showing that (i) implies (ii). Clearly, if A is Hurwitz, then it is also invertible. So it suffices to show that $-A^{-1}$ is non-negative. Pick $\varepsilon > 0$ and define $\mathcal{A}_{\varepsilon,A} = I_n + \varepsilon A$, that is, $(-\varepsilon A) = (I_n - \mathcal{A}_{\varepsilon,A})$. Because A is Metzler, ε can be selected small enough so that $\mathcal{A}_{\varepsilon,A} \geq 0$. Moreover, because the spectrum of A is strictly in the left half plane, one can verify that, for ε small enough, $\mathrm{spec}(\varepsilon A)$ is inside the disk of unit radius centered at the point -1; as illustrated in Figure 9.4. In turn, this last property implies
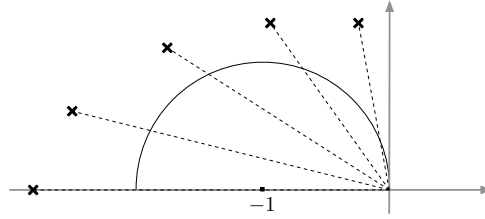

Figure 9.4: For any $\lambda \in \mathbb{C}$ with strictly negative real part, there exists ε such that the segment from the origin to $\varepsilon\lambda$ is inside the disk of unit radius centered at the point -1.

that $\mathrm{spec}(I_n + \varepsilon A)$ is strictly inside the disk of unit radius centered at the origin, that is, $\rho(\mathcal{A}_{\varepsilon,A}) < 1$. We now adopt the Neumann series as defined in Exercise E2.12: because $\rho(\mathcal{A}_{\varepsilon,A}) < 1$, we know that $(I_n - \mathcal{A}_{\varepsilon,A}) = (-\varepsilon A)$ is invertible and that

$$(-\varepsilon A)^{-1} = (I_n - \mathcal{A}_{\varepsilon,A})^{-1} = \sum_{k=0}^{\infty} \mathcal{A}_{\varepsilon,A}^k. \tag{9.1}$$

Note now that the right-hand side is non-negative because it is the sum of non-negative matrices. In summary, we have shown that A is invertible and that $-A^{-1} \geq 0$. This statement proves that (i) implies (ii).

Next we show that (ii) implies (i). We know A is Metzler, invertible and satisfies $-A^{-1} \geq 0$. By the Perron-Frobenius Theorem 9.4 for Metzler matrices, we know there exists $v \geq \mathbb{0}_n$, $v \neq \mathbb{0}_n$, satisfying $Av = \lambda_{\mathrm{Metzler}}v$, where $\lambda_{\mathrm{Metzler}} = \max\{\Re(\lambda) \mid \lambda \in \mathrm{spec}(A)\}$. Clearly, A invertible implies $\lambda_{\mathrm{Metzler}} \neq 0$ and, moreover, $v = \lambda_{\mathrm{Metzler}}A^{-1}v$. Now, we know v is non-negative and $A^{-1}v$ is non-positive. Hence, $\lambda_{\mathrm{Metzler}}$ must be negative and, in turn, A is Hurwitz. This statement establishes the equivalence between (ii) implies (i)

Finally, regarding the equivalence between statement (ii) and statement (iii), note that, if $-A^{-1} \geq 0$ and $b \geq \mathbb{0}_n$, then clearly $x^* = -A^{-1}b \geq \mathbb{0}_n$ solves $Ax^* + b = \mathbb{0}_n$. This proves that (ii) implies (iii). Vice versa, if statement (iii) holds, then let x_i^* be the non-negative solution of $Ax_i^* = -\mathbb{e}_i$ and let X be the non-negative matrix with columns $x_1^*, \ldots, x_n^*$. Therefore, we know $AX = -I_n$ so that A is invertible, $-X$ is its inverse, and $-A^{-1} = -(-X) = X$ is non-negative. This statement proves that (iii) implies (ii).

Finally, the statement that $-A^{-1} > 0$ for each Metzler, Hurwitz and irreducible matrix A is proved as follows. Because A is irreducible, the matrix $\mathcal{A}_{\varepsilon,A} = I_n + \varepsilon A$ is non-negative (for ε sufficiently small) and primitive. Therefore, the right-hand side of equation (9.1) is strictly positive. ∎

This theorem about Metzler matrices immediately leads to the following corollary about positive affine systems, which extends the results in Exercise E7.2.

Corollary 9.6 (Existence, positivity and stability of equilibria for positive affine systems). *Consider a continuous-time positive affine system $\dot{x} = Ax + b$, where A is Metzler and b is non-negative. If the matrix A is Hurwitz, then*

(i) *the system has a unique equilibrium point $x^* \in \mathbb{R}^n$, that is, a unique solution to $Ax^* + b = \mathbb{0}_n$,*

(ii) *the equilibrium point x^* is non-negative, and*

(iii) *all trajectories converge asymptotically to x^*.*

We will provide an extension of Theorem 9.5 after introducing Lyapunov theory in Chapter 14.

9.3 Compartmental systems

In this section, motivated by the examples in Section 9.1, we study an important class of positive affine systems.

A *compartmental system* is a dynamical system in which material is stored at individual locations and is transferred along the edges of directed graph, called the *compartmental digraph*; see Figure 9.5b. The
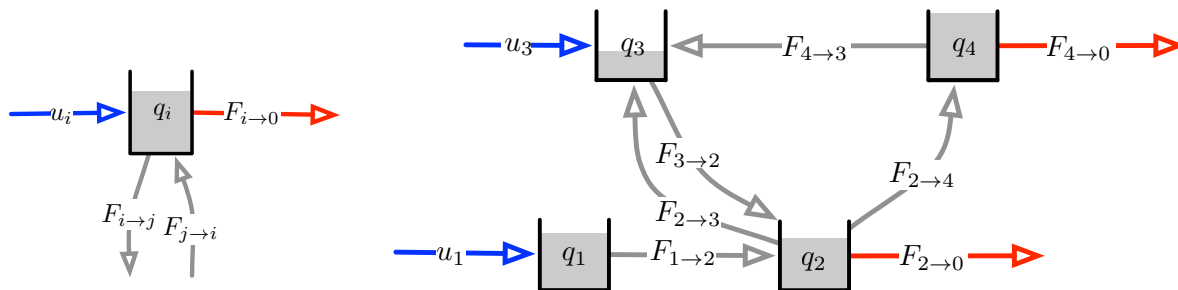

Figure 9.5: A compartment and a compartmental system

"storage" nodes are referred to as *compartments*; each compartment contains a time-varying quantity $q_i(t)$. Each directed arc (i, j) represents a *mass flow* (or *flux*), denoted $F_{i \to j}$, from compartment i to compartment

j. The compartmental system interacts with its surrounding environment via inputs and output flows, denoted in figure by blue and red arcs respectively: the inflow from the environment into compartment i is denoted by u_i and the outflow from compartment i into the environment is denoted by $F_{i \to 0}$.

In summary, a (nonlinear) compartmental system is described by a directed graph G_F, by maps $F_{i \to j}$ for all edges (i, j) of G_F, and by inflow and outflow maps. (The compartmental digraph has no self-loops.) The dynamic equations of the compartmental system are obtained by the *instantaneous flow balance* at each compartment. In other words, asking that the rate of accumulation at each compartment equals the net inflow rate we obtain:

$$\dot{q}_i(t) = \sum_{j=1, j \neq i}^{n} (F_{j \to i} - F_{i \to j}) - F_{i \to 0} + u_i. \tag{9.2}$$

In general, the flow along (i, j) is a function of the entire system state $q = (q_1, \dots, q_n)$ and of time t, so that $F_{i \to j} = F_{i \to j}(q, t)$.

Remarks 9.7 (Basic properties).   (i)   *The mass in each of the compartments as well as the mass flowing along each of the edges must be non-negative at all times (recall we assume $u_i \geq 0$). Specifically, we require the mass flow functions to satisfy*

$$F_{i \to j}(q, t) \geq 0 \text{ for all } (q, t), \text{ and} \qquad F_{i \to j}(q, t) = 0 \text{ for all } (q, t) \text{ such that } q_i = 0. \tag{9.3}$$

Under these conditions, if at some time t_0 one of the compartments has no mass, that is, $q_i(t_0) = 0$ and $q(t_0) \in \mathbb{R}_{\geq 0}^n$, it follows that $\dot{q}_i(t_0) = \sum_{j=1, j \neq i}^{n} F_{j \to i}(q(t_0), t_0) + u_i \geq 0$ so that q_i does not become negative. The compartmental system (9.2) *is therefore a positive system, as introduced in Definition 9.1.*

(ii)   *If $M(q) = \sum_{i=1}^{n} q_i = \mathbb{1}_n^\mathsf{T} q$ denotes the total mass in the system, then along the solutions of* (9.2)

$$\frac{d}{dt} M(q(t)) = \mathbb{1}_n^\mathsf{T} \dot{q}(t) = \underbrace{- \sum_{i=1}^{n} F_{i \to 0}(q(t), t)}_{\text{outflow into environment}} + \underbrace{\sum_{i=1}^{n} u_i}_{\text{inflow from environment}}. \tag{9.4}$$

This equality implies that the total mass $t \mapsto M(q(t))$ is constant in systems without inflows and outflows.   ●

Remark 9.8 (Symmetric physical flow systems).   *Many physical compartmental systems are described by symmetric flows that depend upon effort variables and energy stored at nodes. For an insightful treatment of physical and port-Hamiltonian network systems we refer to (van der Schaft, 2015; van der Schaft and Wei, 2012). We here present a brief introduction without outflows and inflows, for simplicity.*

Following (van der Schaft, 2015), we let G be an undirected graph with n nodes and m edges and with oriented incidence matrix $B \in \mathbb{R}^{n \times m}$ and proceed as follows:

(i)   *for an oriented edge (i, j), let u_{ij} denote the total flow from i to j (that is, $u_{ij} = F_{i \to j} - F_{j \to i}$) so that the flow vector is $u \in \mathbb{R}^m$. Given storage q_i at each node i, mass conservation implies $\dot{q} = Bu \in \mathbb{R}^n$; (if instead the nodes have no storage, then mass conservation implies $Bu = \mathbb{0}_n$, which is consistent with Kirchhoff's current law as stated in Exercise 8.7.)*

(ii)   *typically, the flow through an edge u_{ij} is proportional to an "effort on the edge" e_{ij}, that is, $u_{ij} = -c_{ij} e_{ij}$, for a "conductance constant" $c_{ij} > 0$. In vector form, $u = -Ce \in \mathbb{R}^m$;*

(iii) *typically, the ege effort e_{ij} is the difference of node effort variables, that is, $e = B^\mathsf{T} e_{\mathrm{nodes}} \in \mathbb{R}^m$, for nodal effort variables $e_{\mathrm{nodes}} \in \mathbb{R}^n$;*

(iv) *finally, node efforts are determined by the storage variables according to:*

$$e_{\mathrm{nodes}} = \frac{\partial H}{\partial q}(q) \in \mathbb{R}^n, \tag{9.5}$$

where $H(q)$ is the total stored energy. Typically, $H(q) = \sum_{i=1}^{n} H_i(q_i)$, where $H_i(q_i)$ denotes the energy stored at node i.

In summary, the symmetric physical compartmental system obeys

$$\dot{q} = Bu = -BCe = -BCB^\mathsf{T} e_{\mathrm{nodes}} = -BCB^\mathsf{T} \frac{\partial H}{\partial q}(q) = -L\frac{\partial H}{\partial q}(q), \tag{9.6}$$

where L is the conductance-weighted Laplacian matrix of the compartmental graph.

For example, consider a hydraulic flow network among n fluid reservoirs. The liquid stored at the reservoirs is given by a vector $q \in \mathbb{R}_{\geq 0}^n$. Assume there exists an energy function H_i (possibly the same function at all locations) such that $\frac{\partial H_i}{\partial q_i}(q_i)$ is the pressure at reservoir i. Assume that the liquid flow along the pipe from head reservoir i to tail reservoir j is proportional to the difference between the pressure at i and the pressure at j. Then equation (9.6) describes the mass balance equation among the reservoirs. ●

Linear compartmental systems

Loosely speaking, a compartmental system is linear if it has (i) constant non-negative inflow from the environment and (ii) all other flows depend linearly upon the mass in the originating compartment.

Definition 9.9 (Linear compartmental systems). *A* linear compartmental system *with n compartments is a triplet (F, f_0, u) consisting of*

(i) *a non-negative $n \times n$ matrix $F = (f_{ij})_{i,j \in \{1,\dots,n\}}$ with zero diagonal, called the* flow rate matrix,

(ii) *a vector $f_0 \geq \mathbb{0}_n$, called the* outflow rates vector, *and*

(iii) *a vector $u \geq \mathbb{0}_n$, called the* inflow vector.

The flow rate matrix F is the adjacency matrix of the compartmental digraph G_F (a weighted digraph without self-loops).

The flow rate matrix F encodes the following information: the nodes are the compartments $\{1, \dots, n\}$, there is an edge (i, j) if there is a flow from compartment i to compartment j, and the weight f_{ij} of the (i, j) edge is the corresponding flow rate constant. In a linear compartmental system,

$$\begin{aligned} F_{i \to j}(q, t) &= f_{ij} q_i, \quad \text{for } j \in \{1, \dots, n\}, \\ F_{i \to 0}(q, t) &= f_{0i} q_i, \quad \text{and} \\ u_i(q, t) &= u_i. \end{aligned}$$

Indeed, this model is also referred to as *donor-controlled flow*. Note that this model satisfies the physically-meaningful contraints (9.3). The affine dynamics describing a linear compartmental system is

$$\dot{q}_i(t) = -\Big(f_{0i} + \sum_{j=1, j\neq i}^{n} f_{ij}\Big)q_i(t) + \sum_{j=1, j\neq i}^{n} f_{ji}q_j(t) + u_i. \tag{9.7}$$

Definition 9.10 (Compartmental matrix). *The* compartmental matrix $C = (c_{ij})_{i,j \in \{1,...,n\}}$ *of a compartmental system* (F, f_0, u) *is defined by*

$$c_{ij} = \begin{cases} f_{ji}, & \text{if } i \neq j, \\ -f_{0i} - \sum_{h=1, h\neq i}^{n} f_{ih}, & \text{if } i = j. \end{cases}$$

Equivalently, if $L_F = \mathrm{diag}(F\mathbb{1}_n) - F$ *is the Laplacian matrix of the compartmental digraph,*

$$C = -L_F^\mathsf{T} - \mathrm{diag}(f_0) = F^\mathsf{T} - \mathrm{diag}(F\mathbb{1}_n + f_0). \tag{9.8}$$

In what follows it is convenient to call *compartmental* any matrix C with the following properties:

(i) C is Metzler, that is, $c_{ij} \geq 0$, for $i \neq j$,

(ii) C has non-positive diagonal entries, that is, $c_{ii} \leq 0$ for all i, and

(iii) C is *column diagonally dominant*, that is, $|c_{ii}| \geq \sum_{h=1, h\neq i}^{n} c_{hi}$ for all i.

With the notion of compartmental matrix, the dynamics of the linear compartmental system (9.7) can be written as

$$\dot{q}(t) = Cq(t) + u. \tag{9.9}$$

Moreover, since $L_F\mathbb{1}_n = \mathbb{0}_n$, we know $\mathbb{1}_n^\mathsf{T} C = -f_0^\mathsf{T}$ and, consistently with equation (9.4), we know $\frac{d}{dt}M(q(t)) = -f_0^\mathsf{T} q(t) + \mathbb{1}_n^\mathsf{T} u$.

Remark 9.11 (Symmetric flows). *The donor-controlled model entails a flow* $f_{ij}q_i$ *from i to j and a flow* $f_{ji}q_j$ *from j to i. If the flow rates are equal* $f_{ij} = f_{ji}$, *then the resultant flow as measured from i to j is* $f_{ij}(q_i - q_j)$, *i.e., proportional to the difference in stored quantities. The flow rate matrix F is often symmetric in physical networks.*   •

Algebraic and graphical properties of linear compartmental systems

In this section we present useful properties of compartmental matrices, that are related to those enjoyed by Laplacian and Metzler matrices.

Lemma 9.12 (Spectral properties of compartmental matrices). *For a compartmental system* (F, f_0, u) *with compartmental matrix C,*

(i) *if* $\lambda \in \mathrm{spec}(C)$, *then either* $\lambda = 0$ *or* $\Re(\lambda) < 0$, *and*

(ii) *C is invertible if and only if C is Hurwitz (i.e.,* $\Re(\lambda) < 0$ *for all* $\lambda \in \mathrm{spec}(C)$*).*

Proof. Statement (i) is akin the result in Lemma 6.5 and can be proved by an application of the Geršgorin Disks Theorem 2.8. We invite the reader to fill out the details in Exercise E9.5. Statement (i) immediately implies statement (ii). ∎

Next, we introduce some useful graph-theoretical notions, illustrated in Figure 9.6. In the compartmental digraph, a set of compartments S is

(i) *outflow-connected* if there exists a directed path from every compartment in S to the environment, that is, to a compartment j with a positive flow rate constant $f_{0j} > 0$,

(ii) *inflow-connected* if there exists a directed path from the environment to every compartment in S, that is, from a compartment i with a positive inflow $u_i > 0$,

(iii) a *trap* if there is no directed path from any of the compartments in S to the environment or to any compartment outside S, and

(iv) a *simple trap* is a trap that has no traps inside it.

It is immediate to realize the following equivalence: the system is outflow connected (i.e., all compartments are outflow-connected) if and only if the system contains no trap.
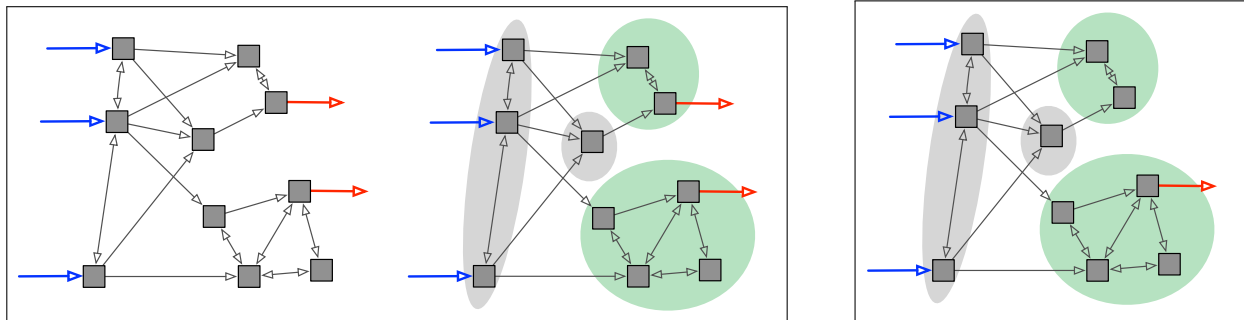
(a) An example compartmental system and its strongly connected components: this system is outflow-connected because its two sinks in the condensation digraph are outflow-connected.

(b) This compartmental system is not outflow-connected because one of its sink strongly-connected components is a trap.

Figure 9.6: Outflow-connectivity and traps in compartmental system

Theorem 9.13 (Algebraic graph theory of compartmental systems). *Consider the linear compartmental system* (F, f_0, u) *with dynamics* (9.9) *with compartmental matrix* C *and compartmental digraph* G_F. *The following statements are equivalent:*

(i) *the system is outflow-connected,*

(ii) *each sink of the condensation of* G_F *is outflow-connected, and*

(iii) *the compartmental matrix* C *is Hurwitz.*

Moreover, the sinks of the condensation of G_F *that are not outflow-connected are precisely the simple traps of the system and their number equals the multiplicity of 0 as a semisimple eigenvalue of* C.

Proof. The equivalence between statements (i) and (ii) is immediate.

To establish the equivalence between (ii) and (iii), we first consider the case in which G_F is strongly connected and at least one compartment has a strictly positive outflow rate. Therefore, the Laplacian matrix L_F of G_F and the compartmental matrix $C = -L_F^\mathsf{T} - \operatorname{diag}(f_0)$ are irreducible. Pick $0 < \varepsilon < 1/\max_i |c_{ii}|$, and define $A = I_n + \varepsilon C^\mathsf{T}$. Because of the definition of ε, the matrix A is non-negative and irreducible. We compute its row-sums as follows:

$$A\mathbb{1}_n = \mathbb{1}_n + \varepsilon(-L_F - \operatorname{diag}(f_0))\mathbb{1}_n = \mathbb{1}_n - \varepsilon f_0.$$

Therefore, A is row-substochastic, i.e., all its row-sums are at most 1 and one row-sum is strictly less than 1. Moreover, because A is irreducible, Corollary 4.11 implies that $\rho(A) < 1$. Now, let $\lambda_1, \ldots, \lambda_n$ denote the eigenvalues of A. Because $A = I_n + \varepsilon C^\mathsf{T}$, we know that the eigenvalues $\eta_1, \ldots, \eta_n$ of C satisfy $\lambda_i = 1 + \varepsilon \eta_i$ so that $\max_i \Re(\lambda_i) = 1 + \varepsilon \max_i \Re(\eta_i)$. Finally, we note that $\rho(A) < 1$ implies $\max_i \Re(\lambda_i) < 1$ so that

$$\max_i \Re(\eta_i) = \frac{1}{\varepsilon}\Big(\max_i \Re(\lambda_i) - 1 \Big) < 0.$$

This concludes the proof that if G is strongly connected, then F has eigenvalues with strictly negative real part. The converse is easy to prove by contradiction: if $f_0 = \mathbb{0}_n$, then the matrix C has zero row-sums, but this is a contradiction with the assumption that C is invertible.

Next, to prove the equivalence between (ii) and (iii) for a graph G_F whose condensation digraph has an arbitrary number of sinks, we proceed as in the proof of Theorem 6.6: we reorder the compartments as described in Exercise E3.1 so that the Laplacian matrix L_F is block lower-triangular as in equation (6.5). We then define an appropriately small ε and the matrix $A = I_n - \varepsilon C^\mathsf{T}$ as above. We leave the remaining details to the reader.

An alternative clever proof strategy for the equivalence between (ii) and (iii) is given as follows. Define the matrix

$$C_{\text{augmented}} = \begin{bmatrix} C & \mathbb{0}_n \\ f_0^\mathsf{T} & 0 \end{bmatrix} \in \mathbb{R}^{(n+1)\times(n+1)},$$

and consider the augmented linear system $\dot{x} = C_{\text{augmented}}x$ with $x \in \mathbb{R}^{n+1}$. Note that $L_{\text{augmented}} = -C_{\text{augmented}}^\mathsf{T}$ is the Laplacian matrix of the *augmented graph* $G_{\text{augmented}}$, whose nodes $\{1, \ldots, n, n+1\}$ include the n compartments and the environment as $(n+1)$st node, and whose edges are the edges of the compartmental graph G_F as well as the outflow edges to the environment node. Note that the environment node $n+1$ in the digraph $G_{\text{augmented}}$ is the only globally reachable node of $G_{\text{augmented}}$ if and only if the compartmental digraph G_F is outflow connected. Assume now that statement (ii) is true. Then, Theorem 7.4 implies

$$\lim_{t\to\infty} e^{-L_{\text{augmented}}t} = \mathbb{1}_{n+1}e_{n+1}^\mathsf{T},$$

which, taking a transpose operation, immediately implies $\lim_{t\to\infty} e^{-C_{\text{augmented}}t} = e_{n+1}\mathbb{1}_{n+1}^\mathsf{T}$. We now can easily compute

$$\lim_{t\to\infty} \begin{bmatrix} q(t) \\ x_{n+1}(t) \end{bmatrix} = e_{n+1}\mathbb{1}_{n+1}^\mathsf{T} \begin{bmatrix} q(0) \\ x_{n+1}(0) \end{bmatrix}$$

$$\implies \lim_{t\to\infty} q(t) = \mathbb{0}_n \quad \lim_{t\to\infty} x_{n+1}(t) = \mathbb{1}_n^\mathsf{T} q(0) + x_{n+1}(0).$$

In other words, all mass in the system reaches asymptotically the environment and the mass in all compartments converge exponentially fast to zero. This occurs for all initial conditions if and only if the matrix C is Hurwitz. Hence we have established that statement (ii) implies statement (iii). We leave the converse to the reader. ∎

Dynamic properties of linear compartmental systems

Consider a linear compartmental system (F, f_0, u) with compartmental matrix C and compartmental digraph G_F. Assuming the system has at least one trap, we define the *reduced compartmental system* $(F_{\mathrm{rd}}, f_{0,\mathrm{rd}}, u_{\mathrm{rd}})$ as follows: remove all traps from G_F and regard the edges into the trapping compartments as outflow edges into the environment, e.g., see Figure 9.7.

(a) A compartmental system that is not outflow-connected

(b) The corresponding reduced compartmental system
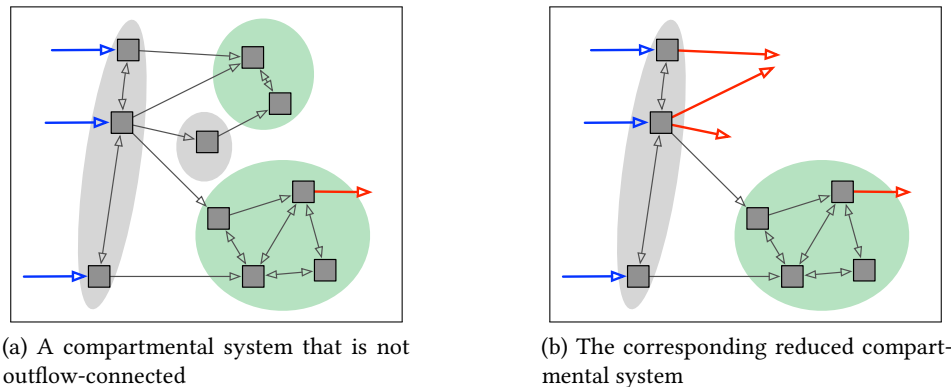
Figure 9.7: An example reduced compartmental system

We now state our main result about the asymptotic behavior of linear compartmental systems.

Theorem 9.14 (Asymptotic behavior of compartmental systems). *The linear compartmental system* (F, f_0, u) *with compartmental matrix* C *and compartmental digraph* G_F *has the following possible asymptotic behaviors:*

 (i) *if the system is outflow-connected, then the compartmental matrix* C *is invertible, every solution tends exponentially to the unique equilibrium* $q^* = -C^{-1}u \geq \mathbb{0}_n$, *and in the ith compartment* $q_i^* > 0$ *if and only if the ith compartment is inflow-connected to a positive inflow;*

 (ii) *if the system contains one or more simple traps, then:*

   a) *the reduced compartmental system* $(F_{\mathrm{rd}}, f_{0,\mathrm{rd}}, u_{\mathrm{rd}})$ *is outflow-connected and all its solutions converge exponentially fast to the unique non-negative equilibrium* $-C_{\mathrm{rd}}^{-1}u_{\mathrm{rd}}$, *for* $C_{\mathrm{rd}} = F_{\mathrm{rd}}^{\mathsf{T}} - \mathrm{diag}(F_{\mathrm{rd}}\mathbb{1}_n + f_{0,\mathrm{rd}})$;

   b) *any simple trap* H *contains non-decreasing mass along time. If* H *is inflow-connected to a positive inflow, then the mass inside* H *goes to infinity. Otherwise, the mass inside* H *converges to a scalar multiple of the right eigenvector corresponding to the eigenvalue* 0 *of the compartmental submatrix for* H.

Proof. Statement (i) is an immediate consequence of Corollary 9.6. We leave the proof of statement (ii) to the reader. ∎

9.4 Table of asymptotic behaviors for averaging and positive systems

Dynamics	Assumptions & Asymptotic Behavior	References
averaging system $x(k+1) = Ax(k)$ A row-stochastic	the associated digraph has a globally reachable node $\implies$ $\lim_{k\to\infty} x(k) = (w^{\mathsf{T}}x(0))\mathbb{1}_n$ where $w \geq 0$ is the left eigenvector of A with eigenvalue 1 satisfying $\mathbb{1}_n^{\mathsf{T}}w = 1$	Convergence properties: Theorem 5.1. Examples: opinion dynamics & averaging in Chapter 1
affine system $x(k+1) = Ax(k) + b$	A convergent (that is, its spectral radius is less than 1) $\implies$ $\lim_{k\to\infty} x(k) = (I_n - A)^{-1}b$	Convergence properties: Exercise E2.10. Examples: Friedkin-Johnsen system in Exercise E5.7
positive affine system $x(k+1) = Ax(k) + b$ $A \geq 0, b \geq \mathbb{0}_n$	$x(0) \geq \mathbb{0}_n \implies x(k) \geq \mathbb{0}_n$ for all k, and A convergent (that is, $\|\lambda\| < 1$ for all $\lambda \in \operatorname{spec}(A)$) $\implies$ $\lim_{k\to\infty} x(k) = (I_n - A)^{-1}b \geq \mathbb{0}_n$	Positivity properties: Exercise E9.9 Examples: Leslie population model in Exercise E4.14

Table 9.1: Discrete-time systems

Dynamics	Assumptions & Asymptotic Behavior	References
averaging system $\dot{x}(t) = -Lx(t)$ L Laplacian matrix	the associated digraph has a globally reachable node $\implies$ $\lim_{t\to\infty} x(t) = (w^{\mathsf{T}}x(0))\mathbb{1}_n$ where $w \geq 0$ is the left eigenvector of L with eigenvalue 0 satisfying $\mathbb{1}_n^{\mathsf{T}}w = 1$	Convergence properties: Theorem 7.4. Examples: Flocking system in Section 7.1.2
affine system $\dot{x}(t) = Ax(t) + b$	A Hurwitz (that is, its spectral abscissa is negative) $\implies$ $\lim_{t\to\infty} x(t) = -A^{-1}b$	Convergence properties: Exercise E7.2
positive affine system $\dot{x}(t) = Ax(t) + b$ A Metzler, $b \geq \mathbb{0}_n$	$x(0) \geq \mathbb{0}_n \implies x(t) \geq \mathbb{0}_n$ for all t, and A Hurwitz (that is, $\Re(\lambda) < 0$ for all $\lambda \in \operatorname{spec}(A)$) $\implies$ $\lim_{t\to\infty} x(t) = -A^{-1}b \geq \mathbb{0}_n$	Positivity properties: Theorem 9.3 and Corollary 9.6. Example: compartmental systems in Section 9.1.

Table 9.2: Continuous-time systems

9.5 Appendix: A static nonlinear flow problem

In this appendix, we consider a static compartmental flow system, where a commodity (e.g., power or water) is transported through a network (e.g., a power grid or a piping system). We model this scenario with an undirected and connected graph with n nodes and m edges. With each node we associate an external supply/demand variable (positive for a source and negative for a sink) y_i and assume that the overall network is balanced: $\sum_{i=1}^{n} y_i = 0$. We also associate a potential variable x_i with every node (e.g., voltage or pressure), and assume the flow of commodity between two connected nodes i and j depends on the potential difference as $f_{ij}(x_i - x_j)$, where f_{ij} is a strictly increasing function satisfying $f_{ij}(0) = 0$. For example, for piping systems and power grids these functions f_{ij} are given by the rational Hazen-Williams flow and the trigonometric power flow, which are both monotone in the region of interest. By balancing the flow at each node (akin to the Kirchhoff's current law), we obtain at node i

$$y_i = \sum_{j=1}^{n} a_{ij} f_{ij}(x_i - x_j), \quad i \in \{1, \ldots, n\},$$

where $a_{ij} \in \{0, 1\}$ is the (i, j) element of the network adjacency matrix. In vector notation, the flow balance is

$$y = Bf(B^\mathsf{T} x),$$

where the map $f : \mathbb{R}^m \to \mathbb{R}^m$ has components f_{ij}. Consider also the associated linearized problem $y = BB^\mathsf{T} x = Lx$, where L is the network Laplacian matrix, where we implicitly assumed $f'_{ij}(0) = 1$. The flows in the linear problem are obtained as

$$B^\mathsf{T} x^\star = B^\mathsf{T} L^\dagger y,$$

where $L^\dagger$ is the Moore-Pennrose pseudoinverse of L; see Exercises E6.8 and E6.10.

In what follows, we restrict ourselves to an acyclic network and show that the nonlinear solution can be obtained from the solution of the linear problem. We formally replace the flow $f(B^\mathsf{T} x)$ by a new edge variable $v := f(B^\mathsf{T} x) \in \mathbb{R}^m$ and arrive at

$$y = Bv, \tag{9.10a}$$

$$v = f(B^\mathsf{T} x). \tag{9.10b}$$

In the acyclic case, $\operatorname{kernel}(B) = \{\mathbb{0}_m\}$ and necessarily $v \in \operatorname{image}(B^\mathsf{T})$, or $v = B^\mathsf{T} w$ for some $w \in \mathbb{R}^n$. Thus, equation (9.10a) reads as $y = Bv = BB^\mathsf{T} w = Lw$ and its solution is $w = L^\dagger y$. Equation (9.10b) then reads as $f(B^\mathsf{T} x) = v = B^\mathsf{T} w = B^\mathsf{T} L^\dagger y$, and its unique solution (due to monotonicity) is

$$B^\mathsf{T} x^\star = f^{-1}(B^\mathsf{T} L^\dagger y).$$

9.6 Historical notes and further reading

This chapter is inspired by the excellent text (Walter and Contreras, 1999) and the tutorial treatment in (Jacquez and Simon, 1993); see also the texts (Luenberger, 1979; Farina and Rinaldi, 2000; Haddad et al.,

2010). Additional results on Metzler matrices are available in (Berman and Plemmons, 1994; Santesso and Valcher, 2007). For nonlinear extensions of the material in this chapter, including recent studies of traffic networks, we refer to (Como et al., 2013; Coogan and Arcak, 2015).

Several other properties of positive affine systems and Metzler matrices are reviewed in (Berman and Plemmons, 1994).

9.7 Exercises

E9.1 **The matrix exponential of a Metzler matrix.** In this exercise we extend and adapt Theorem 7.2 about the matrix exponential of a Laplacian matrix to the setting of Metzler matrices.

Let M be an $n \times n$ Metzler matrix with minimum diagonal entry $m_{\min} = \min\{m_{11}, \ldots, m_{nn}\}$. As usual, associate to M a digraph G without self-loops in the natural way, that is, (i, j) is an edge if and only if $m_{ij} > 0$. Prove that

(i) $\exp(M) \geq e^{m_{\min}} I_n \geq 0$,   for any digraph G,

(ii) $\exp(M) e_j > 0$,   for a digraph G whose j-th node is globally reachable,

(iii) $\exp(M) > 0$,   for a strongly connected digraph G (i.e., for an irreducible M).

Morever, prove that, for any square matrix A,

(iv) $\exp(At) \geq 0$ for all $t \geq 0$ if and only if A is Metzler.

E9.2 **Proof of the Perron-Frobenius Theorem for Metzler matrices.** Prove Theorem 9.4.

E9.3 **Metzler invariance under non-negative change of basis.** Consider a positive system with Metzler matrix A and constant input $b \geq 0$:
$$\dot{x} = Ax + b.$$

Show that, under the change of basis
$$z = T^{-1}x,$$

with T invertible and $T^{-1} \geq 0$, the transformed matrix $T^{-1}AT$ is also Metzler.

E9.4 **Monotonicity properties of positive systems.** Consider the continuous-time positive affine system
$$\dot{x} = Ax + b,$$

where A is Metzler and b is non-negative.

(i) Let $x(t, x_0)$ denote the solution from initial condition $x_0 \in \mathbb{R}^n_{\geq 0}$ at time 0. Show that
$$\mathbb{0}_n \leq x_0 \leq x_1 \quad \implies \quad x(t, x_0) \leq x(t, x_1) \quad \text{for all time } t \geq 0.$$

(ii) Let $\dot{x} = \widehat{A}x + \widehat{b}$ be a second continuous-time positive affine system. Assume that A and $\widehat{A}$ are Hurwitz and, by Corollary 9.6, let x^* and $\widehat{x}^*$ denote the equilibrium points of the two systems. Show that
$$A \geq \widehat{A} \quad \text{and} \quad b \geq \widehat{b} \quad \implies \quad x^* \geq \widehat{x}^*.$$

E9.5 **Establishing the spectral properties of compartmental matrices.** Prove Lemma 9.12 about the spectral properties of compartmental matrices.

E9.6 **Simple traps and strong connectivity.** Show that a compartmental system that has no outflows and that is a simple trap, is strongly connected.

E9.7 **Sufficient condition for a Metzler matrix to be Hurwitz.** For $n \geq 2$, given a Metzler matrix $M \in \mathbb{R}^{n \times n}$, let $v = M \mathbb{1}_n \in \mathbb{R}^n$ denote its vector of row sums. Show that

(i) $M - \operatorname{diag}(v)$ is a Metzler matrix with zero row sums,

(ii) if M is irreducible and v is nonpositive with at least one entry strictly negative, then M is Hurwitz, and

(iii) if S is a symmetric irreducible Metzler matrix with $S\mathbb{1}_n = \mathbb{0}_n$, then, for any $i \in \{1, \ldots, n\}$ and $\varepsilon > 0$, all eigenvalues of $A - \varepsilon e_i e_i^{\mathsf{T}}$ are negative.

E9.8   **On Metzler matrices and compartmental systems with growth and decay.** Let M be an $n \times n$ symmetric Metzler matrix. Recall Lemma 9.12 and define $v \in \mathbb{R}^n$ by $M = -L + \mathrm{diag}(v)$, where L is a symmetric Laplacian matrix. Show that:

  (i) if M is Hurwitz, then $\mathbb{1}_n^\mathsf{T} v < 0$.

Next, assume $n = 2$ and assume v has both non-negative and non-positive entries. (If v is non-negative, lack of stability can be established from statement (i); if v is non-positive, stability can be established via Theorem 9.13.) Show that

  (ii) there exist non-negative numbers f, d and g such that, modulo a permutation, M can be written in the form:
  $$M = -f \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} + \begin{bmatrix} g & 0 \\ 0 & -d \end{bmatrix} = \begin{bmatrix} (g - f) & f \\ f & (-d - f) \end{bmatrix},$$

  (iii) M is Hurwitz if and only if
  $$d > g \quad \text{and} \quad f > \frac{gd}{d - g}.$$

  Note: *The inequality $d > g$ (for $n = 2$) is equivalent to the inequality $\mathbb{1}_n^\mathsf{T} v < 0$ in statement (i). In the interpretation of compartmental systems with growth and decay rates, f is a flow rate, d is a decay rate and g is a growth rate. With this interpretation, the statement (iii) is then interpreted as follows: M is Hurwitz if and only if the decay rate is larger than the growth rate and the flow rate is sufficiently large.*

E9.9   **Non-negative inverse.** Let A be a non-negative square matrix and show that the following statements are equivalent:

  (i) $\lambda > \rho(A)$, and

  (ii) the matrix $(\lambda I_n - A)$ is invertible and its inverse $(\lambda I_n - A)^{-1}$ is non-negative.

  Moreover, show that

  (iii) if A is irreducible and $\lambda > \rho(A)$, then $(\lambda I_n - A)^{-1}$ is positive.

  (Given a square matrix A, the map $\lambda \mapsto (\lambda I_n - A)^{-1}$ is sometimes referred to as the *resolvent* of A.)

E9.10   **Grounded Laplacian matrices.** Let G be a weighted undirected graph with Laplacian $L \in \mathbb{R}^{n \times n}$. Select a set S of $s \geq 1$ nodes and call them *grounded nodes*. Given S, the *grounded Laplacian matrix* $L_{\mathrm{grounded}} \in \mathbb{R}^{(n-s) \times (n-s)}$ is the principal submatrix of L obtained by removing the s rows and columns corresponding to the grounded nodes. In other words, if the grounded nodes are nodes $\{n - s + 1, \dots, n\}$ and L is partitioned in block matrix form

$$L = \begin{bmatrix} L_{11} & L_{12} \\ L_{12}^\mathsf{T} & L_{22} \end{bmatrix}, \qquad \text{with } L_{11} \in \mathbb{R}^{(n-s) \times (n-s)} \text{ and } L_{22} \in \mathbb{R}^{s \times s},$$

then $L_{\mathrm{grounded}} = L_{11}$. Show the following statements:

  (i) If G is connected, then

      a) L_{grounded} is positive definite,
      b) $L_{\mathrm{grounded}}^{-1}$ is non-negative, and
      c) the eigenvector associated with the smallest eigenvalue of L_{grounded} can be selected non-negative.

  (ii) If additionally the graph obtained by removing from G the nodes in S and all the corresponding edges is connected, then

      d) $L_{\mathrm{grounded}}^{-1}$ is positive, and
      e) the eigenvector associated with the smallest eigenvalue of L_{grounded} is unique and positive (up to rescaling).

Hint: Show that $-L_{\text{grounded}}$ *is a compartmental matrix.*

Note: *For more information on grounded Laplacian matrices we refer to (Dörfler and Bullo, 2013; Pirani and Sundaram, 2016; Xia and Cao, 2017).*

E9.11 **Mean Residence Time for a particle in a compartmental system.** Consider an outflow-connected compartmental system with irreducible matrix C and $\mu(C) < 0$. Let v is the dominant eigenvector of C, that is, $Cv = \mu(C)v$, $\mathbb{1}_n^\mathsf{T} v = 1$, and $v > 0$.

   Assume a tagged particle is randomly located inside the compartmental system at time 0 with probability mass function v. The *mean residence time (mrt)* of the tagged particle is the expected time that the particle remains inside the compartmental system.

   Using the definition of expectation, the mean residence time is

$$\text{mrt} = \int_0^\infty t\, \mathbb{P}[\text{particle leaves at time } t]\, dt.$$

Let us also take for granted that:

$$\mathbb{P}[\text{particle leaves at time } t] = -\left(\frac{d}{dt}\mathbb{P}[\text{particle inside at time } t]\right).$$

Show that

$$\text{mrt} = -\frac{1}{\mu(C)}.$$

E9.12 **Resistive circuits as compartmental systems (Dörfler et al., 2017).** Consider a resistive circuit with shunt capacitors at each node as in figure below (see also in Section 7.1.3). Assume that the circuit is connected. Attach to at least one node $j \in \{1, \ldots, n\}$ a current source generating an injected current $c_{\text{injected at } j} > 0$, and connect to at least one node $i \in \{1, \ldots, n\}$ a positive resistor to ground.

   (i) Model the resulting system as a compartmental system, i.e., identify the conserved quantity and write the compartmental matrix, the inflow vector and the outflow rate vector, and

   (ii) show that there exists a unique steady state that is positive and globally-asymptotically stable.
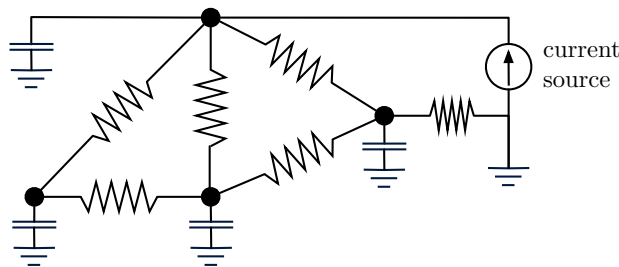
E9.13 **Solutions of partial differential equations (Luenberger, 1979, Chapter 6).** The electric potential V within a two-dimensional domain is governed by the Laplace's partial differential equation:

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} = 0, \tag{E9.1}$$

combined with the value of V along the boundary of the enclosure; see the left image in Figure E9.1. (A similar setup with a time-varying spatial quantity and free boundary conditions was described in Section 7.1.4.)
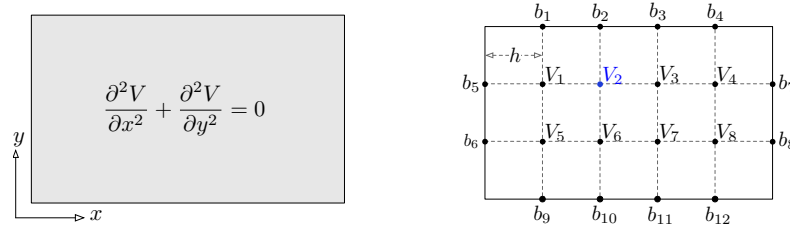
Figure E9.1: Laplace's equation over a rectangular enclosure and a regular Cartesian grid.

For arbitrary enclosures and boundary conditions, it is impossible to solve the Laplace's equation in closed form. An approximate solution is computed by (i) introducing a regular Cartesian grid of points with spacing h, e.g., see the right image in Figure E9.1, and (ii) approximating the second-order derivatives by second-order finite differences. Specifically, at node 2 of the grid, we have along the x direction

$$\frac{\partial^2 V}{\partial x^2}(V_2) \approx \frac{1}{h^2}(V_3 - V_2) - \frac{1}{h^2}(V_2 - V_1) = \frac{1}{h^2}(V_3 + V_2 - 2V_2),$$

so that equation (E9.1) is approximated as follows:

$$0 = \frac{\partial^2 V}{\partial x^2}(V_2) + \frac{\partial^2 V}{\partial y^2}(V_2) \approx \frac{1}{h^2}(V_1 + V_3 + V_6 + b_2 - 4V_2) \quad \Longrightarrow \quad 4V_2 = V_1 + V_3 + V_6 + b_2.$$

This approximation translates into the matrix equation:

$$4V = A_{\text{grid}}V + C_{\text{grid-boundary}}b, \tag{E9.2}$$

where $V \in \mathbb{R}^n$ is the vector of unknown potentials, $b \in \mathbb{R}^m$ is the vector of boundary conditions, $A_{\text{grid}} \in \{0,1\}^{n \times n}$ is the binary adjacency matrix of the *(interior) grid graph* (that is, $(A_{\text{grid}})_{ij} = 1$ if and only if the interior nodes i and j are connected), and $C_{\text{grid-boundary}} \in \{0,1\}^{n \times m}$ is the connection matrix between interior and boundary nodes (that is, $(C_{\text{grid-boundary}})_{i\alpha} = 1$ if and only if grid interior node i is connected with boundary node α). Show that

(i) A_{grid} is irreducible but not primitive,

(ii) $\rho(A_{\text{grid}}) < 4$,

   Hint: *Recall Theorem 4.9.*

(iii) there exists a unique solution V^* to equation (E9.2),

(iv) the unique solution V^* satisfies $V^* \geq \mathbb{0}_n$ if $b \geq \mathbb{0}_m$, and

(v) each solution to the following iteration converges to V^*:

$$4V(k+1) = A_{\text{grid}}V(k) + C_{\text{grid-boundary}}b,$$

whereby, at each step, the value of V at each node is updated to be equal to the average of its neighboring nodes.

E9.14 **Irreducible Metzler matrices and positive vectors.** Let $M \in \mathbb{R}^{n \times n}$ be an irreducible Metzler matrix, let $\mu(M)$ be its spectral abscissa (i.e., its dominant eigenvalue), and let $x > \mathbb{0}_n$ be a positive vector. Show that

(i) if $Mx < \eta x$ for some $\eta \in \mathbb{R}$, then $\mu(M) < \eta$;

(ii) if $Mx = \eta x$ for some $\eta \in \mathbb{R}$, then $\mu(M) = \eta$; and

(iii) if $Mx > \eta x$ for some $\eta \in \mathbb{R}$, then $\mu(M) > \eta$.

Hint: *Read Section 2.1 in (Varga, 2009).*

E9.15 **Discrete-time compartmental systems.** Perform the following steps.

    (i) Provide a proper definition of a discrete-time compartmental system without inflows or outflows.

    (ii) Show that the system dynamics takes the form $x(k+1) = A^\mathsf{T} x(k)$, where A is row stochastic.

    (iii) Show that the system admits a globally exponentially stable equilibrium if each sink of the digraph associated to A is aperiodic.

Part II

Topics in Averaging Systems

Convergence Rates, Scalability and Optimization

In this chapter we discuss the convergence rate of averaging algorithms. We focus on discrete-time systems and their convergence factors. The study of continuous-time systems is analogous. We also perform a scalability analysis for an example system and discuss some interesting optimization problems.

Before proceeding, we recall a few basic facts. Given a square matrix A,

(i) the spectral radius of A is $\rho(A) = \max\{|\lambda| \mid \lambda \in \mathrm{spec}(A)\}$;

(ii) the p-induced norm of A, for $p \in \mathbb{N} \cup \{\infty\}$, is

$$\|A\|_p = \max\left\{\|Ax\|_p \mid x \in \mathbb{R}^n \text{ and } \|x\|_p = 1\right\} = \max_{x \neq \mathbb{0}_n} \frac{\|Ax\|_p}{\|x\|_p},$$

and, specifically, the induced 2-norm of A is $\|A\|_2 = \max\{\sqrt{\lambda} \mid \lambda \in \mathrm{spec}(A^\mathsf{T} A)\}$;

(iii) for any p, $\rho(A) \leq \|A\|_p$; and

(iv) if $A = A^\mathsf{T}$, then $\|A\|_2 = \rho(A)$.

Definition 10.1 (Essential spectral radius). *The* essential spectral radius *of a row-stochastic matrix A is*

$$\rho_{\mathrm{ess}}(A) = \begin{cases} 0, & \text{if } \mathrm{spec}(A) = \{1, \ldots, 1\}, \\ \max\{|\lambda| \mid \lambda \in \mathrm{spec}(A) \setminus \{1\}\}, & \text{otherwise.} \end{cases}$$

10.1 Some preliminary calculations and observations

The convergence factor for symmetric row-stochastic matrices   To build some intuition about the general case, we start with a weighted undirected graph G with adjacency matrix A that is row-stochastic and primitive (i.e., the graph G, viewed as a digraph, is strongly connected and aperiodic). We consider the corresponding discrete-time averaging algorithm

$$x(k+1) = Ax(k).$$

Note that G undirected implies that A is symmetric. Therefore, A has real eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ and corresponding orthonormal eigenvectors $v_1, \ldots, v_n$. Because A is row-stochastic, $\lambda_1 = 1$ and $v_1 = \mathbb{1}_n/\sqrt{n}$. Next, along the same lines of the modal decomposion given in Section 2.1, we know that the solution can be decoupled into n independent evolution equations as

$$x(k) = \text{average}(x(0))\mathbb{1}_n + \lambda_2^k(v_2^\mathsf{T}x(0))v_2 + \cdots + \lambda_n^k(v_n^\mathsf{T}x(0))v_n.$$

Moreover, A being primitive implies that $\max\{|\lambda_2|, \ldots, |\lambda_n|\} < 1$. Specifically, for a symmetric and primitive A, we have $\rho_{\text{ess}}(A) = \max\{|\lambda_2|, |\lambda_n|\} < 1$. Therefore, as predicted by Corollary 2.15

$$\lim_{k \to \infty} x(k) = \mathbb{1}_n\mathbb{1}_n^\mathsf{T}x(0)/n = \text{average}(x(0))\mathbb{1}_n.$$

To upper bound the error, since the vectors $v_1, \ldots, v_n$ are orthonormal, we compute

$$\left\|x(k) - \text{average}(x(0))\mathbb{1}_n\right\|_2 = \left\|\sum_{j=2}^n \lambda_j^k(v_j^\mathsf{T}x(0))v_j\right\|_2 = \sqrt{\sum_{j=2}^n |\lambda_j|^{2k}\left\|(v_j^\mathsf{T}x(0))v_j\right\|_2^2}$$

$$\leq \rho_{\text{ess}}(A)^k\sqrt{\sum_{j=2}^n \left\|(v_j^\mathsf{T}x(0))v_j\right\|_2^2} = \rho_{\text{ess}}(A)^k\left\|x(0) - \text{average}(x(0))\mathbb{1}_n\right\|_2, \quad (10.1)$$

where the second and last equalities are Pythagoras Theorem.

In summary, we have learned that, for symmetric matrices, the essential spectral radius $\rho_{\text{ess}}(A) < 1$ is the *convergence factor* to average consensus, i.e., the factor determining the exponential convergence of the error to zero. (The wording "convergence factor" is for discrete-time systems, whereas the wording "convergence rate" is for continuous-time systems.)

A note on convergence factors for asymmetric matrices

The behavior of asymmetric row-stochastic matrices is more complex than of symmetric ones. For large even n, consider the asymmetric positive matrix

$$A_{\text{large-gain}} = \frac{1}{2n}\mathbb{1}_n\mathbb{1}_n^\mathsf{T} + \frac{1}{2}\left(\mathbb{1}_{1:n/2}e_1^\mathsf{T} + \mathbb{1}_{n/2:n}e_n^\mathsf{T}\right),$$

where $\mathbb{1}_{1:n/2}$ (resp. $\mathbb{1}_{n/2:n}$) is the vector whose first (resp. second) $n/2$ entries are equal to 1 and whose second (resp. first) $n/2$ entries are equal to 0. We visualize the digraph associated to this matrix in Figure 10.1.

The matrix $A_{\text{large-gain}}$ is row-stochastic because, given $\mathbb{1}_n^\mathsf{T}\mathbb{1}_n = n$ and $e_j^\mathsf{T}\mathbb{1}_n = 1$ for all j, we compute

$$A_{\text{large-gain}}\mathbb{1}_n = \frac{1}{2}\mathbb{1}_n + \frac{1}{2}(\mathbb{1}_{1:n/2} + \mathbb{1}_{n/2:n})1 = \mathbb{1}_n.$$

Therefore, Corollary 2.15 implies that every solution to $x(k+1) = A_{\text{large-gain}}x(k)$ converges to consensus and Exercise E1.1 implies that $k \mapsto V_{\text{max-min}}(x(k))$ is non-increasing. Nevertheless, the 2-norm of the deviation from consensus can easily increase. For example, take $x(0) = e_1 - e_n$ and compute

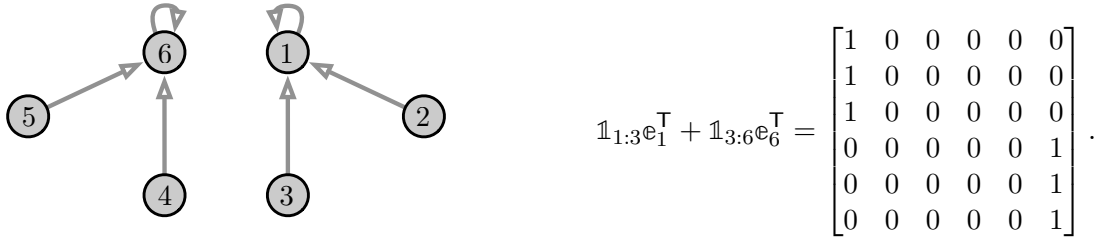$$x(1) = A_{\text{large-gain}}x(0) = \frac{1}{2}\mathbb{1}_{1:n/2} - \frac{1}{2}\mathbb{1}_{n/2:n}.$$

$$
\mathbb{1}_{1:3}\mathbb{e}_1^\mathsf{T} + \mathbb{1}_{3:6}\mathbb{e}_6^\mathsf{T} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.
$$

Figure 10.1: The unweighted digraph associated to the matrix $\mathbb{1}_{1:n/2}\mathbb{e}_1^\mathsf{T} + \mathbb{1}_{n/2:n}\mathbb{e}_n^\mathsf{T}$, for $n = 6$. This digraph is the union of two disjoint stars. The weighted digraph associated to $A_{\text{large-gain}}$ is the superposition of these two stars with a complete digraph.

Because $\text{average}(x(0)) = \text{average}(x(1)) = 0$, we compute

$$
\|x(0) - \text{average}(x(0))\mathbb{1}_n\|_2 = \sqrt{2} \quad \text{and} \quad \|x(1) - \text{average}(x(1))\|_2 = \frac{1}{2}\|\mathbb{1}_{1:n/2} - \mathbb{1}_{n/2:n}\|_2 = \frac{1}{2}\sqrt{n}.
$$

In other words, the 2-norm of $x(k) - \text{average}(x(k))$ along the averaging system defined by $A_{\text{large-gain}}$ grows to be at least of order $\sqrt{n}$ (starting from $\mathcal{O}(1)$).[1] The problem is that the eigenvalues (alone) of a non-symmetric matrix do not fully describe the state amplification that may take place during a transient period of time.

10.2 Convergence factors for row-stochastic matrices

Consider a discrete-time averaging algorithm (distributed linear averaging)

$$
x(k + 1) = Ax(k),
$$

where A is doubly-stochastic and not necessarily symmetric. If A is primitive (i.e., the associated digraph is aperiodic and strongly connected), we know

$$
\lim_{k \to \infty} x(k) = \text{average}(x(0))\mathbb{1}_n = \left(\mathbb{1}_n\mathbb{1}_n^\mathsf{T}/n\right)x(0).
$$

We now define two possible notions of convergence factors. The *per-step convergence factor* is

$$
r_{\text{step}}(A) = \sup_{x(k) \neq x_{\text{final}}} \frac{\|x(k + 1) - x_{\text{final}}\|_2}{\|x(k) - x_{\text{final}}\|_2},
$$

where $x_{\text{final}} = \text{average}(x(0))\mathbb{1}_n = \text{average}(x(k))\mathbb{1}_n$ and where the supremum is taken over any possible sequence. Moreover, the *asymptotic convergence factor* is

$$
r_{\text{asym}}(A) = \sup_{x(0) \neq x_{\text{final}}} \lim_{k \to \infty} \left(\frac{\|x(k) - x_{\text{final}}\|_2}{\|x(0) - x_{\text{final}}\|_2}\right)^{1/k}.
$$

Given these definitions the preliminary calculations in the previous Section 10.1, we can now state our main results.

---

[1]Here and in what follows, $\mathcal{O}(x)$ is a scalar function upper bounded by a constant times x.

Theorem 10.2 (Convergence factor and solution bounds). *Let A be doubly-stochastic and primitive.*

  (i) *The convergence factors of A satisfy*

$$r_{\text{step}}(A) = \|A - \mathbb{1}_n \mathbb{1}_n^\mathsf{T}/n\|_2,$$
$$r_{\text{asym}}(A) = \rho_{\text{ess}}(A) = \rho(A - \mathbb{1}_n \mathbb{1}_n^\mathsf{T}/n) < 1. \tag{10.2}$$

  Moreover, $r_{\text{asym}}(A) \leq r_{\text{step}}(A)$, and $r_{\text{step}}(A) = r_{\text{asym}}(A)$ if A is symmetric.

  (ii) *For any initial condition $x(0)$ with corresponding $x_{\text{final}} = \text{average}(x(0))\mathbb{1}_n$,*

$$\|x(k) - x_{\text{final}}\|_2 \leq r_{\text{step}}(A)^k \|x(0) - x_{\text{final}}\|_2, \tag{10.3}$$
$$\|x(k) - x_{\text{final}}\|_2 \leq c_\varepsilon (r_{\text{asym}}(A) + \varepsilon)^k \|x(0) - x_{\text{final}}\|_2, \tag{10.4}$$

  where $\varepsilon > 0$ is an arbitrarily small constant and c_ε is a sufficiently large constant independent of $x(0)$.

Note: A sufficient condition for $r_{\text{step}}(A) < 1$ is given in Exercise E10.1.

Before proving Theorem 10.2, we introduce an interesting intermediate result. For $x_{\text{final}} = \text{average}(x(0))\mathbb{1}_n$, the *disagreement vector* is the error signal

$$\delta(k) = x(k) - x_{\text{final}}. \tag{10.5}$$

Lemma 10.3 (Disagreement or error dynamics). *Given a doubly-stochastic matrix A, the disagreement vector $\delta(k)$ satisfies*

  (i) *$\delta(k) \perp \mathbb{1}_n$ for all k,*

  (ii) *$\delta(k + 1) = \big(A - \mathbb{1}_n \mathbb{1}_n^\mathsf{T}/n\big)\delta(k)$,*

  (iii) *the following properties are equivalent:*

     a) *$\lim_{k \to \infty} A^k = \mathbb{1}_n \mathbb{1}_n^\mathsf{T}/n$, (that is, the averaging algorithm achieves average consensus)*
     b) *A is primitive, (that is, the digraph is aperiodic and strongly connected)*
     c) *$\rho(A - \mathbb{1}_n \mathbb{1}_n^\mathsf{T}/n) < 1$. (that is, the error dynamics is convergent)*

Proof. To study the error dynamics, note that $\mathbb{1}_n^\mathsf{T} x(k+1) = \mathbb{1}_n^\mathsf{T} A x(k)$ and, in turn, that $\mathbb{1}_n^\mathsf{T} x(k) = \mathbb{1}_n^\mathsf{T} x(0)$; see also Exercise E7.8. Therefore, $\text{average}(x(0)) = \text{average}(x(k))$ and $\delta(k) \perp \mathbb{1}_n$ for all k. This completes the proof of statement (i). To prove statement (ii), we compute

$$\delta(k + 1) = A x(k) - x_{\text{final}} = A x(k) - (\mathbb{1}_n \mathbb{1}_n^\mathsf{T}/n)x(k) = \big(A - \mathbb{1}_n \mathbb{1}_n^\mathsf{T}/n\big)x(k),$$

and the equation in statement (ii) follows from $\big(A - \mathbb{1}_n \mathbb{1}_n^\mathsf{T}/n\big)\mathbb{1}_n = \mathbb{0}_n$.

Next, let us prove the equivalence among the three properties. From Perron–Frobenius Theorem 2.12 for primitive matrices in Chapter 2 and from Corollary 2.15, we know that A primitive (statement (iii)b) implies average consensus (statement (iii)a). The converse is true because $\mathbb{1}_n \mathbb{1}_n^\mathsf{T}/n$ is a positive matrix and, by the definition of limit, there must exist k such that each entry of A^k becomes positive.

Finally, we prove the equivalence between statement (iii)a and (iii)c. First, note that $P = I_n - \mathbb{1}_n\mathbb{1}_n^{\mathsf{T}}/n$ is a projection matrix, that is, $P = P^2$. This can be easily verified by expanding the matrix power P^2. Second, let us prove a useful identity:

$$
\begin{aligned}
A^k - \mathbb{1}_n\mathbb{1}_n^{\mathsf{T}}/n &= A^k(I_n - \mathbb{1}_n\mathbb{1}_n^{\mathsf{T}}/n) && \text{(because A row-stochastic)} \\
&= A^k(I_n - \mathbb{1}_n\mathbb{1}_n^{\mathsf{T}}/n)^k && \text{(because $I_n - \mathbb{1}_n\mathbb{1}_n^{\mathsf{T}}/n$ is a projection)} \\
&= \left(A(I_n - \mathbb{1}_n\mathbb{1}_n^{\mathsf{T}}/n)\right)^k = \left(A - \mathbb{1}_n\mathbb{1}_n^{\mathsf{T}}/n\right)^k.
\end{aligned}
$$

The statement follows from taking the limit as $k \to \infty$ in this identity and by recalling that a matrix is convergent if and only if its spectral radius is less then one. ∎

We are now ready to prove the main theorem in this section.

Proof of Theorem 10.2. Regarding the equalities (10.2), the formula for r_{step} is an consequence of the definition of induced 2-norm:

$$
\begin{aligned}
r_{\text{step}}(A) &= \sup_{x(k)\neq x_{\text{final}}} \frac{\|x(k+1) - x_{\text{final}}\|_2}{\|x(k) - x_{\text{final}}\|_2} \\
&= \sup_{\delta(k)\perp\mathbb{1}_n} \frac{\|\delta(k+1)\|_2}{\|\delta(k)\|_2} = \sup_{\delta(k)\perp\mathbb{1}_n} \frac{\|(A - \mathbb{1}_n\mathbb{1}_n^{\mathsf{T}}/n)\delta(k)\|_2}{\|\delta(k)\|_2} = \sup_{y\neq\mathbb{0}_n} \frac{\|(A - \mathbb{1}_n\mathbb{1}_n^{\mathsf{T}}/n)y\|_2}{\|y\|_2},
\end{aligned}
$$

where the last equality follows from $(A - \mathbb{1}_n\mathbb{1}_n^{\mathsf{T}}/n)\mathbb{1}_n = \mathbb{0}_n$.

The equality $r_{\text{asym}}(A) = \rho(A - \mathbb{1}_n\mathbb{1}_n^{\mathsf{T}}/n)$ is a consequence of the error dynamics in Lemma 10.3, statement (ii).

Next, note that $\rho(A) = 1$ is a simple eigenvalue and A is semi-convergent. Hence, by Exercise E2.2 on the Jordan normal form of A, there exists a nonsingular T such that

$$
A = T \begin{bmatrix} 1 & \mathbb{0}_{n-1}^{\mathsf{T}} \\ \mathbb{0}_{n-1} & B \end{bmatrix} T^{-1},
$$

where $B \in \mathbb{R}^{(n-1)\times(n-1)}$ is convergent, that is, $\rho(B) < 1$. Moreover we know $\rho_{\text{ess}}(A) = \rho(B)$.

Usual properties of similarity transformations imply

$$
A^k = T \begin{bmatrix} 1 & \mathbb{0}_{n-1}^{\mathsf{T}} \\ \mathbb{0}_{n-1} & B^k \end{bmatrix} T^{-1}, \qquad \implies \qquad \lim_{k\to\infty} A^k = T \begin{bmatrix} 1 & \mathbb{0}_{n-1}^{\mathsf{T}} \\ \mathbb{0}_{n-1} & \mathbb{0}_{(n-1)\times(n-1)} \end{bmatrix} T^{-1}.
$$

Because A is doubly-stochastic and primitive, we know $\lim_{k\to\infty} A^k = \mathbb{1}_n\mathbb{1}_n^{\mathsf{T}}/n$ so that A can be decomposed as

$$
A = \mathbb{1}_n\mathbb{1}_n^{\mathsf{T}}/n + T \begin{bmatrix} 0 & \mathbb{0}_{n-1}^{\mathsf{T}} \\ \mathbb{0}_{n-1} & B \end{bmatrix} T^{-1},
$$

and conclude with $\rho_{\text{ess}}(A) = \rho(B) = \rho(A - \mathbb{1}_n\mathbb{1}_n^{\mathsf{T}}/n)$. This concludes the proof of the equalities (10.2).

The bound (10.3) is an immediate consequence of the definition of induced norm.

Finally, we leave to the reader the proof of the bound (10.4) in Exercise E10.3. Note that the arbitrarily-small positive parameter ε is required because the eigenvalue corresponding to the essential spectral radius may have an algebraic multiplicity strictly larger than its geometric multiplicity. ∎

Note: the matrix $\mathbb{1}_n\mathbb{1}_n^{\mathsf{T}}$ is studied in Exercise E2.13 as the rank-one projection matrix J_A associated to a primitive matrix A. Using methods and results from that exercise one can generalize the treatment in this section to row-stochastic instead of doubly-stochastic matrices.

10.3   Cumulative quadratic disagreement for symmetric matrices

The previous convergence metrics (per-step convergence factor and asymptotic convergence factor) are *worst-case* convergence metrics (both are defined with a supremum operation) that are achieved only for particular initial conditions, e.g., the performance predicted by the asymptotic metric $r_{\mathrm{asym}}(A)$ is achieved when $x(0) - x_{\mathrm{final}}$ is aligned with the eigenvector associated to $\rho_{\mathrm{ess}}(A) = \rho(A - \mathbb{1}_n\mathbb{1}_n^{\mathsf{T}}/n)$.

In what follows we study and appropriate *average* and *transient* performance We consider an averaging algorithm

$$x(k + 1) = Ax(k),$$

defined by a row-stochastic matrix A and subject to random initial conditions x_0 satisfying

$$\mathbb{E}[x_0] = \mathbb{0}_n, \quad \text{and} \quad \mathbb{E}[x_0 x_0^{\mathsf{T}}] = I_n.$$

Recall the disagreement vector $\delta(k)$ defined in (10.5) and the associated disagreement dynamics

$$\delta(k + 1) = \left(A - \mathbb{1}_n\mathbb{1}_n^{\mathsf{T}}/n\right)\delta(k),$$

and observe that the initial conditions of the disagreement vector $\delta(0)$ satisfy

$$\mathbb{E}[\delta(0)] = \mathbb{0}_n \quad \text{and} \quad \mathbb{E}[\delta(0)\delta(0)^{\mathsf{T}}] = I_n - \mathbb{1}_n\mathbb{1}_n^{\mathsf{T}}/n.$$

To define an average transient and asymptotic performance of this averaging algorithm, we define the *cumulative quadratic disagreement* of the matrix A by

$$\mathcal{J}_{\mathrm{cum}}(A) = \lim_{K \to \infty} \frac{1}{n} \sum_{k=0}^{K} \mathbb{E}\left[\|\delta(k)\|_2^2\right]. \tag{10.6}$$

Theorem 10.4 (Cumulative quadratic disagreement for symmetric matrices). *The cumulative quadratic disagreement* (10.6) *of a row-stochastic, primitive, and symmetric matrix A satisfies*

$$\mathcal{J}_{\mathrm{cum}}(A) = \frac{1}{n} \sum_{\lambda \in \mathrm{spec}(A) \setminus \{1\}} \frac{1}{1 - \lambda^2}.$$

Proof. Pick a terminal time $K \in \mathbb{N}$ and define $\mathcal{J}_K(A) = \frac{1}{n} \sum_{k=0}^{K} \mathbb{E}\left[\|\delta(k)\|_2^2\right]$. From the definition (10.6) and the disagreement dynamics, we compute

$$\mathcal{J}_K(A) = \frac{1}{n} \sum_{k=0}^{K} \mathrm{trace}\left(\mathbb{E}\left[\delta(k)\delta(k)^{\mathsf{T}}\right]\right)$$

$$= \frac{1}{n} \sum_{k=0}^{K} \mathrm{trace}\left(\left(A - \mathbb{1}_n\mathbb{1}_n^{\mathsf{T}}/n\right)^k \mathbb{E}\left[\delta(0)\delta(0)^{\mathsf{T}}\right] \left(\left(A - \mathbb{1}_n\mathbb{1}_n^{\mathsf{T}}/n\right)^k\right)^{\mathsf{T}}\right)$$

$$= \frac{1}{n} \sum_{k=0}^{K} \mathrm{trace}\left(\left(A - \mathbb{1}_n\mathbb{1}_n^{\mathsf{T}}/n\right)^k \left(\left(A - \mathbb{1}_n\mathbb{1}_n^{\mathsf{T}}/n\right)^k\right)^{\mathsf{T}}\right).$$

Because A is symmetric, also the matrix $A - \mathbb{1}_n\mathbb{1}_n^\mathsf{T}/n$ is symmetric and can be diagonalized as $A - \mathbb{1}_n\mathbb{1}_n^\mathsf{T}/n = Q\Lambda Q^\mathsf{T}$, where Q is orthonormal and Λ is a diagonal matrix whose diagonal entries are the elements of $\mathrm{spec}\big(A - \mathbb{1}_n\mathbb{1}_n^\mathsf{T}/n\big) = \{0\} \cup \mathrm{spec}(A) \setminus \{1\}$. It follows that

$$
\begin{aligned}
\mathcal{J}_K(A) &= \frac{1}{n}\sum_{k=0}^K \mathrm{trace}\Big(Q\Lambda^k Q^\mathsf{T}\big(Q\Lambda^k Q^\mathsf{T}\big)^\mathsf{T}\Big) \\
&= \frac{1}{n}\sum_{k=0}^K \mathrm{trace}\big(\Lambda^k \cdot \Lambda^k\big) &&\text{(because } \mathrm{trace}(AB) = \mathrm{trace}(BA)) \\
&= \frac{1}{n}\sum_{k=0}^K \sum_{\lambda\in\mathrm{spec}(A)\setminus\{1\}} \lambda^{2k} \\
&= \frac{1}{n}\sum_{\lambda\in\mathrm{spec}(A)\setminus\{1\}} \frac{1 - \lambda^{2(K-1)}}{1 - \lambda^2}. &&\text{(because of the geometric series)}
\end{aligned}
$$

The formula for $\mathcal{J}_{\mathrm{cum}}$ follows from taking the limit as $K \to \infty$ and recalling that A primitive implies $\rho_{\mathrm{ess}}(A) < 1$. ∎

Note: All eigenvalues of A appear in the computation of the cumulative quadratic disagreement (10.6), not only the dominant eigenvalue as in the asymptotic convergence factor.

10.4 Circulant network examples and scalability analysis

In general it is difficult to compute explicitly the second largest eigenvalue magnitude for an arbitrary matrix. There are some graphs with *constant* essential spectral radius, independent of the network size n. For example, a complete graph with identical weights and doubly stochastic adjacency matrix $A = \mathbb{1}_n\mathbb{1}_n^\mathsf{T}/n$ has $\rho_{\mathrm{ess}}(A) = 0$. In this case, the associated averaging algorithm converges in a single step.

Next, we present an interesting family of examples where all eigenvalues are known. Recall the cyclic balancing problem from Section 1.4, where each bug feels an attraction towards the closest counterclockwise and clockwise neighbors and Exercise E4.17 on circulant matrices. Given the angular distances between bugs $d_i = \theta_{i+1} - \theta_i$, for $i \in \{1, \dots, n\}$ (with the usual convention that $d_{n+1} = d_1$ and $d_0 = d_n$), the closed-loop system is $d(k + 1) = A_{n,\kappa} d(k)$, where $\kappa \in [0, 1/2[$, and

$$
A_{n,\kappa} = \begin{bmatrix}
1 - 2\kappa & \kappa & 0 & \cdots & 0 & \kappa \\
\kappa & 1 - 2\kappa & \kappa & \ddots & \ddots & 0 \\
0 & \kappa & 1 - 2\kappa & \ddots & \ddots & \vdots \\
\vdots & \ddots & \ddots & \ddots & \ddots & 0 \\
0 & \ddots & & \ddots & \kappa & 1 - 2\kappa & \kappa \\
\kappa & 0 & \cdots & 0 & \kappa & 1 - 2\kappa
\end{bmatrix}.
$$

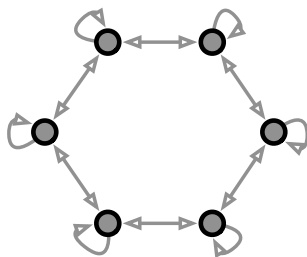This matrix is circulant, that is, each row-vector is equal to the preceding row-vector rotated one element

Figure 10.2: Digraph associated to the circulant matrix $A_{n,\kappa}$, for $n = 6$.

to the right. The associated digraph is illustrated in the Figure 10.2. From Exercise E4.17, the eigenvalues of $A_{n,\kappa}$ can be computed to be (not ordered in magnitude)

$$\lambda_i = 2\kappa \cos \frac{2\pi(i - 1)}{n} + (1 - 2\kappa), \quad \text{for } i \in \{1, \ldots, n\}. \tag{10.7}$$

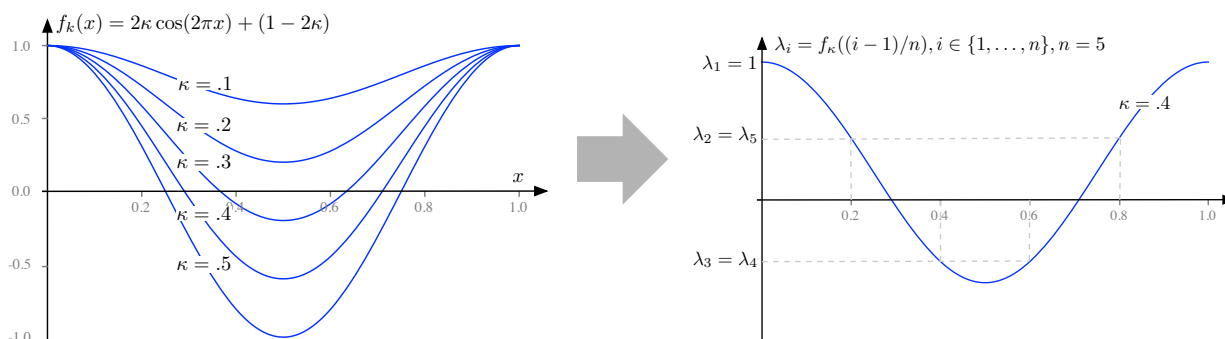An illustration is given in Figure 10.3. For n even (similar results hold for n odd), plotting the eigenvalues

Figure 10.3: The eigenvalues of $A_{n,\kappa}$ as given in equation (10.7). The left figure illustrate also the case of $\kappa = .5$, even if that value is strictly outside the allowed range $\kappa \in [0, .5[$.

on the segment $[-1, 1]$ shows that

$$\rho_{\text{ess}}(A_{n,\kappa}) = \max\{|\lambda_2|, |\lambda_{n/2+1}|\},$$

where

$$\lambda_2 = 2\kappa \cos \frac{2\pi}{n} + (1 - 2\kappa), \text{ and } \lambda_{n/2+1} = 1 - 4\kappa.$$

If we fix $\kappa \in]0, 1/2[$ and consider sufficiently large values of n, then $|\lambda_2| > |\lambda_{n/2+1}|$. In the limit of large graphs $n \to \infty$, the Taylor expansion $\cos(x) = 1 - x^2/2 + \mathcal{O}(x^4)$ leads to

$$\rho_{\text{ess}}(A_{n,\kappa}) = 1 - 4\pi^2 \kappa \frac{1}{n^2} + \mathcal{O}\left(\frac{1}{n^4}\right). \tag{10.8}$$

Note that $\rho_{\text{ess}}(A_{n,\kappa}) < 1$ for any n, but the separation from $\rho_{\text{ess}}(A_{n,\kappa})$ to 1, called the *spectral gap*, shrinks with $1/n^2$.

In summary, this discussion leads to the broad statement that certain large-scale graphs have slow convergence factors.

10.5 Appendix: Accelerated consensus algorithm

The averaging algorithm $x(k + 1) = Ax(k)$ may converge slowly as seen in Section 10.4 due to a large $\rho_{\text{ess}}(A)$. In this section we propose a simple modification of averaging that is known to be faster. The *accelerated consensus algorithm* is defined by

$$x(k + 1) = \beta A \, x(k) + (1 - \beta) \, x(k - 1), \qquad \text{for } k \in \mathbb{Z}_{\geq 0}, \tag{10.9}$$

where the initial conditions are $x(0) = x(-1) := x_0$, the matrix $A \in \mathbb{R}^{n \times n}$ is symmetric, primitive, and row-stochastic, and $\beta \in \mathbb{R}$ is a parameter to be chosen.

This iteration has some basic properties. We define the iteration matrix

$$T_\beta = \begin{bmatrix} \beta A & (1 - \beta)I_n \\ I_n & \mathbb{0}_{n \times n} \end{bmatrix} \in \mathbb{R}^{2n \times 2n}.$$

One can show that $T_\beta \mathbb{1}_{2n} = \mathbb{1}_{2n}$ for all β, and that T_β is semiconvergent if and only if $\rho_{\text{ess}}(T_\beta) < 1$. Moreover, similar to the result in (10.4) one can show that, for an appropriate value of β, the asymptotic convergence factor for this accelerated iteration is equal to $\rho_{\text{ess}}(T_\beta)$. Accordingly, in what follows, we optimize the convergence speed of the algorithm by minimizing $\rho_{\text{ess}}(T_\beta)$ with respect to β. We formally state these results and more in the following theorem.

Theorem 10.5 (Convergence and optimization of the accelerated consensus algorithm). *Consider the accelerated consensus algorithm* (10.9) *with* $x(0) = x(-1) = x_0$, $A \in \mathbb{R}^{n \times n}$ *symmetric, primitive, and row-stochastic matrix, and* $\beta \in \mathbb{R}$. *The following statements hold:*

(i) for all $\beta \in \mathbb{R}$, *the set of fixed points of* T_β *is* $\{\alpha \mathbb{1}_{2n} \mid \alpha \in \mathbb{R}\}$ *and, if* $\lim_{k \to \infty} x(k)$ *exists, then it is equal to* $\text{average}(x_0)\mathbb{1}_n$;

(ii) the following conditions are equivalent:

a) T_β *is semi-convergent,*

b) $\rho_{\text{ess}}(T_\beta) < 1$, *and*

c) $\beta \in (0, 2)$;

(iii) for $\beta \in (0, 2)$, *along the accelerated consensus iteration* (10.9)

$$\left\| x(k) - \text{average}(x_0)\mathbb{1}_n \right\|_2 \leq c_\varepsilon (\rho_{\text{ess}}(T_\beta) + \varepsilon)^k \left\| x(0) - \text{average}(x_0)\mathbb{1}_n \right\|_2,$$

where $\varepsilon > 0$ *is an arbitrarily small constant and* c_ε *is a sufficiently large constant independent of* x_0;

(iv) *the optimal convergence rate of the accelerated consensus algorithm is*

$$\min_{\beta \in (0,2)} \rho_{\text{ess}}(T_\beta) = \frac{\rho_{\text{ess}}(A)}{1 + \sqrt{1 - \rho_{\text{ess}}(A)^2}}, \tag{10.10}$$

which is obtained at

$$\beta^* = \operatorname{argmin}_{\beta \in (0,2)} \rho_{\text{ess}}(T_\beta) = \frac{2}{1 + \sqrt{1 - \rho_{\text{ess}}(A)^2}} \in (1, 2). \tag{10.11}$$

Note: A key advantage of the accelerated consensus algorithm is it is faster than standard averaging in two senses: First, it is immediate to see that $\rho_{\text{ess}}(T_{\beta^*}) = \frac{\rho_{\text{ess}}(A)}{1 + \sqrt{1 - \rho_{\text{ess}}(A)^2}} < \rho_{\text{ess}}(A)$. Second, Exercise E10.10 shows that performance improves also in its asymptotic order; for example, for averaging algorithms over circulant matrices, the spectral gap of order $1/n$ instead of order $1/n^2$. One important drawback of the accelerated consensus algorithm is that computation of optimal gain requires knowledge of the essential spectral radius of A.

Proof of Theorem 10.5. Regarding statement (i), we let $x^\star = \lim_{k \to \infty} x(k)$ and take the limit in both left and right hand side of the accelerated consensus algorithm (10.9) to obtain $x^* = \beta A x^* + (1 - \beta) x^*$, that is, after simple manipulations $x^\star = A x^\star$. Under the given assumptions on the matrix A and by employing the Perron-Frobenius Theorem, we obtain that $x^\star = \alpha \mathbb{1}_n$ for some $\alpha \in \mathbb{R}$. Observe also that $x(t) = \alpha \mathbb{1}_n$ is a conserved quantity for the accelerated consensus algorithm (10.9). Thus, when left-multiplying $x(t) = \alpha \mathbb{1}_n$ by $\mathbb{1}_n^\mathsf{T}$ and evaluating the result for $t = 0$, we obtain $\alpha = \operatorname{average}(x_0)$. This concludes the proof of statement (i).

Next, we prove statement (ii). We start by analyzing the matrix T_β with methods similar to those adopted for the second-order Laplacian flow in Section 7.4. The symmetric matrix A can be expressed as $A = U\Lambda U^\mathsf{T}$, where U is a unitary matrix and $\Lambda = \operatorname{diag}(\{\lambda_i\}_{i=1}^n)$ collects the eigenvalues of the matrix A. A similarity transformation with the matrix U leads us to

$$\begin{bmatrix} U & 0 \\ 0 & U \end{bmatrix}^\mathsf{T} T_\beta \begin{bmatrix} U & 0 \\ 0 & U \end{bmatrix} = \begin{bmatrix} U & 0 \\ 0 & U \end{bmatrix}^\mathsf{T} \begin{bmatrix} \beta A & (1-\beta)I_n \\ I_n & 0 \end{bmatrix} \begin{bmatrix} U & 0 \\ 0 & U \end{bmatrix} = \begin{bmatrix} \beta \Lambda & (1-\beta)I_n \\ I_n & 0 \end{bmatrix}.$$

By appropriately permuting the entries of this matrix, we arrive at

$$\Gamma = \begin{bmatrix} \Gamma_1 & 0 & \dots & 0 \\ 0 & \Gamma_2 & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & \Gamma_n \end{bmatrix}, \qquad \text{where } \Gamma_i = \begin{bmatrix} \beta \lambda_i & 1-\beta \\ 1 & 0 \end{bmatrix}, \ i \in \{1, \dots, n\}.$$

Note that the similarity transformation via the matrix U and the permutation (which is itself a similarity transformation) the spectra of Γ and T_β are identical. We can, hence, analyze the matrix Γ to investigate the convergence rates. For a given index $i \in \{1, \dots, n\}$, the eigenvalues of Γ_i are the roots of

$$v_i^2 - (\beta \lambda_i) v_i + \beta - 1 = 0, \tag{10.12}$$

which are given by

$$v_{1,2_i} = \frac{\beta\,\lambda_i \pm \sqrt{\beta^2\,\lambda_i^2 - 4\beta + 4}}{2}\,. \tag{10.13}$$

For the system to converge to steady-state consensus, all eigenvalues $v_{1,2_i}$, $i \in \{1,\ldots,n\}$, should lie within the unit disc, with only one eigenvalue on the unit circle. For Γ_n with $\lambda_n = 1$, we note that the eigenvalues are $\{1,\, \beta - 1\}$. Therefore, a necessary convergence condition for $\beta \in \mathbb{R}$ is

$$-1 < \beta - 1 < 1 \quad \text{or} \quad 0 < \beta < 2\,. \tag{10.14}$$

For the other block matrices Γ_i, $i \in \{2,\ldots,n\}$, the eigenvalues are given by equation (10.13), the sum of the roots by $(v_{1_i} + v_{2_i} = \beta\,\lambda_i)$, and the product of the roots by $(v_{1_i} \cdot v_{2_i} = \beta - 1)$. We consider the following cases:

a) Assume Γ_i has real-valued roots: For the roots to lie within the unit circle, we require $|v_{1_i}| < 1$, $|v_{2,i}| < 1$, and $v_{1_i}^2 + v_{2_i}^2 < 2$ for all $i \in \{1,\ldots,n\}$. Regarding the latter:

$$\begin{aligned}
v_{1_i}^i + v_{2_i}^2 &= (v_{1_i} + v_{2_i})^2 - 2 \cdot v_{1_i} \cdot v_{2_i} < 2 \\
\iff \quad & \beta^2\lambda_i^2 - 2\beta + 2 < 2 \\
\iff \quad & \beta^2 - 2\beta < 0 \qquad\qquad (\text{as } |\lambda_i| < 1) \\
\iff \quad & \beta(\beta - 2) < 0 \text{ or } \beta \in (0,2).
\end{aligned} \tag{10.15}$$

We now verify $|v_{1_i}| < 1$, $|v_{2,i}| < 1$. For Γ_n with $\lambda_n = 1$, the eigenvalues are $\{v_{1_n}, v_{2_n}\} = \{1,\, \beta - 1\}$ and thus $|v_{1_n}| < 1$, $|v_{2,n}| < 1$. For Γ_i, $i \in \{1,\ldots,n-1\}$, with $|\lambda_i| < 1$, it can explicitly calculated that $|v_{1_i}| < 1$, $|v_{2,i}| < 1$ if $\beta \in (0,2)$.

b) Assume Γ_i has complex conjugate roots: As the coefficients of equation (10.12) are all real (β is real and λ_i is real as the matrix A is symmetric), the complex-conjugate roots have the same magnitude. We require the magnitudes to be strictly less than 1:

$$|v_{1_i}| = |v_{2_i}| = \sqrt{\beta - 1} < 1 \quad \implies \quad 0 < (\beta - 1) < 1 \text{ or } \beta \in (0,2). \tag{10.16}$$

Equations (10.14), (10.15), and (10.16) together imply that the iteration converges for values of $\beta \in (0,2)$. This concludes the proof of statement (ii).

Regarding statement (iii), it is an immediate consequence of Exercise E10.3 and some ad-hoc bounds. We leave it to the reader to fill out the details.

Finally, we prove statement (iv). In order to minimize the modulus of the eigenvalues of Γ_i, we choose β such that the discriminant in the expression (10.13) becomes zero:

$$\beta^2\,\lambda_i^2 - 4\beta + 4 = 0\,. \tag{10.17}$$

Let us keep the index $i \in \{1,\ldots,n-1\}$ fixed. Two possible values of β arise from equation (10.17):

$$\beta \in \left\{ \frac{2}{1 + \sqrt{1 - \lambda_i^2}},\ \frac{2}{1 - \sqrt{1 - \lambda_i^2}} \right\},$$

Because the second root may lead to a value of β outside the existence interval $(0, 2)$, we restrict ourselves to the optimal selection (for the index i) of the gain β as

$$\beta = \frac{2}{1 + \sqrt{1 - \lambda_i^2}} \, .$$

Among all choices of β for different indices $i \in \{1, \ldots, n-1\}$, we note that $\beta^\star = 2/(1 + \sqrt{1 - \rho_{\mathrm{ess}}(A)^2})$ as in equation (10.11) is the optimal choice to minimize the maximum magnitude of $|v_{1,2_i}|$ for $i \in \{1, \ldots, n-1\}$. Furthermore, since $1 > \rho_{\mathrm{ess}}(A) \geq 0$, we have $2 > \beta^\star \geq 1$, and thus the magnitudes of all eigenvalues of Γ is strictly less than 1, except for the the eigenvalue at 1. The magnitudes of the other eigenvalues of Γ for $\beta = \beta^\star$ are

$$\Big\{ \underbrace{1, |\beta^\star - 1|\}}_{\Gamma_n}, \ \underbrace{\{|\sqrt{\beta^\star - 1}|, \ |\sqrt{\beta^\star - 1}|\}}_{\Gamma_{n-1}}, \ \underbrace{\{|v_{1_{n-2}}(\beta^\star)|, \ |v_{2_{n-2}}(\beta^\star)|}_{\Gamma_{n-2}}, \ \ldots, \ \underbrace{\{|v_{1_1}(\beta^\star)|, \ |v_{2_1}(\beta^\star)|\}}_{\Gamma_1} \Big\} .$$

(10.18)

Furthermore, it can be verified that for $\beta = \beta^\star$ we have identical magnitudes $|v_{1_i}(\beta^\star)| = |v_{2_i}(\beta^\star)| = \sqrt{\beta^\star - 1}$ for all $i \in \{1, \ldots, n - 2\}$. Finally, note that $\sqrt{\beta^\star - 1} \geq |\beta^\star - 1| = \beta^\star - 1$ so that

$$\rho_{\mathrm{ess}}(T_{\beta^*}) = \rho_{\mathrm{ess}}(\Gamma) = \sqrt{\beta^\star - 1} = \frac{\rho_{\mathrm{ess}}(A)}{1 + \sqrt{1 - \rho_{\mathrm{ess}}(A)^2}} < \rho_{\mathrm{ess}}(A).$$

∎

10.6   Appendix: Design of fastest distributed averaging

We are interested in optimization problems of the form:

minimize $r_{\mathrm{asym}}(A)$ or $r_{\mathrm{step}}(A)$

subject to A compatible with a digraph G, doubly-stochastic and primitive

where A is compatible with G if its only non-zero entries correspond to the edges E of the graph. In other words, if $E_{ij} = \mathrm{e}_i \mathrm{e}_j^\mathsf{T}$ is the matrix with entry (i, j) equal to one and all other entries equal to zero, then $A = \sum_{(i,j) \in E} a_{ij} E_{ij}$ for arbitrary weights $a_{ij} \in \mathbb{R}$. We refer to such problems as fastest distributed averaging (FDAs) problems.

Note: In what follows, we remove the constraint $A \geq 0$ to widen the set of matrices of interest. Accordingly, we remove the constraint of A being primitive. Convergence to average consensus is guaranteed by (1) achieving convergence factors less than 1, (2) subject to row-sums and column-sums equal to 1.

Problem 10.6 (Asymmetric FDA with asymptotic convergence factor).

minimize $\rho\big(A - \mathbb{1}_n \mathbb{1}_n^\mathsf{T}/n\big)$

subject to $A = \sum_{(i,j) \in E} a_{ij} E_{ij}, \ \ A\mathbb{1}_n = \mathbb{1}_n, \ \ \mathbb{1}_n^\mathsf{T} A = \mathbb{1}_n^\mathsf{T}$

The asymmetric FDA is a hard optimization problem. Even though the constraints are linear, the objective function, i.e., the spectral radius of a matrix, is not convex (and, additionally, not even Lipschitz continuous).

Problem 10.7 (Asymmetric FDA with per-step convergence factor).

$$\text{minimize } \left\| A - \mathbb{1}_n \mathbb{1}_n^\mathsf{T}/n \right\|_2$$
$$\text{subject to } A = \sum_{(i,j)\in E} a_{ij} E_{ij}, \;\; A\mathbb{1}_n = \mathbb{1}_n, \;\; \mathbb{1}_n^\mathsf{T} A = \mathbb{1}_n^\mathsf{T}$$

Problem 10.8 (Symmetric FDA problem).

$$\text{minimize } \rho\big(A - \mathbb{1}_n \mathbb{1}_n^\mathsf{T}/n\big)$$
$$\text{subject to } A = \sum_{(i,j)\in E} a_{ij} E_{ij}, \;\; A = A^\mathsf{T}, \;\; A\mathbb{1}_n = \mathbb{1}_n$$

Recall here that $A = A^\mathsf{T}$ implies $\rho(A) = \|A\|_2$.

Both Problems 10.7 and 10.8 are convex and can be rewritten as so-called *semi-definite programs* (SDPs); see (Xiao and Boyd, 2004). An SDP is an optimization problem where (1) the variable is a positive semidefinite matrix, (2) the objective function is linear, and (3) the constraints are affine equations. SDPs can be efficiently solved by software tools such as CVX; see (Grant and Boyd, 2016).

10.7 Historical notes and further reading

The main ideas in Sections 10.1 and 10.2 are taken from (Olshevsky and Tsitsiklis, 2009; Garin and Schenato, 2010; Fagnani, 2014).

A recent breakthrough in achieving linear time average consensus on fixed graphs (not reviewed here) is given by Olshevsky (2014).

The cumulative quadratic disagreement in Section 10.3 is taken from (Carli et al., 2009). Theorem 10.4 may be extended to the setting of normal matrices, as opposed to symmetric, as illustrated in (Carli et al., 2009); it is not known how to compute the cumulative quadratic disagreement for arbitrary doubly-stochastic primitive matrices.

Regarding Section 10.4, for more results on the study of circulant matrices and on the elegant settings of Cayley graphs we refer to (Davis, 1979; Carli et al., 2008b).

The accelerated consensus algorithm (10.9) is rooted in momentum methods for optimization (Polyak, 1964), and it has been applied to averaging algorithms for example in (Muthukrishnan et al., 1998; Bof et al., 2016).

10.8   Exercises

E10.1   **Induced norm of certain doubly stochastic matrices.** Assume A is doubly stochastic, primitive and has a strictly-positive diagonal. Show that
$$r_{\text{step}}(A) = \|A - \mathbb{1}_n \mathbb{1}_n^{\mathsf{T}}/n\|_2 < 1.$$

E10.2   **Spectrum of $A - \mathbb{1}_n \mathbb{1}_n^{\mathsf{T}}/n$.** Consider a matrix A doubly stochastic, primitive and symmetric. Assume $\lambda_1 \geq \cdots \geq \lambda_n$ are its real eigenvalue with corresponding orthonormal eigenvectors $v_1, \ldots, v_n$. Show that the matrix $A - \mathbb{1}_n \mathbb{1}_n^{\mathsf{T}}/n$ has eigenvalues $0, \lambda_2 \geq \cdots \geq \lambda_n$ with eigenvectors $v_1, \ldots, v_n$.

E10.3   **Bounds on the norm of a matrix power.** Given a matrix $B \in \mathbb{R}^{n \times n}$ and an index $k \in \mathbb{N}$, show that

(i)   there exists $c > 0$ such that
$$\|B^k\|_2 \leq c\, k^{n-1} \rho(B)^k,$$

(ii)   for all $\varepsilon > 0$, there exists $c_\varepsilon > 0$ such that
$$\|B^k\|_2 \leq c_\varepsilon (\rho(B) + \varepsilon)^k.$$

Hint: Adopt the Jordan normal form

E10.4   **Spectral gap of regular cycle graphs.** A k-*regular cycle graph* is an undirected cycle graph with n-nodes each connected to itself and its $2k$ nearest neighbors with a uniform weight equal to $1/(2k + 1)$. The associated doubly-stochastic adjacency matrix $A_{n,k}$ is a circulant matrix with first row given by
$$A_{n,k}(1,:) = \begin{bmatrix} \frac{1}{2k+1} & \cdots & \frac{1}{2k+1} & 0 & \cdots & 0 & \frac{1}{2k+1} & \cdots & \frac{1}{2k+1} \end{bmatrix}.$$

Using the results in Exercise E4.17, compute

(i)   the eigenvalues of $A_{n,k}$ as a function of n and k;
(ii)   the limit of the spectral gap for fixed k as $n \to \infty$; and
(iii)   the limit of the spectral gap for $2k = n - 1$ as $n \to \infty$.

E10.5   **Properties of the spectral radius.** For any $A \in \mathbb{C}^{n \times n}$ and any matrix norm, show

(i) $\rho(A) \leq \|A\|$, and
(ii) $\rho(A) \leq \|A^k\|^{1/k}$ for all k,
(iii) $\rho(A) = \lim_{k \to \infty} \|A^k\|^{1/k}$.

Next, for any $A \in \mathbb{C}^{n \times n}$, let $|A|$ denote the matrix with entries $|a_{ij}|$, and for any real matrices B, C, let $B \leq C$ mean $b_{ij} \leq c_{ij}$ for each i and j. Show

(iv)   if $|A| \leq B$, then $\rho(A) \leq \rho(|A|) \leq \rho(B)$.

Hint: Peruse (Meyer, 2001, Chapter 7).

E10.6 $\mathcal{H}_2$ **performance of balanced averaging in continuous time.** Consider the continuous-time averaging dynamics with disturbance
$$\dot{x}(t) = -Lx(t) + w(t),$$
where $L = L^{\mathsf{T}}$ is the Laplacian matrix of an undirected and connected graph and $w(t)$ is an exogenous disturbance input signal. Pick a matrix $Q \in \mathbb{R}^{p \times n}$ satisfying $Q\mathbb{1}_n = \mathbb{0}_p$ and define the output signal $y(t) = Qx(t) \in \mathbb{R}^p$ as the solution from zero initial conditions $x(0) = \mathbb{0}_n$. Define the *system $\mathcal{H}_2$ norm* from w to y by
$$\|\mathcal{H}\|_2^2 = \int_0^\infty y(t)^{\mathsf{T}} y(t)dt = \int_0^\infty x(t)^{\mathsf{T}} Q^{\mathsf{T}} Q x(t)dt = \text{trace}\left(\int_0^\infty H(t)^{\mathsf{T}} H(t)dt\right),$$
where $H(t) = Qe^{-Lt}$ is the so-called *impulse response matrix*.

(i) Show $\|\mathcal{H}\|_2 = \sqrt{\mathrm{trace}(P)}$, where P is the solution to the so-called *Lyapunov equation*

$$LP + PL = Q^\mathsf{T}Q. \tag{E10.1}$$

(ii) Show $\|\mathcal{H}\|_2 = \sqrt{\mathrm{trace}\left(L^\dagger Q^\mathsf{T}Q\right)/2}$, where $L^\dagger$ is the pseudoinverse of L.

(iii) Define *short-range* and *long-range output matrices* Q_{sr} and Q_{lr} by $Q_{\mathrm{sr}}^\mathsf{T}Q_{\mathrm{sr}} = L$ and $Q_{\mathrm{lr}}^\mathsf{T}Q_{\mathrm{lr}} = I_n - \frac{1}{n}\mathbb{1}_n\mathbb{1}_n^\mathsf{T}$, respectively. Show:

$$\|\mathcal{H}\|_2^2 = \begin{cases} n - 1, & \text{for } Q = Q_{\mathrm{sr}}, \\ \displaystyle\sum_{i=2}^{n} \frac{1}{\lambda_i(L)}, & \text{for } Q = Q_{\mathrm{lr}}. \end{cases}$$

Hint: *The $\mathcal{H}_2$ norm has several interesting interpretations, including the total output signal energy in response to a unit impulse input or the root mean square of the output signal in response to a white noise input with identity covariance. You may find useful Theorem 7.4 and Exercise E6.8.*

E10.7 **Convergence rate for the Laplacian flow.** Consider a weight-balanced, strongly connected digraph G with self-loops, degree matrices $D_{\mathrm{out}} = D_{\mathrm{in}} = I_n$, doubly-stochastic adjacency matrix A, and Laplacian matrix L. Consider the associated Laplacian flow

$$\dot{x}(t) = -Lx(t).$$

For $x_{\mathrm{ave}} := \frac{\mathbb{1}_n^\mathsf{T}x(0)}{n}$, define the *disagreement vector* by $\delta(t) = x(t) - x_{\mathrm{ave}}\mathbb{1}_n$.

(i) Show that the average $t \mapsto \frac{\mathbb{1}_n^\mathsf{T}x(t)}{n}$ is conserved and that, consequently, $\mathbb{1}_n^\mathsf{T}\delta(t) = 0$ for all $t \geq 0$.

(ii) Derive the matrix E describing the disagreeement dynamics

$$\dot{\delta}(t) = E\delta(t).$$

(iii) Describe the spectrum $\mathrm{spec}(E)$ of E as a function of the spectrum $\mathrm{spec}(A)$ of the doubly-stochastic adjacency matrix A associated with G. Show that $\mathrm{spec}(E)$ has a simple eigenvalue at $\lambda = 0$ with corresponding normalized eigenvector $v_1 := \mathbb{1}_n/\sqrt{n}$.

(iv) The Jordan form J of E can be described as follows

$$E = P \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & J_2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & J_m \end{bmatrix} P^{-1} =: \begin{bmatrix} c_1 & \tilde{C} \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & \tilde{J} \end{bmatrix} \begin{bmatrix} r_1 \\ \tilde{R} \end{bmatrix},$$

where c_1 is the first column of P and r_1 is the first row of P^{-1}. Show that

$$\delta(t) = \tilde{C}\exp(\tilde{J}t)\tilde{R}\delta(0).$$

(v) Use statements (iii) and (iv) to show that, for all $\varepsilon > 0$, there exists $C_\varepsilon > 0$ satisfying

$$\|\delta(t)\| \leq C_\varepsilon(e^\mu + \varepsilon)^t\|\delta(0)\|,$$

where $\mu = \max\{\Re(\lambda) - 1 \mid \lambda \in \mathrm{spec}(A)\setminus\{1\}\} < 0$. Show that, if $A = A^\mathsf{T}$, then $\mu \leq \rho_{\mathrm{ess}}(A) - 1$.

Hint: *Use arguments similar to those in Exercise E10.3 and in the proof of Theorem 7.4.*

E10.8 **Convergence factors in digraphs with equal out-degree.** Consider the unweighted digraphs in the figure below with their associated discrete-time averaging systems $x(t+1) = A_1 x(t)$ and $x(t+1) = A_2 x(t)$. For which digraph is the worst-case discrete-time consensus protocol (i.e., the evolution starting from the worst-case initial condition) guaranteed to converge faster? Assign to each edge the same weight equal to $\frac{1}{3}$.
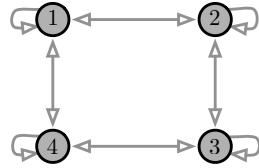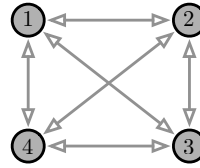
(a) Digraph 1   (b) Digraph 2

E10.9 **Convergence estimates.** Consider a discrete-time averaging system with 4 agents, state variable $x \in \mathbb{R}^4$, dynamics $x(k+1) = Ax(k)$, and averaging matrix $A = \sum_{i=1}^{3} \alpha_i v_i v_i^\mathsf{T} \in \mathbb{R}^{4\times4}$ with

$$\alpha_1 = 1, \alpha_2 = \frac{1}{2}, \alpha_3 = \frac{1}{4}, \quad v_1 = \frac{1}{2}\begin{bmatrix}1\\1\\1\\1\end{bmatrix}, v_2 = \frac{1}{\sqrt{2}}\begin{bmatrix}0\\1\\0\\-1\end{bmatrix}, v_3 = \frac{1}{\sqrt{2}}\begin{bmatrix}1\\0\\-1\\0\end{bmatrix}.$$

(i) Verify A is row-stochastic, symmetric and primitive.

(ii) Suppose $x(0) = [0, 8, 2, 2]^\mathsf{T}$. It is possible that $x(3) = [4, 3, 2, 3]^\mathsf{T}$?

E10.10 **Scalability of accelerated consensus.**

(i) Prove the following series expansion around $x = 0$:

$$f(x) = \frac{1 - x}{1 + \sqrt{1 - (1-x)^2}} = 1 - \sqrt{2}\sqrt{x} + o(x).$$

Next, consider a sequence of increasing dimension row-stochastic matrices $\{A_n \in \mathbb{R}^{n \times n}\}_{n \in \mathbb{N}}$, and the corresponding accelerated consensus algorithms with sequence of optimal iteration matrices $\{T_{\beta^*,n} \in \mathbb{R}^{2n \times 2n}\}_{n \in \mathbb{N}}$.

(ii) Prove that, if $\rho_{\text{ess}}(A_n) = 1 - g(n)$ with $g(n) = o(n)$ as $n \to \infty$, then the following series expansion holds as $n \to \infty$:

$$\rho_{\text{ess}}(T_{\beta^*,n}) = 1 - \sqrt{2}\sqrt{g(n)} + o(g(n)).$$

(iii) Show that, for circulant matrices $\{A_n\}_n$ with spectral radius given in equation (10.8) in Section 10.4, there exists a constant c such that the accelerated consensus algorithm satisfies

$$\rho_{\text{ess}}(T_{\beta^*,n}) = 1 - c\frac{1}{n} + \mathcal{O}\left(\frac{1}{n^2}\right).$$

Time-varying Averaging Algorithms

In this chapter we discuss time-varying averaging systems, that is, systems in which the row-stochastic matrix is a function of time. We provide sufficient conditions on the sequence of digraphs associated to the sequence of row-stochastic matrices for consensus to be achieved. We focus mainly on the discrete-time setting, but present the main result also for continuous-time systems.

11.1 Examples and models of time-varying discrete-time algorithms

In time-varying or time-varying algorithms the averaging row-stochastic matrix is not constant throughout time, but instead changes values and, possibly, switches among a finite number of values. Here are examples of discrete-time averaging algorithms with switching matrices.

Example 11.1 (Shared Communication Channel). Given a communication digraph $G_{\text{shared-comm}}$, at each communication round, only one node can transmit to all its out-neighbors over a common bus and every receiving node will implement a single averaging step. For example, if agent j receives the message from agent i, then agent j will implement:

$$x_j^+ := \frac{1}{2}(x_i + x_j). \tag{11.1}$$

Each node is allocated a communication slot in a periodic deterministic fashion, e.g., in a *round-robin scheduling*, where the n agents are numbered and, for each i, agent i talks only at times $i, n+i, 2n+i, \ldots, kn+i$ for $k \in \mathbb{Z}_{\geq 0}$. For example, in Figure 11.1 we illustrate the communication digraph and in Figure 11.2 the resulting round-robin communication protocol.

$G_{\text{shared-comm}}$
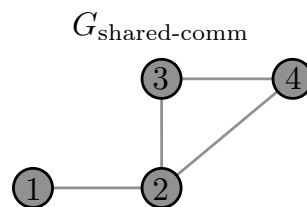
Figure 11.1: Example communication digraph

time = $1, 5, 9, \ldots$ time = $2, 6, 10, \ldots$ time = $3, 7, 11, \ldots$ time = $4, 8, 12, \ldots$
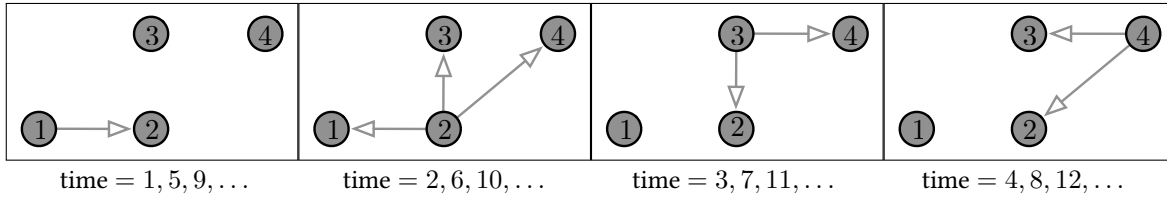
Figure 11.2: Round-robin communication protocol.

Formally, let A_i denote the averaging matrix corresponding to the transmission by agent i to its out-neighbors. With round robin scheduling, we have

$$x(n + 1) = A_n A_{n-1} \cdots A_1 x(1).$$ •

Example 11.2 (Asynchronous Execution). Imagine each node has a different clock, so that there is no common time schedule. Suppose that messages are safely delivered even if transmitting and receiving agents are not synchronized. Each time an agent wakes up, the available information from its neighbors varies. At an iteration instant for agent i, assuming agent i has new messages/information from agents $i_1, \ldots, i_m$, agent i will implement:

$$x_i^+ := \frac{1}{m+1} x_i + \frac{1}{m+1}(x_{i_1} + \cdots + x_{i_m}).$$

Given arbitrary clocks, one can consider the set of times at which one of the n agents performs an iteration. Then the system is a discrete-time averaging algorithm. It is possible to carefully characterize all possible sequences of events (who transmitted to agent i when it wakes up). •

11.2 Models of time-varying averaging algorithms

Consider a sequence of row-stochastic matrices $\{A(k)\}_{k \in \mathbb{Z}_{\geq 0}}$, or equivalently a time-varying row-stochastic matrix $k \mapsto A(k)$. The associated *time-varying averaging algorithm* is the discrete-time dynamical system

$$x(k + 1) = A(k)x(k), \quad k \in \mathbb{Z}_{\geq 0}. \tag{11.2}$$

We let $\{G(k)\}_{k \in \mathbb{Z}_{\geq 0}}$ be the sequence of weighted digraphs associated to the matrices $\{A(k)\}_{k \in \mathbb{Z}_{\geq 0}}$.

Note that $(1, \mathbb{1}_n)$ is an eigenpair for each matrix $A(k)$. Hence, all points in the consensus set $\{\alpha \mathbb{1}_n \mid \alpha \in \mathbb{R}\}$ are equilibria for the algorithm. We aim to provide conditions under which each solution converges to consensus.

We start with a useful definition, for two digraphs $G = (V, E)$ and $G' = (V', E')$, *union* of G and G' is defined by

$$G \cup G' = (V \cup V', E \cup E').$$

In what follows, we will need to compute only the union of digraphs with the same set of vertices; in that case, the graph union is essentially defined by the union of the edge sets. Some useful properties of the product of multiple row-stochastic matrices and of the unions of multiple digraphs are presented in Exercise E11.1.

11.3 Convergence over time-varying graphs connected at all times

Let us first consider the case when $A(k)$ induces an undirected, connected, and aperiodic graph $G(k)$ at each time k.

Theorem 11.3 (Convergence under connectivity at all times). *Let $\{A(k)\}_{k\in\mathbb{Z}_{\geq 0}}$ be a sequence of symmetric and doubly-stochastic matrices with associated digraphs $\{G(k)\}_{k\in\mathbb{Z}_{\geq 0}}$ so that*

(AC1) *each non-zero edge weight $a_{ij}(k)$, including the self-loops weights $a_{ii}(k)$, is larger than a constant $\varepsilon > 0$; and*

(AC2) *each graph $G(k)$ is connected.*

Then the solution to $x(k + 1) = A(k)x(k)$ converges exponentially fast to $\mathrm{average}\big(x(0)\big)\mathbb{1}_n$.

Note: Assumption (AC1) prevents the weights from becoming arbitrarily close to zero as $k \to \infty$ and, as we show below, ensures that $\rho_{\mathrm{ess}}(A(k))$ is upper bounded by a number strictly lower than 1 at every time $k \in \mathbb{Z}_{\geq 0}$. To gain some intuition into what can go wrong, consider a sequence of symmetric and doubly-stochastic averaging matrices $\{A(k)\}_{k\in\mathbb{Z}_{\geq 0}}$ with entries given by

$$A(k) = \begin{bmatrix} \exp(-1/(k+1)^\alpha) & 1 - \exp(-1/(k+1)^\alpha) \\ 1 - \exp(-1/(k+1)^\alpha) & \exp(-1/(k+1)^\alpha) \end{bmatrix}$$

for $k \in \mathbb{Z}_{\geq 0}$ and exponent $\alpha \geq 1$. These matrices fail to satisfy Assumption (AC1). For any $\alpha \geq 1$ and for k, we know the $\rho_{\mathrm{ess}}(A(k) < 1$. For any $\alpha \geq 1$ and for $k \to \infty$, this matrix converges to $A_\infty = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ with spectrum $\mathrm{spec}(A_\infty) = \{-1, +1\}$ and essential spectral radius $\rho_{\mathrm{ess}}(A_\infty) = 1$. One can show that,

  (i) for $\alpha = 1$, the convergence of $A(k)$ to A_∞ is so slow that $\{x(k)\}_k$ converges to $\mathrm{average}(x(0))\mathbb{1}_n$,

 (ii) for $\alpha > 1$, the convergence of $A(k)$ to A_∞ is so fast that $\{x(k)\}_k$ oscillates indefinitely. [1]

Proof of Theorem 11.3. At fixed n, there exist only a finite number of possible connected unweighted graphs and, for each given graph, the set of matrices with edge weights in the interval $[\varepsilon, 1]$ is compact. It is known that the following maps are continuous: the function from a matrix to its eigenvalues, the function from a complex number to its magnitude, and the function from $n - 1$ non-negative numbers to their maximum. Hence, by composition, the essential spectral radius ρ_{ess} is a continuous function of the matrix entries defined over a compact set and, therefore, it attains its maximum value. Because each essential spectral radius of each possible weighted graph is strictly less than 1, so is its maximum value. In summary, we now know that, under assumptions (AC1) and (AC2), there exists a $c \in [0, 1[$ so that $\rho_{\mathrm{ess}}(A(k)) \leq c < 1$ for all $k \in \mathbb{Z}_{\geq 0}$. Recall the notion of the disagreement vector $\delta(k) = x(k) - \mathrm{average}(x(0))\mathbb{1}_n$ and define $V(\delta) = \|\delta\|_2^2$. It is immediate to compute

$$V(\delta(k+1)) = V(A(k)\delta(k)) = \|A(k)\delta(k)\|_2^2 \leq \rho_{\mathrm{ess}}(A(k))^2\|\delta(k)\|_2^2 \leq c^2 V(\delta(k)).$$

It follows that $V(\delta(k)) \leq c^{2k}V(\delta(0))$ or $\|\delta(k)\|_2 \leq c^k\|\delta(0)\|_2$, that is, $\delta(k)$ converges to zero exponentially fast. Equivalently, as $k \to \infty$, $x(k)$ converges exponentially fast to $\mathrm{average}\big(x(0)\big)\mathbb{1}_n$. ∎

---

[1] A simplified version of this example is the scalar iteration $x(k+1) = \exp(-1/(k+1)^\alpha)x(k)$ whose solution satisfies $\log(x(k)) = -\sum_{\kappa=0}^{k-1} \frac{1}{(\kappa+1)^\alpha} + \log(x_0)$. For $\alpha = 1$, $\lim_{k\to\infty} \log(x(k))$ diverges to $-\infty$, and $\lim_{k\to\infty} x(k)$ converges to zero. Instead, for $\alpha > 1$, $\lim_{k\to\infty} \log(x(k))$ exists finite, and thus $\lim_{k\to\infty} x(k)$ does not converge to zero.

This proof is based on a positive "energy function" that decreases along the system's evolutions (we postpone a careful discussion of Lyapunov theory to Chapter 14). The same quadratic function is useful also for sequences of primitive row-stochastic matrices $\{A(k)\}_{k\in\mathbb{Z}_{\geq 0}}$ with a common dominant left eigenvector, see Exercise E11.5. More general cases require a different type (not quadratic) of "decreasing energy" fuctions.

11.4 Convergence over time-varying digraphs connected over time

We are now ready to state the main result in this chapter.

Theorem 11.4 (Consensus for time-varying algorithms (Moreau, 2005)). *Let* $\{A(k)\}_{k\in\mathbb{Z}_{\geq 0}}$ *be a sequence of row-stochastic matrices with associated digraphs* $\{G(k)\}_{k\in\mathbb{Z}_{\geq 0}}$*. Assume that*

(A1) *each digraph* $G(k)$ *has a self-loop at each node;*

(A2) *each non-zero edge weight* $a_{ij}(k)$*, including the self-loops weights* $a_{ii}(k)$*, is larger than a constant* $\varepsilon > 0$*; and*

(A3) *there exists a duration* $\delta \in \mathbb{N}$ *such that, for all times* $k \in \mathbb{Z}_{\geq 0}$*, the digraph* $G(k) \cup \cdots \cup G(k+\delta-1)$ *contains a globally reachable node.*

Then

(i) *there exists a non-negative* $w \in \mathbb{R}^n$ *normalized to* $w_1 + \cdots + w_n = 1$ *such that* $\lim_{k\to\infty} A(k) \cdot A(k-1) \cdot \cdots \cdot A(0) = \mathbb{1}_n w^\mathsf{T}$*;*

(ii) *the solution to* $x(k+1) = A(k)x(k)$ *converges exponentially fast to* $\left(w^\mathsf{T} x(0)\right)\mathbb{1}_n$*;*

(iii) *if additionally each matrix in the sequence is doubly-stochastic, then* $w = \frac{1}{n}\mathbb{1}_n$ *so that*

$$\lim_{k\to\infty} x(k) = \operatorname{average}\!\left(x(0)\right)\mathbb{1}_n.$$

Note: In a sequence with property (A2), edges can appear and disappear, but the weight of each edge (that appears an infinite number of times) does not go to zero as $k \to \infty$.

Note: This result is analogous to the time-invariant result that we saw in Chapter 5. The existence of a globally reachable node is the connectivity requirement in both cases.

Note: Assumption (A3) is a *uniform* connectivity requirement, that is, any interval of length δ must have the connectivity property. In equivalent words, the connectivity property holds for any contiguous interval of duration δ.

Example 11.5 (Shared communication channel with round robin scheduling). Consider the shared communication channel model with round-robin scheduling. Assume the algorithm is implemented over a communication graph $G_{\text{shared-comm}}$ that is strongly connected.

Consider now the assumptions in Theorem 11.4. Assumption (A1) is satisfied because in equation (11.1) the self-loop weight is equal to $1/2$. Similarly, Assumption (A2) is satisfied because the edge weight is equal to $1/2$. Finally, Assumption (A3) is satisfied with duration δ selected equal to n, because after n rounds each node has transmitted precisely once and so all edges of the communication graph $G_{\text{shared-comm}}$ are

present in the union graph. Therefore, the algorithm converges to consensus. However, the algorithm does not converge to average consensus since it is false that the averaging matrices are doubly-stochastic.

Note: round robin is not necessarily the only scheduling protocol with convergence guarantees. Indeed, consensus is achieved so long as each node is guaranteed a transmission slot once every bounded period of time.   •

Next, we provide a second theorem on convergence over time-varying averaging systems, whereby we assume the matrix to be symmetric and the corresponding graphs to be connected over time.

Theorem 11.6 (Consensus for symmetric time-varying algorithms). *Let $\{A(k)\}_{k \in \mathbb{Z}_{\geq 0}}$ be a sequence of symmetric row-stochastic matrices with associated undirected graphs $\{G(k)\}_{k \in \mathbb{Z}_{\geq 0}}$. Let the matrix sequence $\{A(k)\}_{k \in \mathbb{Z}_{\geq 0}}$ satisfy Assumptions (A1) and (A2) in Theorem 11.4 as well as*

(A4) *for all $k \in \mathbb{Z}_{\geq 0}$, the graph $\cup_{\tau \geq k} G(\tau)$ is connected.*

Then

(i) *$\lim_{k \to \infty} A(k) \cdot A(k-1) \cdot \cdots \cdot A(0) = \frac{1}{n} \mathbb{1}_n \mathbb{1}_n^{\mathsf{T}}$;*

(ii) *each solution to $x(k+1) = A(k)x(k)$ converges exponentially fast to $\mathrm{average}\big(x(0)\big)\mathbb{1}_n$.*

Note: this result is analogous to the time-invariant result that we saw in Chapter 5. For symmetric row-stochastic matrices and undirected graphs, the connectivity of an appropriate graph is the requirement in both cases.

Note: Assumption (A3) in Theorem 11.4 requires the existence of a finite time-interval of duration δ so that the union graph $\cup_{k \leq \tau \leq k+\delta-1} G(\tau)$ contains a globally reachable node for all times $k \geq 0$. This assumption is weakened in the symmetric case in Theorem 11.6 to Assumption (A4) requiring that the union graph $\cup_{\tau \geq k} G(\tau)$ is connected for all times $k \geq 0$.

Finally, we conclude this section with an instructive example.

Example 11.7 (Uniform connectivity is required for non-symmetric matrices). We have learned that, for asymmetric matrices, a uniform connectivity property (A3) is required, whereas for symmetric matrices, uniform connectivity is not required (see (A4)). Here is a counter-example from (Hendrickx, 2008) showing that Assumption (A3) cannot be relaxed for asymmetric graphs. Initialize a group of $n = 3$ agents to

$$x_1 < -1, \quad x_2 < -1, \quad x_3 > +1.$$

Step 1: Perform $x_1^+ := (x_1 + x_3)/2$, $x_2^+ := x_2$, $x_3^+ := x_3$ a number of times δ_1 until

$$x_1 > +1, \quad x_2 < -1, \quad x_3 > +1.$$

Step 2: Perform $x_1^+ := x_1$, $x_2^+ := x_2$, $x_3^+ := (x_2 + x_3)/2$ a number of times δ_2 until

$$x_1 > +1, \quad x_2 < -1, \quad x_3 < -1.$$

Step 3: Perform $x_1^+ := x_1$, $x_2^+ := (x_1 + x_2)/2$, $x_3^+ := x_3$ a number of times δ_3 until

$$x_1 > +1, \quad x_2 > +1, \quad x_3 < -1.$$

And repeat this process.
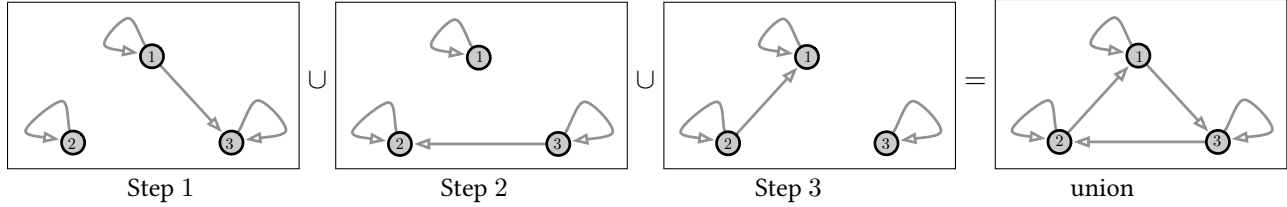
Observe that on steps $1, 7, 15, \ldots$, the variable x_1 is made to become larger than $+1$ by computing averages with $x_3 > +1$. Note that every time this happens the variable $x_3 > +1$ is increasingly smaller and closer to $+1$. Hence, $\delta_1 < \delta_7 < \delta_{15} < \ldots$, that is, it takes more steps for x_1 to become larger than $+1$. Indeed, one can formally show the following:

(i) The agents do not converge to consensus.

(ii) Hence, one of the assumptions of Theorem 11.4 must be violated.

(iii) It is easy to see that (A1) and (A2) are satisfied.

(iv) Regarding connectivity, note that, for all $k \in \mathbb{Z}_{\geq 0}$, the digraph $\cup_{\tau \geq k} G(\tau)$ contains a globally reachable node. However, this property is not quite equivalent to Assumption (A3).

(v) Assumption (A3) in Theorem 11.4 must be violated: there does not exist a duration $\delta \in \mathbb{N}$ such that, for all $k \in \mathbb{Z}_{\geq 0}$, the digraph $G(k) \cup \cdots \cup G(k + \delta - 1)$ contains a globally reachable node.

(vi) Indeed, one can show that $\lim_{k \to \infty} \delta_k = \infty$ so that, as we keep iterating Steps 1+2+3, their duration grows unbounded.   •

11.5   A new analysis method for convergence to consensus

It is well known that, for time-varying systems, the analysis of eigenvalues is not appropriate anymore. In the following example, two matrices with spectral radius equal to $1/2$ are multiplied to obtain a spectral radius larger than 1:

$$\begin{bmatrix} \frac{1}{2} & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{2} & 0 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} \frac{5}{4} & 0 \\ 0 & 0 \end{bmatrix}.$$

This example explains how it is not possible to predict the convergence of arbitrary products of matrices, just based on their spectral radii. In other words, we need to work harder and with sharper tools.

11.5.1   The max-min function and row-stochastic matrices

In what follows we present a new analysis method for the convergence to consensus for the discrete-time averaging system. Before establishing the results for time-varying averaging systems, it is instructive to rederive the convergence results for time-invariant averaging systems.

We start our analysis by defining the *max-min function* $V_{\text{max-min}} : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ by

$$V_{\text{max-min}}(x) = \max(x_1, \ldots, x_n) - \min(x_1, \ldots, x_n)$$
$$= \max_{i \in \{1,\ldots,n\}} x_i - \min_{i \in \{1,\ldots,n\}} x_i.$$

Note that:

   (i) $V_{\text{max-min}}(x) \geq 0$, and

  (ii) $V_{\text{max-min}}(x) = 0$ if an only if $x = \alpha \mathbb{1}_n$ for some $\alpha \in \mathbb{R}$.

The following result is a generalization of Exercise E1.1.

Lemma 11.8 (Monotonicity and bounded evolutions). *If A is row-stochastic, then for all $x \in \mathbb{R}^n$*

$$V_{\text{max-min}}(Ax) \leq V_{\text{max-min}}(x).$$

For any sequence of row-stochastic matrices, the solution $x(k)$ of the corresponding time-varying averaging algorithm satisfies, from any initial condition $x(0)$ and at any time k,

$$V_{\text{max-min}}(x(k)) \leq V_{\text{max-min}}(x(0)), \quad \text{and}$$
$$\min x(0) \leq \min x(k) \leq \min x(k+1) \leq \max x(k+1) \leq \max x(k) \leq \max x(0).$$

Proof. For the maximum, let us compute:

$$\max_i (Ax)_i = \max_i \sum_{j=1}^n a_{ij} x_j \leq \max_i \sum_{j=1}^n a_{ij} \left(\max_h x_h \right) = \left(\max_i \sum_{j=1}^n a_{ij} \right) \left(\max_h x_h \right) = 1 \cdot \max_i x_i.$$

Similarly, for the minimum,

$$\min_i (Ax)_i = \min_i \sum_{j=1}^n a_{ij} x_j \geq \min_i \sum_{j=1}^n a_{ij} \left(\min_h x_h \right) = \left(\min_i \sum_{j=1}^n a_{ij} \right) \left(\min_h x_h \right) = 1 \cdot \min_i x_i.$$

                                                                                                                ■

Next, given an n-dimensional row-stochastic matrix A, we define its *column-maximum row-minimum entry*, denoted $\gamma(A)$, by

$$\gamma(A) = \max_{j \in \{1,\ldots,n\}} \min_{i \in \{1,\ldots,n\}} a_{ij} \in [0, 1]. \tag{11.3}$$

It is useful to clarify this definition and explain how to compute this quantity: for each column j the quantity $b_j = \min_i a_{ij}$ is the smallest entry over the n rows, and then $\gamma(A) = \max_j b_j$ is the largest of these entries over the n columns.

    The next lemma provides an alternative proof method for the convergence to consensus of row stochastic matrices.

Lemma 11.9 (Alternative convergence analysis for discrete-time averaging). *Given an n-dimensional row-stochastic matrix A, the following statements hold:*

  (i) *for all $x \in \mathbb{R}^n$, the max-min function satisfies*

$$V_{\text{max-min}}(Ax) \leq \big(1 - \gamma(A)\big)V_{\text{max-min}}(x);$$

  (ii) $\gamma(A) > 0$ *if and only if A has a strictly positive column;*

  (iii) *the following properties of A are equivalent:*

   a) *the digraph associated to A contains a globally reachable node and the subgraph of globally reachable nodes is aperiodic,*

   b) *there exists an index $h \in \mathbb{N}$ such that A^h has a positive column, and*

   c) *A is semiconvergent to a rank-one matrix.*

The lemma immediately implies the following statement: if A satisfies any (and therefore all) of the properties in statement (iii), then $k \mapsto V_{\text{max-min}}(x(k))$ converges exponentially fast to zero in the sense that, for all time $k \in \mathbb{N}$ and for an index h as in statement (iii)b,

$$V_{\text{max-min}}(x(k)) \leq \underbrace{\big(1 - \gamma(A^h)\big)}_{<1}^{\lfloor k/h \rfloor} V_{\text{max-min}}(x(0)).$$

Proof of Lemma 11.9. Statement (i) is trivial if $\max_{j \in \{1,\dots,n\}} \min_{i \in \{1,\dots,n\}} a_{ij} = 0$; e.g., see Exercise E1.1. Hence, let us prove the statement when $\max_{j \in \{1,\dots,n\}} \min_{i \in \{1,\dots,n\}} a_{ij} = a_{\bar{i}\bar{j}} > 0$. We compute

$$
\begin{aligned}
V_{\text{max-min}}(Ax) &= \max_i \sum_{p=1}^n a_{ip}x_p - \min_i \sum_{p=1}^n a_{ip}x_p \\
&= \max_i \Big(\sum_{p=1,p\neq j}^n a_{ip}x_p + a_{ij}x_j \Big) - \min_i \Big(\sum_{p=1,p\neq j}^n a_{ip}x_p + a_{ij}x_j \Big) \\
&\leq \min_j \Big[\max_i \Big(\sum_{p=1,p\neq j}^n a_{ip}x_{\max} + a_{ij}x_j \Big) - \min_i \Big(\sum_{p=1,p\neq j}^n a_{ip}x_{\min} + a_{ij}x_j \Big) \Big],
\end{aligned}
$$

where, after using the bounds $x_{\min} \leq x_p \leq x_{\max}$, we minimize the right hand side as a function of j. From the latter equation we obtain

$$V_{\text{max-min}}(Ax) = \min_j \max_i \Big((1-a_{ij})x_{\max} + a_{ij}x_j \Big) - \max_j \min_i \Big((1-a_{ij})x_{\min} + a_{ij}x_j \Big)$$

and, noting that $\min_j \max_i (1 - a_{ij}) = 1 - \max_j \min_i a_{ij} = 1 - a_{\bar{i}\bar{j}}$,

$$
\begin{aligned}
&= \Big((1-a_{\bar{i}\bar{j}})x_{\max} + a_{\bar{i}\bar{j}}x_{\bar{j}} \Big) - \Big((1-a_{\bar{i}\bar{j}})x_{\min} + a_{\bar{i}\bar{j}}x_{\bar{j}} \Big) \\
&= (1-a_{\bar{i}\bar{j}})\big(x_{\max} - x_{\min}\big) = (1-a_{\bar{i}\bar{j}})V_{\text{max-min}}(x).
\end{aligned}
$$

Statement (ii) is an immediate consequence of the definition of $\gamma(A)$. Indeed, if each column j has an entry equal to zero, then the quantity $b_j = \min_i a_{ij} = 0$ for all j and, and in turn, $\gamma(A) = \max_j b_j = 0$.

Regarding statement (iii), the equivalence between (iii)a and (iii)b is a generalization of Theorem 4.7 (given as Exercise E4.12) and the equivalence between (iii)a and (iii)c is given in Theorem 5.1. ∎

11.5.2 Connectivity over time

Before presenting the convergence to consensus proof for time-varying averaging systems, we provide one more useful result. This result allows us to manipulate our assumption of connectivity over time.

Lemma 11.10 (Global reachability over time). *Given a sequence of digraphs $\{G(k)\}_{k \in \mathbb{Z}_{\geq 0}}$ such that each digraph $G(k)$ has a self-loop at each node, the following two properties are equivalent:*

(i) *there exists a duration $\delta \in \mathbb{N}$ such that, for all times $k \in \mathbb{Z}_{\geq 0}$, the digraph $G(k) \cup \cdots \cup G(k + \delta - 1)$ contains a directed spanning tree;*

(ii) *there exists a duration $\Delta \in \mathbb{N}$ such that, for all times $k \in \mathbb{Z}_{\geq 0}$, there exists a node $j = j(k)$ that reaches all nodes $i \in \{1, \ldots, n\}$ over the interval $\{k, k + \Delta - 1\}$ in the following sense: there exists a sequence of nodes $\{j, h_1, \ldots, h_{\Delta-1}, i\}$ such that (j, h_1) is an edge at time k, (h_1, h_2) is an edge at time $k + 1$, ..., $(h_{\Delta-2}, h_{\Delta-1})$ is an edge at time $k + \Delta - 2$, and $(h_{\Delta-1}, i)$ is an edge at time $k + \Delta - 1$;*

or, equivalently, for the reverse digraph,

(iii) *there exists a duration $\delta \in \mathbb{N}$ such that, for all times $k \in \mathbb{Z}_{\geq 0}$, the digraph $G(k) \cup \cdots \cup G(k + \delta - 1)$ contains a globally reachable node;*

(iv) *there exists a duration $\Delta \in \mathbb{N}$ such that, for all times $k \in \mathbb{Z}_{\geq 0}$, there exists a node j reachable from all nodes $i \in \{1, \ldots, n\}$ over the interval $\{k, k + \Delta - 1\}$ in the following sense: there exists a sequence of nodes $\{j, h_1, \ldots, h_{\Delta-1}, i\}$ such that (h_1, j) is an edge at time k, (h_2, h_1) is an edge at time $k + 1$, ..., $(h_{\Delta-1}, h_{\Delta-2})$ is an edge at time $k + \Delta - 2$, and $(i, h_{\Delta-1})$ is an edge at time $k + \Delta - 1$.*

Note: It is sometimes easy to see if a sequence of digraphs satisfies properties (i) and (iii). Property (iv) is directly useful in the analysis later in the chapter. Regarding the proof of the lemma, it is easy to check that (ii) implies (i) and that (iv) implies (iii) with $\delta = \Delta$. The converse is left as Exercise E11.3.

11.5.3 Proof of Theorem 11.4: the max-min function is exponentially decreasing

We are finally ready to prove Theorem 11.4. We start by noting that Assumptions (A1) and (A3) imply property Lemma 11.10(iv) about the existence of a duration Δ with certain properties. Next, without loss of generality, we assume that at some time $h\Delta$, for some $h \in \mathbb{N}$, the solution $x(h\Delta)$ is not equal to a multiple of $\mathbb{1}_n$ and, therefore, satisfies $V_{\text{max-min}}(x(h\Delta)) > 0$. Clearly,

$$x((h+1)\Delta) = A((h+1)\Delta - 1) \cdots A(h\Delta + 1) \cdot A(h\Delta)\, x(h\Delta)$$
$$=: \mathcal{A}x(h\Delta).$$

By Assumption (A3), we know that there exists a node j reachable from all nodes i over the interval $\{h\Delta, (h+1)\Delta - 1\}$ in the following sense: there exists a sequence of nodes $\{j, h_1, \ldots, h_{\Delta-1}, i\}$ such that all following edges exist in the sequence of digraphs: (h_1, j) at time $h\Delta$, (h_2, h_1) at time $h\Delta + 1$, ..., $(i, h_{\Delta-1})$ at time $(h+1)\Delta - 1$. Therefore, Assumption (A2) implies

$$a_{h_1,j}\big(h\Delta\big) \geq \varepsilon, \ \ a_{h_2,h_1}\big(h\Delta + 1\big) \geq \varepsilon, \ \ \ldots, \ \ a_{i,h_{\Delta-1}}\big((h+1)\Delta - 1\big) \geq \varepsilon,$$

and therefore their product satisfies

$$a_{i,h_{\Delta-1}}\big((h+1)\Delta - 1\big) \cdot a_{h_{\Delta-1},h_{\Delta-2}}\big((h+1)\Delta - 2\big) \cdots a_{h_2,h_1}\big(h\Delta + 1\big) \cdot a_{h_1,j}\big(h\Delta\big) \geq \varepsilon^{\Delta}.$$

Remarkably, this product is one term in the (i,j) entry of the row-stochastic matrix $\mathcal{A} = A((h+1)\Delta - 1) \cdots A(h\Delta)$. In other words, Assumption (A3) implies $\mathcal{A}_{ij} \geq \varepsilon^{\Delta}$.

Hence, for all nodes i, given globally reachable node j during interval $\{h\Delta, (h+1)\Delta\}$, we compute

$$
\begin{aligned}
x_i\big((h+1)\Delta\big) &= \mathcal{A}_{i,j}x_j(h\Delta) + \sum\nolimits_{p\neq j,p=1}^{n} \mathcal{A}_{i,p}x_p(h\Delta) && \text{(by definition)} \\
&\leq \mathcal{A}_{i,j}x_j(h\Delta) + (1 - \mathcal{A}_{i,j})\max\big(x(h\Delta)\big) && \text{(because } x_p(h\Delta) \leq \max\big(x(h\Delta)\big)\text{)} \\
&\leq \min_{\mathcal{A}_{i,j}}\Big(\mathcal{A}_{i,j}x_j(h\Delta) + (1 - \mathcal{A}_{i,j})\max\big(x(h\Delta)\big)\Big) && \text{(because } x_j(h\Delta) \leq \max\big(x(h\Delta)\big)\text{)} \\
&\leq \varepsilon^{\Delta}x_j(h\Delta) + (1 - \varepsilon^{\Delta})\max\big(x(h\Delta)\big).
\end{aligned}
$$

A similar argument leads to

$$x_i\big((h+1)\Delta\big) \geq \varepsilon^{\Delta}x_j(h\Delta) + (1 - \varepsilon^{\Delta})\min\big(x(h\Delta)\big),$$

so that

$$
\begin{aligned}
V_{\text{max-min}}\big(x((h+1)\Delta)\big) &= \max_i x_i\big((h+1)\Delta\big) - \min_i x_i\big((h+1)\Delta\big) \\
&\leq \Big(\varepsilon^{\Delta}x_j(h\Delta) + (1 - \varepsilon^{\Delta})\max\big(x(h\Delta)\big)\Big) - \Big(\varepsilon^{\Delta}x_j(h\Delta) + (1 - \varepsilon^{\Delta})\min\big(x(h\Delta)\big)\Big) \\
&\leq (1 - \varepsilon^{\Delta})V_{\text{max-min}}\big(x(h\Delta)\big).
\end{aligned}
$$

This final inequality, together with Lemma 11.8, proves exponential convergence of the cost function $k \mapsto V_{\text{max-min}}(x(k))$ to zero and convergence of $x(k)$ to a multiple of $\mathbb{1}_n$. We leave the other statements in Theorem 11.4 to the reader and refer to (Moreau, 2005; Hendrickx, 2008) for further details.

11.6   Time-varying algorithms in continuous-time

We now briefly consider the continuous-time linear time-varying system

$$\dot{x}(t) = -L(t)x(t).$$

We associate a time-varying graph $G(t)$ (without self loops) to the time-varying Laplacian $L(t)$ in the usual manner.

For example, in Chapter 7, we discussed how the heading in some flocking models is described by the continuous-time Laplacian flow:

$$\dot{\theta} = -L\theta,$$

where each θ is the heading of a bird, and where L is the Laplacian of an appropriate weighted digraph G: each bird is a node and each directed edge (i,j) has weight $1/d_{\text{out}}(i)$. We discussed also the need to consider time-varying graphs: birds average their heading only with other birds within sensing range, but this sensing relationship may change with time.

Recall that the solution to a continuous-time time-varying system can be given in terms of the state transition matrix:

$$x(t) = \Phi(t,0)x(0),$$

We refer to (Hespanha, 2009) for the proper definition and study of the state transition matrix.

Theorem 11.11 (Consensus for time-varying algorithms in continuous time). *Let* $t \mapsto A(t)$ *be a time-varying adjacency matrix with associated time-varying digraph* $t \mapsto G(t)$, $t \in \mathbb{R}_{\geq 0}$. *Assume*

(A1) *each non-zero edge weight* $a_{ij}(t)$ *is larger than a constant* $\varepsilon > 0$,

(A2) *there exists a duration* $T > 0$ *such that, for all* $t \in \mathbb{R}_{\geq 0}$, *the digraph associated to the adjancency matrix*

$$\int_t^{t+T} L(\tau)d\tau$$

contains a globally reachable node.

Then

(i) *there exists a non-negative* $w \in \mathbb{R}^n$ *normalized to* $w_1 + \cdots + w_n = 1$ *such that the state transition matrix* $\Phi(t,0)$ *associated to* $-L(t)$ *satisfies* $\lim_{t \to \infty} \Phi(t,0) = \mathbb{1}_n w^\mathsf{T}$,

(ii) *the solution to* $\dot{x}(t) = -L(t)x(t)$ *converges exponentially fast to* $\left(w^\mathsf{T} x(0)\right)\mathbb{1}_n$,

(iii) *if additionally, the* $\mathbb{1}_n^\mathsf{T} L(t) = \mathbb{0}_n^\mathsf{T}$ *for almost all times* t *(that is, the digraph is weight-balanced at all times, except a set of measure zero), then* $w = \frac{1}{n}\mathbb{1}_n$ *so that*

$$\lim_{t \to \infty} x(t) = \operatorname{average}\left(x(0)\right)\mathbb{1}_n.$$

11.7 Historical notes and further reading

The main theorem in this chapter is due to Moreau (2005). Note that Theorem 11.4 provides only sufficient condition for consensus in time-varying averaging systems. For results on necessary and sufficient conditions we refer the reader to the recent works (Blondel and Olshevsky, 2014; Xia and Cao, 2014) and references therein. The proof of Theorem 11.4 is inspired by the presentation in (Hendrickx, 2008, Theorem 9.2).

In the context of time-varying averaging systems, other relevant references on first and second order, discrete and continuous time systems include (Tsitsiklis, 1984; Tsitsiklis et al., 1986; Hong et al., 2006, 2007; Cao et al., 2008; Carli et al., 2008b).

For references on time-varying continuous-time averaging systems we refer to (Moreau, 2004; Lin et al., 2007; Hendrickx and Tsitsiklis, 2013).

11.8   Exercises

E11.1   **On the product of stochastic matrices ([Jadbabaie et al., 2003]).** Let $k \geq 2$ and $A_1, A_2, \ldots, A_k$ be non-negative $n \times n$ matrices with positive diagonal entries. Let $a_{\min}$ (resp. $a_{\max}$) be the smallest (resp. largest) diagonal entry of $A_1, A_2, \ldots, A_k$ and let $G_1, \ldots, G_k$ be the digraphs associated with $A_1, \ldots, A_k$.
Show that

(i) $A_1 A_2 \cdots A_k \geq \left(\dfrac{a_{\min}^2}{2a_{\max}} \right)^{k-1} (A_1 + A_2 + \cdots + A_k)$, and

(ii) if the digraph $G_1 \cup \ldots \cup G_k$ is strongly connected, then the matrix $A_1 \cdots A_k$ is irreducible.

Hint: Set $A_i = a_{\min} I_n + B_i$ for a non-negative B_i, and show statement (i) by induction on k.

E11.2   **Products of primitive matrices with positive diagonal.** Let $A_1, A_2, \ldots, A_{n-1}$ be primitive $n \times n$ matrices with positive diagonal entries. Show that $A_1 A_2 \cdots A_{n-1} > 0$.

E11.3   **A simple proof.** Prove Lemma 11.10.
Hint: You will want to use Exercise E3.6.

E11.4   **Alternative sufficient condition.** As in Theorem 11.4, let $\{A(k)\}_{k \in \mathbb{Z}_{\geq 0}}$ be a sequence of row-stochastic matrices with associated digraphs $\{G(k)\}_{k \in \mathbb{Z}_{\geq 0}}$. Prove that the same asymptotic properties in Theorem 11.4 hold true under the following Assumption (A5), instead of Assumptions (A1), (A2), and (A3):

(A5) there exists a node j such that, for all times $k \in \mathbb{Z}_{\geq 0}$, each edge weight $a_{ij}(k)$, $i \in \{1, \ldots, n\}$, is larger than a constant $\varepsilon > 0$.

In other words, Assumption (A5) requires that all digraphs $G(k)$ contain all edges $a_{ij}(k)$, $i \in \{1, \ldots, n\}$, and that all these edges have weights larger than a strictly positive constant.
Hint: Modify the proof of Theorem 11.4.

E11.5   **Convergence over digraphs strongly-connected at all times.** Consider a sequence $\{A(k)\}_{k \in \mathbb{Z}_{\geq 0}}$ of row-stochastic matrices with associated digraphs $\{G(k)\}_{k \in \mathbb{Z}_{\geq 0}}$ so that

(A1) each non-zero edge weight $a_{ij}(k)$, including the self-loops weights $a_{ii}(k)$, is larger than a constant $\varepsilon > 0$;

(A2) each digraph $G(k)$ is strongly connected and aperiodic point-wise in time; and

(A3) there is a positive vector $w \in \mathbb{R}^n$ satisfying $\mathbb{1}_n^\mathsf{T} w = 1$ and $w^\mathsf{T} A(k) = w^\mathsf{T}$ for all $k \in \mathbb{Z}_{\geq 0}$.

Without relying on Theorem 11.4, show that the solution to $x(k+1) = A(k)x(k)$ satisfies to $\lim_{k \to \infty} x(k) = (w^\mathsf{T} x(0)) \mathbb{1}_n$.
Hint: Search for a non-negative cost function decreasing along the dynamics, as in the proof of Theorem 11.3.

Randomized Averaging Algorithms

In this chapter we discuss averaging algorithms defined by sequences of random stochastic matrices. In other words, we imagine that at each discrete instant, the averaging matrix is selected randomly according to some stochastic model. We refer to such algorithms as *randomized averaging algorithms*. Randomized averaging algorithms are well behaved and easy to study in the sense that much information can be learned simply from the expectation of the averaging matrix.

12.1 Examples of randomized averaging algorithms

Consider the following models of randomized averaging algorithms.

Uniform Symmetric Gossip. Given an undirected graph G, at each iteration, select uniformly likely one of the graph edges, say agents i and j talk, and they both perform $(1/2, 1/2)$ averaging, that is:
$$x_i(k+1) = x_j(k+1) := \frac{1}{2}\big(x_i(k) + x_j(k)\big).$$

Packet Loss in Communication Network. Given a strongly connected and aperiodic digraph, at each communication round, packets travel over directed edges and, with some likelihood, each edge may drop the packet. (If information is not received, then the receiving node can either do no update whatsoever, or adjust its averaging weights to compensate for the packet loss).

Broadcast Wireless Communication. Given a digraph, at each communication round, a randomly-selected node transmits to all its out-neighbors. (Here we imagine that simultaneous transmissions are prohibited by wireless interference.)

Opinion Dynamics with Stochastic Interactions and Prominent Agents. (Somehow similar to uniform gossip) Given an undirected graph and a probability $0 < p < 1$, at each iteration, select uniformly likely one of the graph edges and perform: with probability p both agents perform the $(1/2, 1/2)$ update, and with probability $(1-p)$ only one agent performs the update and the "prominent agent" does not.

Note that, in the second, third and fourth example models, the row-stochastic matrices at each iteration are not symmetric in general, even if the original digraph was undirected.

12.2   A brief review of probability theory

We briefly review a few basic concepts from probability theory and refer the reader for example to (Breiman, 1992).

- Loosely speaking, a *random variable* $X : \Omega \to E$ is a measurable function from the *set of possible outcomes* Ω to some set E which is typically a subset of $\mathbb{R}$.

- The *probability* of an event (i.e., a subset of possible outcomes) is the measure of the likelihood that the event will occur. An event occurs *almost surely* if it occurs with probability equal to 1.

- The random variable X is called *discrete* if its image is finite or countably infinite. In this case, X is described by a *probability mass function* assigning a probability to each value in the image of X.

  Specifically, if X takes value in $\{x_1, \dots, x_M\} \subset \mathbb{R}$, then the probability mass function $p : \{x_1, \dots, x_M\} \to [0, 1]$ satisfies $p_X(x_i) \geq 0$ and $\sum_{i=1}^n p_X(x_i) = 1$, and determines the probability of X being equal to x_i by $\mathbb{P}[X = x_i] = p_X(x_i)$.

- The random variable X is called *continuous* if its image is uncountably infinite. If X is an absolutely continuous function, X is described by a *probability density function* assigning a probability to intervals in the image of X.

  Specifically, if X takes value in $\mathbb{R}$, then the probability density function $f_X : \mathbb{R} \to [0, 1]$ satisfies $f(x) \geq 0$ and $\int_{\mathbb{R}} f(x)dx = 1$, and determines the probability of X taking value in the interval $[a, b]$ by $\mathbb{P}[a \leq X \leq b] = \int_a^b f(x)dx$.

- The *expected value* of a discrete variable is $\mathbb{E}[X] = \sum_{i=1}^M x_i p_X(x_i)$.

  The *expected value* of a continuous variable is $\mathbb{E}[X] = \int_{-\infty}^\infty x f_X(x)dx$.

- A (finite or infinite) sequence of random variables is *independent and identically distributed (i.i.d.)* if each random variable has the same probability mass/distribution as the others and all are mutually independent.

12.3   Randomized averaging algorithms

In this section we consider random sequences of row stochastic sequences. Accordingly, let $A(k)$ be the row-stochastic averaging matrix occurring randomly at time k and $G(k)$ be its associated graph. We then consider the stochastic linear system

$$x(k + 1) = A(k)x(k).$$

We are now ready to present the main result of this chapter.

Theorem 12.1 (Consensus for randomized algorithms). *Let $\{A(k)\}_{k \in \mathbb{Z}_{\geq 0}}$ be a sequence of random row-stochastic matrices with associated digraphs $\{G(k)\}_{k \in \mathbb{Z}_{\geq 0}}$. Assume*

(A1) *the sequence of variables $\{A(k)\}_{k \in \mathbb{Z}_{\geq 0}}$ is i.i.d.,*

(A2) *at each time k, the random matrix $A(k)$ has a strictly positive diagonal so that each digraph in the sequence $\{G(k)\}_{k \in \mathbb{Z}_{\geq 0}}$ has a self-loop at each node, and*

(A3) *the digraph associated to the expected matrix $\mathbb{E}[A(k)]$, for any k, has a globally reachable node.*

Then the following statements hold almost surely:

(i) *there exists a random non-negative vector $w \in \mathbb{R}^n$ with $w_1 + \cdots + w_n = 1$ such that*

$$\lim_{k \to \infty} A(k) \cdot A(k-1) \cdot \cdots \cdot A(0) = \mathbb{1}_n w^{\mathsf{T}} \quad \text{almost surely,}$$

(ii) *as $k \to \infty$, each solution $x(k)$ of $x(k+1) = A(k)x(k)$ satisfies*

$$\lim_{k \to \infty} x(k) = \left(w^{\mathsf{T}} x(0)\right) \mathbb{1}_n \quad \text{almost surely,}$$

(iii) *if additionally each random matrix is doubly-stochastic, then $w = \frac{1}{n}\mathbb{1}_n$ so that*

$$\lim_{k \to \infty} x(k) = \operatorname{average}\left(x(0)\right) \mathbb{1}_n.$$

Note: if each random matrix is doubly-stochastic, then $\mathbb{E}[A(k)]$ is doubly-stochastic. The converse is easily seen to be false.

Note: Assumption (A1) is restrictive and more general conditions are sufficient; see the discussion below in Section 12.4.

12.3.1 Additional results on uniform symmetric gossip algorithms

Recall: given undirected graph G, at each iteration, select uniformly likely one of the graph edges, say agents i and j talk, and they both perform $(1/2, 1/2)$ averaging, that is:

$$x_i(k+1) = x_j(k+1) := \frac{1}{2}\left(x_i(k) + x_j(k)\right).$$

Corollary 12.2 (Convergence for uniform symmetric gossip). *If the graph G is connected, then each solution to the uniform symmetric gossip converges to average consensus with probability* 1.

Proof based on Theorem 12.1. The corollary can be established by verifying that Assumptions (A1)–(A3) in Theorem 12.1 are satisfied. Regarding (A3), note that the graph associated to the expected averaging matrix is G. ∎

We provide also an alternative elegant proof.

Proof based on Theorem 11.6. For any time $k_0 \geq 0$ and any edge (i, j), consider the event "the edge (i, j) is not selected for update at any time larger than k_0." Since the probability that (i, j) is not selected at any time k is $1 - 1/m$, whre m is the number of edges, the probability that (i, j) is not selected at any times after k_0 is

$$\lim_{k \to \infty} \left(1 - \frac{1}{m}\right)^{k - k_0} = 0.$$

With this fact one can verify that all assumptions in Theorem 11.6 are satisfied by the random sequence of matrices almost surely. Hence, almost sure convergence follows. Finally, since each matrix is doubly stochastic, $\operatorname{average}(x(k))$ is preserved, and the solution converges to $\operatorname{average}(x(0))\mathbb{1}_n$. ∎

12.3.2 Additional results on the mean-square convergence factor

Given a sequence of stochastic averaging matrices $\{A(k)\}_{k\in\mathbb{Z}_{\geq 0}}$ and corresponding solutions $x(k)$ to $x(k+1) = A(k)x(k)$, we define the *mean-square convergence factor* by

$$r_{\text{mean-square}}\big(\{A(k)\}_{k\in\mathbb{Z}_{\geq 0}}\big) = \sup_{x(0)\neq x_{\text{final}}} \limsup_{k\to\infty} \left(\mathbb{E}\left[\|x(k) - \text{average}(x(k))\mathbb{1}_n\|_2^2\right] \right)^{1/k}.$$

We now present upper and lower bounds for the mean-square convergence factor.

Theorem 12.3 (Upper and lower bounds on the mean-square convergence factor). *Under the same assumptions as in Theorem 12.1, the mean-square convergence factor satisfies*

$$\rho_{\text{ess}}\big(\mathbb{E}[A(k)]\big)^2 \leq r_{\text{mean-square}} \leq \rho\Big(\mathbb{E}\left[A(k)^\mathsf{T}(I_n - \mathbb{1}_n\mathbb{1}_n^\mathsf{T}/n)A(k)\right]\Big).$$

12.4 Historical notes and further reading

In this chapter we present results from (Fagnani and Zampieri, 2008; Tahbaz-Salehi and Jadbabaie, 2008; Garin and Schenato, 2010) that build on classic references such as (Chatterjee and Seneta, 1977; Cogburn, 1984). Specifically, references for the main Theorem 12.1 are (Tahbaz-Salehi and Jadbabaie, 2008) and (Fagnani and Zampieri, 2008). Note that Assumption (A1) is restrictive and more general conditions are sufficient. For example, Tahbaz-Salehi and Jadbabaie (2010) treat the case of a sequence of row-stochastic matrices generated by an ergodic and stationary random process. Related analysis and modeling results are presented in (Hatano and Mesbahi, 2005; Bajović et al., 2013; Matei et al., 2013; Touri and Nedić, 2014; Ravazzi et al., 2015).

For a comprehensive analysis of the mean-square convergence factor we refer to (Fagnani and Zampieri, 2008, Proposition 4.4).

A detailed analysis of the uniform symmetric gossip model is given by Boyd et al. (2006). A detailed analysis of the model with stochastic interactions and prominent agents is given by (Acemoglu and Ozdaglar, 2011); see also (Acemoglu et al., 2013).

In this book we will not discuss averaging algorithms in the presence of quantization effects, we refer the reader instead to (Kashyap et al., 2007; Nedić et al., 2009; Frasca et al., 2009). Similarly, regarding averaging in the presence of noise, we refer to (Xiao et al., 2007; Bamieh et al., 2012; Lovisari et al., 2013; Jadbabaie and Olshevsky, 2015). Finally, regarding averaging in the presence of delays, we refer to (Olfati-Saber and Murray, 2004; Hu and Hong, 2007; Lin and Jia, 2008).

12.5 Table of asymptotic behaviors for averaging systems

Dynamics	Assumptions & Asymptotic Behavior	References
discrete-time: $x(k+1) = Ax(k)$, A row-stochastic adjacency matrix of digraph G	G has a globally reachable node $\implies$ $\lim_{k\to\infty} x(k) = (w^\mathsf{T}x(0))\mathbb{1}_n$, where $w \geq 0$, $w^\mathsf{T}A = w^\mathsf{T}$, and $\mathbb{1}_n^\mathsf{T}w = 1$	Thm 5.1
continuous-time: $\dot{x}(t) = -Lx(t)$, L Laplacian matrix of digraph G	G has a globally reachable node $\implies$ $\lim_{t\to\infty} x(t) = (w^\mathsf{T}x(0))\mathbb{1}_n$, where $w \geq 0$, $w^\mathsf{T}L = \mathbb{0}_n^\mathsf{T}$, and $\mathbb{1}_n^\mathsf{T}w = 1$	Thm 7.4
time-varying discrete-time: $x(k+1) = A(k)x(k)$, $A(k)$ row-stochastic adjacency matrix of digraph $G(k)$, $k \in \mathbb{Z}_{\geq 0}$	(i) at each time k, $G(k)$ has self-loop at each node, (ii) each $a_{ij}(k) \geq 0$ is larger than $\varepsilon > 0$, (iii) there exists duration δ s.t., for all time k, $G(k) \cup \cdots \cup G(k+\delta-1)$ has a globally reachable node $\implies$ $\lim_{k\to\infty} x(k) = (w^\mathsf{T}x(0))\mathbb{1}_n$, where $w \geq 0$, $\mathbb{1}_n^\mathsf{T}w = 1$	Thm 11.4
time-varying symmetric discrete-time: $x(k+1) = A(k)x(k)$, $A(k)$ symmetric stochastic adjacency of $G(k)$, $k \in \mathbb{Z}_{\geq 0}$	(i) at each time k, $G(k)$ has self-loop at each node, (ii) each $a_{ij}(k) \geq 0$ is larger than $\varepsilon > 0$, (iii) for all time k, $\cup_{\tau \geq k} G(\tau)$ is connected $\implies$ $\lim_{k\to\infty} x(k) = \operatorname{average}\big(x(0)\big)\mathbb{1}_n$	Thm 11.6
time-varying continuous-time: $\dot{x}(t) = -L(t)x(t)$, $L(t)$ Laplacian matrix of digraph $G(t)$, $t \in \mathbb{R}_{\geq 0}$	(i) each $a_{ij}(k) \geq 0$ is larger than $\varepsilon > 0$, (ii) there exists duration T s.t., for all time t, digraph associated to $\int_t^{t+T}L(\tau)d\tau$ has a globally reachable node $\implies$ $\lim_{k\to\infty} x(k) = (w^\mathsf{T}x(0))\mathbb{1}_n$, where $w \geq 0$, $\mathbb{1}_n^\mathsf{T}w = 1$	Thm 11.11
randomized discrete-time: $x(k+1) = A(k)x(k)$, $A(k)$ random row-stochastic adjacency matrix of digraph $G(k)$, $k \in \mathbb{Z}_{\geq 0}$	(i) $\{A(k)\}_{k \in \mathbb{Z}_{\geq 0}}$ is i.i.d., (ii) each matrix has strictly positive diagonal, (iii) digraph associated to $\mathbb{E}[A(k)]$ has a globally reachable node, $\implies$ $\lim_{k\to\infty} x(k) = \big(w^\mathsf{T}x(0)\big)\mathbb{1}_n$ almost surely, where $w > 0$ is random vector with $\mathbb{1}_n^\mathsf{T}w = 1$	Thm 12.1

Table 12.1: Averaging systems: definitions, assumptions, asymptotic behavior, and reference

Part III

Nonlinear Systems

Motivating Problems and Systems

In this chapter we begin our study of nonlinear network systems by introducing some example models and problems. Although the models presented are simple and their mathematical analyses are elementary, these models provide the appropriate notation, concepts, and intuition required to consider more realistic and complex models.

13.1 Lotka-Volterra population models

The Lotka-Volterra population models are one the simplest and most widely adopted frameworks for modeling the dynamics of interacting populations in mathematical ecology. These equations were originally developed in (Lotka, 1920; Volterra, 1928). In what follows we introduce various single-species and multi-species model of population dynamics. We start with single-species models. We let $x(t)$ denote the population number or its density at time t. The ratio $\dot{x}/x$ is the average contribution of an individual to the growth of the population.

Single-species constant growth model   In a simplest model, one may assume $\dot{x}/x$ is equal to a constant *growth rate* r. This assumption however leads to exponential growth or decay $x(t) = x(0)\,\mathrm{e}^{rt}$ depending upon whether r is positive or negative. Of course, exponential growth may be reasonable only for short periods of time and violates a reasonable assumption of *bounded resources* for large times.

Single-species logistic growth model   In large populations it is natural to assume that resources would diminish with the growing size of the population. In a very simple model, one may assume $\dot{x}/x = r(1-x/\kappa)$, where $r > 0$ is the intrinsic growth rate and $\kappa > 0$ is called the *carrying capacity*. This assumption leads to the so-called *logistic equation*

$$\dot{x}(t) = rx(t)\big(1 - x(t)/\kappa\big). \tag{13.1}$$

This dynamical system has the following behavior:

(i) there are two equilibrium points 0 and κ,

(ii) the solution is

$$x(t) = \frac{\kappa x(0) \, \mathrm{e}^{rt}}{\kappa + x(0)(\mathrm{e}^{rt} - 1)},$$

(iii) all solutions with $0 < x(0) < \kappa$ are monotonically increasing and converge asymptotically to κ,

(iv) all solutions with $\kappa < x(0)$ are monotonically decreasing and converge asymptotically to κ.

The reader is invited to show these facts and related ones in Exercise E13.1. The evolution of the logistic equation from multiple initial values is illustrated in Figure 13.1.
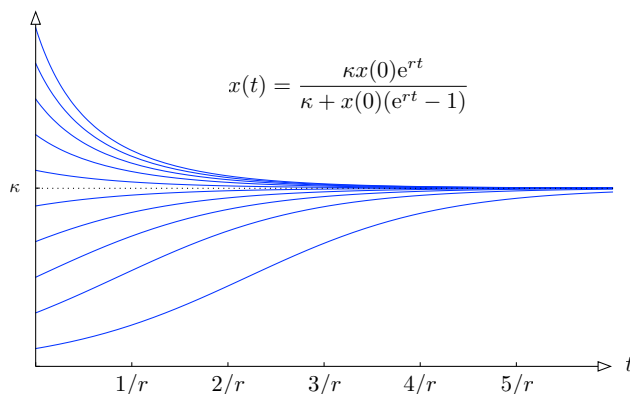
Figure 13.1: Solutions to the logistic equations from 10 initial conditions

Multi-species Lotka-Volterra model with signed interactions   Finally, we consider the case of $n \geq 2$ interacting species. We assume logistic growth model for each species with an additional term due to the interaction with the other species. Specifically, we write the growth rate for species $i \in \{1, \dots, n\}$,

$$\frac{\dot{x}_i}{x_i} = r_i + a_{ii} x_i + \sum_{j=1, j \neq i}^{n} a_{ij} x_j, \tag{13.2}$$

where the first two terms are the logistic equation (so that a_{ii} is typically negative because of bounded resources and the carrying capacity is $\kappa_i = -r_i/a_{ii}$), and the third term is the combined effect of the pairwise interactions with all other species. The vector r is called the *intrisic growth rate*, the matrix $A = [a_{ij}]$ is called the *interaction matrix*, and the ordinary differential equations (13.2) are called the *Lotka-Volterra model* for $n \geq 2$ interacting species. For $x \in \mathbb{R}^n_{\geq 0}$, this model is written in vector form as

$$\dot{x} = \mathrm{diag}(x)\big(Ax + r\big) =: f_{\mathrm{LV}}(x). \tag{13.3}$$

As illustrated in Figure 13.2, for any two species i and j, the sign of a_{ij} and a_{ji} in the interaction matrix A is determined by which of the following three possible types of interaction is being modeled:

(+, +) = mutualism: for $a_{ij} > 0$ and $a_{ji} > 0$, the two species are in symbiosis and cooperation. The presence of species i has a positive effect on the growth of species j and vice versa.
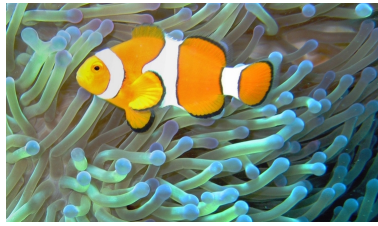
(a) Common clownfish (*Amphiprion ocellaris*) near magnificent sea anemones (*Heteractis magnifica*) on the Great Barrier Reef, Australia. Clownfish and anemones provide an example of ecological mutualism in that each species benefits from the activity of the other. Public domain image from Wikipedia.

(b) The canadian lynx (*Lynx canadensis*) is a major predator of the snowshoe hare (*Lepus americanus*). Historical records of animals captures indicate that the lynx and hare numbers rise and fall periodically; see (Odum, 1959). Public domain image from Rudolfo's Usenet Animal Pictures Gallery (no longer in existence).

(c) Subadult male lion (*Panthera Leo*) and spotted hyena (*Crocuta Crocuta*) compete for the same resources in the Maasai Mara National Reserve in Narok County, Kenya. Picture "Hyänen und Löwe im Morgenlicht" by lubye134, licensed under Creative Commons Attribution 2.0 Generic (BY 2.0).

Figure 13.2: Mutualism, predation and competition in population dynamics

(+,-) = predation: for $a_{ij} > 0$ and $a_{ji} < 0$, the species are in a predator-prey or host-parasite relationship. In other words, the presence of a prey (or host) species j favors the growth of the predator (or parasite) species i, wheres the presence of the predator species has a negative effect on the growth of the prey.

(-,-) = competition: for $a_{ij} < 0$ and $a_{ji} < 0$, the two species compete for a common resources of sorts and have therefore a negative effect on each other.

Note: the typical availability of bounded resources suggests it is ecologically meaningful to assume that the interaction matrix A is Hurwitz and that, to model the setting in which species live in isolation, the diagonal entries a_{ii} are negative.

Scientific questions of interest include:

(i) Does the Lotka-Volterra system have equilibrium points? Are they stable?

(ii) How does the presence of mutualism, predation, and/or competition affect the dynamic behavior?

(iii) Does the model predict extinction or periodic evolution of species?

13.2 Virus propagation models

We now study the diffusion and propagation of infectious diseases over networks. The proposed models may be relevant also in the context of propagation of information/signals in a communication network, spread of rumors over a social network, and diffusion of innovations in competitive economic networks. In the interest of clarity, we begin with "lumped" variables, i.e., variables which represent an entire "well-mixed" population of nodes. We then introduce "distributed" variable models, i.e., network models.

We start by studying three low-dimensional deterministic models in which nodes may be in one of two or three states; see Figure 13.3. For the SI and SIS models, we say that an *epidemic outbreak* takes place if a small initial fraction of infected individuals leads to the contagion of a significant fraction of the population.

We say the system displays an *epidemic threshold* if epidemic outbreaks occur when some combined value of parameters and initial conditions are above critical values.
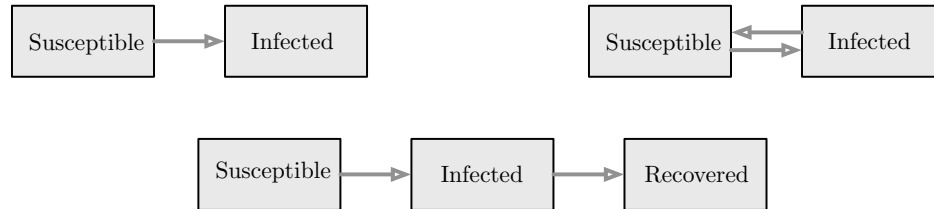
Figure 13.3: The three basic models SI, SIS and SIR for the propagation of an infectious disease

The SI model   Given a population, let $x(t)$ denote the fraction of infected individuals at time $t \in \mathbb{R}_{\geq 0}$. Similarly, let $s(t)$ denote the fraction of susceptible individuals. Clearly, $x(t) + s(t) = 1$ at all times. We model propagation via the following first-order differential equation, called the *susceptible–infected (SI) model*

$$\dot{x}(t) = \beta s(t)x(t) = \beta(1 - x(t))x(t), \tag{13.4}$$

where $\beta > 0$ is the *infection rate*. It is immediate to see that the SI model (13.4) is a logistic equation (13.1) with growth rate $r = \beta$ and carrying capacity $\kappa = 1$. As before, the solution from initial condition $x(0) = x_0 \in [0, 1]$ is

$$x(t) = \frac{x_0\, e^{\beta t}}{1 - x_0 + x_0\, e^{\beta t}}. \tag{13.5}$$

From all positive initial conditions $0 < x_0 < 1$, the solution $x(t)$ is monotonically increasing and converges to the unique equilibrium 1 as $t \to \infty$, as illustrated in Figure 13.4.
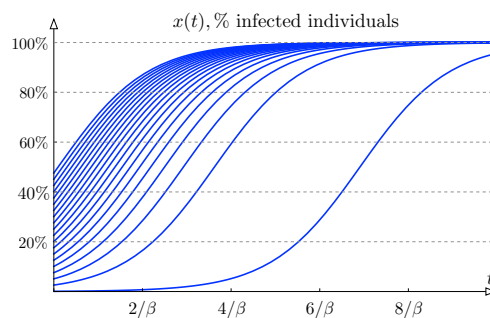
Figure 13.4: Evolution of the fraction of infected individuals in the (lumped deterministic) SI model

The SIS model   Next, we study a model in which individuals recover from the infection, but are susceptible to being re-infected. As in the SI model, the population is divided into two fractions with $s(t) + x(t) = 1$ and β is the infection rate. We model the recovery process via a constant *recovery rate* γ and write the *(susceptible–infected–susceptible) SIS model* as

$$\dot{x} = \beta sx - \gamma x = (\beta - \gamma - \beta x)x. \tag{13.6}$$

This SIS model is again a logistic equation with closed-form solution

$$x(t) = \frac{(\beta - \gamma)x_0}{\beta x_0 - e^{-(\beta-\gamma)t}(\gamma - \beta(1 - x_0))}, \tag{13.7}$$

from initial condition $x(0) = x_0 \in [0, 1]$ and for $\beta \neq \gamma$. Note, however, that there is a change now: it is possible for the carrying capacity $\beta/(\beta - \gamma)$ to be positive or negative. From the solution in equation (13.7) and from the simulations in Figure 13.5, one can observe the following two cases:

(i) if $\beta \leq \gamma$, all trajectories converge to the unique equilibrium $x = 0$ (i.e., the epidemic disappears), and

(ii) if $\beta > \gamma$, then, from all positive initial conditions $x(0) > 0$, all trajectories converge to the unique exponentially stable equilibrium $x = (\beta - \gamma)/\beta < 1$ (epidemic outbreak and steady-state epidemic contagion).

(a) SIS model with $\beta = 1 < \gamma = 2$ (b) SIS model with $\beta = 1 > \gamma = .5$
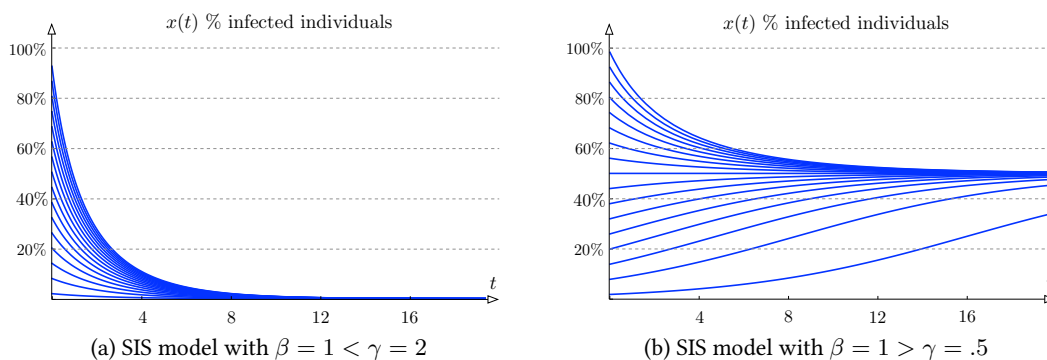
Figure 13.5: Evolution of the fraction of infected individuals in the (lumped deterministic) SIS model

The SIR model   As third and final lumped deterministic model, we study the setting in which individuals recover from the infection and are not susceptible to the epidemics after one round of infection. In other words, we assume the population is divided into three distinct groups: $s(t)$ denotes the fraction of susceptible individuals, $x(t)$ denotes the fraction of infected individuals, and $r(t)$ denotes the fraction of recovered individuals. Clearly, $s(t) + x(t) + r(t) = 1$. We model the recovery process via a constant *recovery rate* γ and write our *(susceptible–infected–recovered) SIR model* as

$$\begin{aligned}
\dot{s}(t) &= -\beta s(t)x(t), \\
\dot{x}(t) &= \beta s(t)x(t) - \gamma x(t), \\
\dot{r}(t) &= \gamma x(t).
\end{aligned} \tag{13.8}$$

Note that the first term is the same infection term as in the SI model and the second term is the same recovery term as in the SIS model.

We postpone the analysis of this mode till later, but illustrate its behavior via a simulation in Figure 13.6. Note the qualitatively different behavior for $\beta/\gamma = 4$ and $\beta/\gamma = 1/4$.

(a) SIR model with $\beta = 2$, $\gamma = 1/4$, initial infected 1%

(b) SIR model with $\beta = 2$, $\gamma = 4$, initial infected 30%
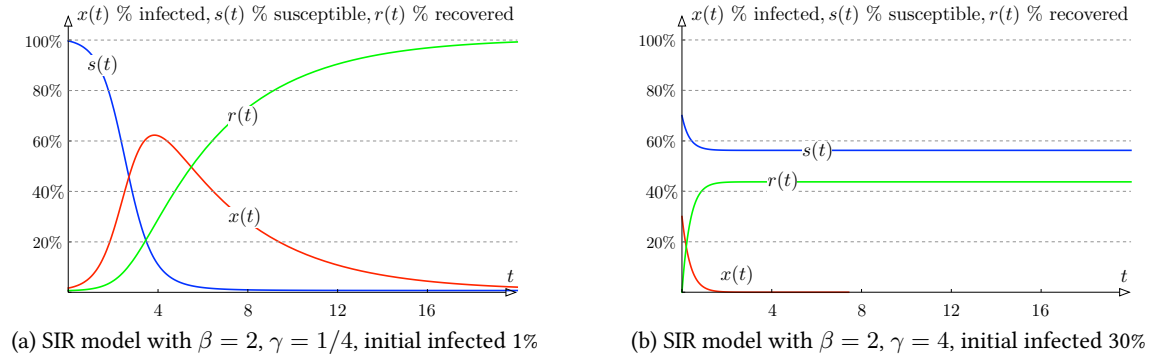
Figure 13.6: Evolution of the fraction of infected, susceptible and recovered individuals in the (lumped deterministic) SIR model (assuming zero recovered individuals at initial time)

The network/multigroup epidemic models   We conclude this section by presenting natural extensions of the three lumped SI/SIS/SIR scalar models to the setting of network multigroup models; originally due to (Lajmanovich and Yorke, 1976; Hethcote, 1978). In other words, we new present deterministic network models for the propagation of epidemics.

Two interpretations of the provided models are possible: if node i is a population of individuals at location i, then x_i can be interpreted as the infected fraction of that population. If node i is a single individual, then x_i can be interpreted as the probability that the individual is infected: $x_i(t) = \mathbb{P}[\text{individual } i \text{ is infected at time } t]$.

Consider an undirected weighted graph of order n with adjacency matrix A and degree matrix $D = \mathrm{diag}(A\mathbb{1}_n)$. The entries of A describe the frequency of contact among individuals; the graph is therefore referred to as a *contact network* (the nodes are individuals, the links are social contacts). Let $x_i(t) \in [0, 1]$ denote the fraction of infected individuals at node $i \in V$ at time $t \in \mathbb{R}_{\geq 0}$. Given an infection rate β, for $x \in [0, 1]^n$, the *network SI model* is

$$\dot{x}_i(t) = \beta(1 - x_i(t)) \sum_{j=1}^n a_{ij} x_j(t). \tag{13.9}$$

Next, given additionally a recovery rate γ, the *network SIS model* is

$$\dot{x}_i(t) = \beta(1 - x_i(t)) \sum_{j=1}^n a_{ij} x_j(t) - \gamma x_i(t). \tag{13.10}$$

And, finally, we consider the SIR model. Let $s_i(t)$, $x_i(t)$, $r_i(t) \in [0, 1]$ denote the fractions of susceptibile, infected and recovered individuals at node $i \in V$ at time $t \in \mathbb{R}_{\geq 0}$. The *network SIR model* is

$$
\begin{aligned}
\dot{s}_i(t) &= -\beta s_i(t) \sum_{j=1}^n a_{ij} x_j(t), \\
\dot{x}_i(t) &= \beta s_i(t) \sum_{j=1}^n a_{ij} x_j(t) - \gamma x_i(t), \\
\dot{r}_i(t) &= \gamma x_i(t).
\end{aligned}
\tag{13.11}
$$

These models are immediately written in equivalent vector form as:

$$\text{network SI:} \qquad \dot{x} = \beta\big(I_n - \text{diag}(x)\big)Ax, \tag{13.12}$$

$$\text{network SIS:} \qquad \dot{x} = \beta\big(I_n - \text{diag}(x)\big)Ax - \gamma x, \tag{13.13}$$

$$\text{network SIR:} \qquad \dot{s} = -\beta\,\text{diag}(s)Ax, \tag{13.14}$$

$$\dot{x} = \beta\,\text{diag}(s)Ax - \gamma x.$$

Note that the SIR system is completely specified by two equations and the constraint $s(t)+x(t)+r(t) = \mathbb{1}_n$. Scientific questions of interest for network epidemic models include:

(i) Do the network models have a behavior similar to the scalar models?

(ii) As a function of the model parameters, what possible asymptotic behaviors (e.g., vanishing infection, steady-state epidemic, full contagion) arise?

(iii) What is the transient propagation of epidemics starting from small initial fractions of infected nodes (epidemic outbreak or monotonically vanishing infection)?

13.3 Kuramoto coupled-oscillator models

In this section we introduce network of coupled oscillators and, in particular, phase-coupled oscillators. We start with two simple definitions. Given a connected, weighted, and undirected graph $G = (\{1,\ldots,n\}, E, A)$ and angles $\theta_1,\ldots,\theta_n$ associated to each node in the network, define the *coupled oscillators model* by

$$\dot{\theta}_i = \omega_i - \sum_{j=1}^{n} a_{ij}\sin(\theta_i - \theta_j), \qquad i \in \{1,\ldots,n\}. \tag{13.15}$$

A special case of this model is due to (Kuramoto, 1975); the *Kuramoto coupled oscillators model* is characterized by a complete homogeneous graph (i.e., a graph with identical edge weights $a_{ij} = K/n$ for all $i, j \in \{1,\ldots,n\}$ and for some *coupling strength K*):

$$\dot{\theta}_i = \omega_i - \frac{K}{n}\sum_{j=1}^{n}\sin(\theta_i - \theta_j), \qquad i \in \{1,\ldots,n\}. \tag{13.16}$$

Note: for $n = 2$, with the shorthands $\omega = \omega_1 - \omega_2$ and $a = a_{12} + a_{21}$, the coupled oscillator model can be written as a one-dimensional system in the difference variable $\theta = \theta_1 - \theta_2$ as:

$$\dot{\theta} = \omega - a\sin(\theta). \tag{13.17}$$

Coupled oscillator models arise naturally in many circumstances; in what follows we present three examples taken from (Dörfler and Bullo, 2014).

Example #1: A spring network on a ring   We start by studying a system of n dynamic particles constrained to rotate around a unit-radius circle and assumed to possibly overlap without ever colliding. Each particle is subject to (1) a non-conservative torque τ_i, (2) a linear damping torque, and (3) a total elastic torque. This system is illustrated in Figure 13.7.

We assume that pairs of interacting particles i and j are coupled through elastic springs with stiffness $k_{ij} > 0$; we set $k_{ij} = 0$ if the particles are not interconnected. The elastic energy stored by the spring between particles at angles θ_i and θ_j is

$$
\begin{aligned}
\mathrm{U}_{ij}(\theta_i, \theta_j) &= \frac{k_{ij}}{2}\mathrm{distance}^2 = \frac{k_{ij}}{2}\big((\cos\theta_i - \cos\theta_j)^2 + (\sin\theta_i - \sin\theta_j)^2\big) \\
&= k_{ij}\big(1 - \cos(\theta_i)\cos(\theta_j) - \sin(\theta_i)\sin(\theta_j)\big) = k_{ij}\big(1 - \cos(\theta_i - \theta_j)\big),
\end{aligned}
$$

so that the elastic torque on particle i is

$$
\mathrm{T}_i(\theta_i, \theta_j) = -\frac{\partial}{\partial\theta_i}\mathrm{U}_{ij}(\theta_i, \theta_j) = -k_{ij}\sin(\theta_i - \theta_j).
$$

Newton's Law applied to this rotating system implies that the network of spring-interconnected particles obeys the dynamics

$$
m_i\ddot{\theta}_i + d_i\dot{\theta}_i = \tau_i - \sum\nolimits_{j=1}^{n} k_{ij}\sin(\theta_i - \theta_j),
$$

where m_i and d_i are inertia and damping coefficients. In the limit of small masses m_i and uniformly-high viscous damping $d = d_i$, that is, $m_i/d \approx 0$, the model simplifies to the coupled oscillator network (13.15)

$$
\dot{\theta}_i = \omega_i - \sum_{j=1}^{n} a_{ij}\sin(\theta_i - \theta_j), \qquad i \in \{1, \dots, n\}.
$$

with natural rotation frequencies $\omega_i = \tau_i/d$ and with coupling strengths $a_{ij} = k_{ij}/d$.
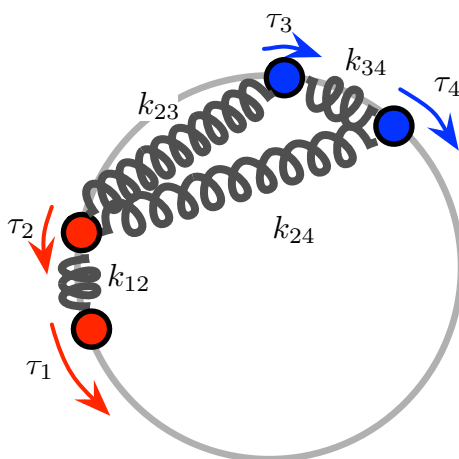
Figure 13.7: Mechanical analog of a coupled oscillator network

Example #2: The structure-preserving power network model As second example we consider an AC power network, visualized in Figure 13.8, with n buses including generators and load buses. We present two simplified models for this network, a static power-balance model and a dynamic continuous-time model.

(a) Line diagram

(b) Equivalent graph representation
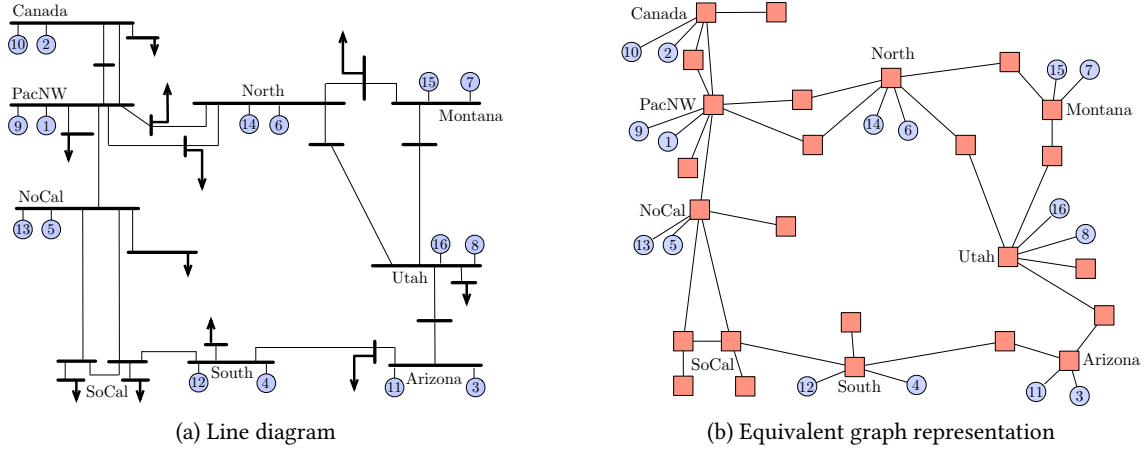
Figure 13.8: A simplified aggregated model with 16 generators and 25 load busses of the Western North American power grid, ofter referred to as the Western Interconnect. This model is often studied in the context of inter-area oscillations (Trudnowski et al., 1991). In the equivalent graph representation, generators are represented by light blue boxes and load buses by light red boxes.

The transmission network is described by an admittance matrix $Y \in \mathbb{C}^{n \times n}$ that is symmetric and sparse with line impedances $Z_{ij} = Z_{ji}$ for each branch $\{i, j\} \in E$. The network admittance matrix is sparse matrix with nonzero off-diagonal entries $Y_{ij} = -1/Z_{ij}$ for each branch $\{i, j\} \in E$; the diagonal elements $Y_{ii} = -\sum_{j=1, j \neq i}^{n} Y_{ij}$ assure zero row-sums.

The static model is described by the following two concepts. Firstly, according to Kirchhoff's current law, the current injection at node i is balanced by the current flows from adjacent nodes:

$$I_i = \sum_{j=1}^{n} \frac{1}{Z_{ij}} (V_i - V_j) = \sum_{j=1}^{n} Y_{ij} V_j \,.$$

Here, I_i and V_i are the *phasor representations* of the nodal current injections and nodal voltages, so that, for example, $V_i = |V_i| \, e^{i\theta_i}$ corresponds to the signal $|V_i| \cos(\omega_0 t + \theta_i)$. (Recall i $= \sqrt{-1}$.) The complex power injection $S_i = V_i \cdot \overline{I}_i$ (where $\overline{z}$ denotes the complex conjugate of $z \in \mathbb{C}$) then satisfies the power balance equation

$$S_i = V_i \cdot \sum_{j=1}^{n} \overline{Y}_{ij} \overline{V}_j = \sum_{j=1}^{n} \overline{Y}_{ij} |V_i| |V_j| e^{i(\theta_i - \theta_j)} \,.$$

Secondly, for a lossless network the real part of the power balance equations at each node is

$$\underbrace{P_i}_{\text{active power injection}} = \sum_{j=1}^{n} \underbrace{a_{ij} \cdot \sin(\theta_i - \theta_j)}_{\text{active power flow from } j \text{ to } i} \,, \quad i \in \{1, \ldots, n\}, \tag{13.18}$$

where $a_{ij} = |V_i||V_j||Y_{ij}|$ denotes the maximum power transfer over the transmission line $\{i, j\}$, and $P_i = \Re(S_i)$ is the active power injection into the network at node i, which is positive for generators and negative for loads. The systems of equations (13.18) are the *active power flow equations* at balance.

Next, we discuss a simplified dynamic model. Many appropriate dynamic models have been proposed for each network node: zeroth order (for so-called constant power loads), first-order models (for so-called frequency-dependent loads and inverter-based generators), and second and higher order for generators; see (Bergen and Hill, 1981). For extreme simplicity, we here assume that every node is described by a first-order integrator with the following intuition: node i speeds up (i.e., θ_i increases) when the power balance at node i is positive, and slows down (i.e., θ_i decreases) when the power balance at node i is negative. This assumption leads immediately to the coupled-oscillators model (13.15) written as:

$$\dot{\theta}_i = P_i - \sum_{j=1}^{n} a_{ij} \sin(\theta_i - \theta_j). \tag{13.19}$$

The systems of equations (13.19) are a first-order simplified version of the so-called coupled swing equations; see (Bergen and Hill, 1981). A more realistic model of power network necessarily include higher-order dynamics for the generators, uncertain load models, mixed resistive-inductive lines, and the modelling of reactive power.

Example #3: Flocking, schooling, and vehicle coordination   As third example, we consider a set of n kinematic particles in the plane $\mathbb{R}^2$, which we identify with the complex plane $\mathbb{C}$. Each particle $i \in \{1, \ldots, n\}$ is characterized by its position $r_i \in \mathbb{C}$, its heading angle $\theta_i \in \mathbb{S}^1$, and a steering control law $u_i(r, \theta)$ depending on the position and heading of itself and other vehicles, see Figure 13.9.(a). For simplicity, we assume that all particles have unit speed. The particle kinematics are then given by

$$\begin{aligned} \dot{r}_i &= e^{\mathrm{i}\theta_i}, \\ \dot{\theta}_i &= u_i(r, \theta), \end{aligned} \tag{13.20}$$

for $i \in \{1, \ldots, n\}$. If no control is applied, then particle i travels in a straight line with orientation $\theta_i(0)$, and if $u_i = \omega_i \in \mathbb{R}$ is a nonzero constant, then particle i traverses a circle with radius $1/|\omega_i|$.

The interaction among the particles is modeled by a interaction graph $G = (\{1, \ldots, n\}, E, A)$ determined by communication and sensing patterns. Interesting motion patterns emerge if the controllers use only relative phase information between neighboring particles. As we will discuss later, we may adopt potential functions-based gradient control strategies (i.e., negative gradient flows) to coordinate the relative heading angles $\theta_i(t) - \theta_j(t)$. As shown in Example #1, an intuitive extension of the quadratic elastic spring potential to the circle is the function $\mathrm{U}_{ij} : \mathbb{S}^1, \mathbb{S}^1 \to \mathbb{R}$ defined by

$$\mathrm{U}_{ij}(\theta_i, \theta_j) = a_{ij}(1 - \cos(\theta_i - \theta_j)),$$

for each edge $\{i, j\}$ of the graph. Note that the potential $\mathrm{U}_{ij}(\theta_i, \theta_j)$ achieves its unique minimum value if the heading angles θ_i and θ_j are synchronized and its unique maximum when θ_i and θ_j are out of phase by angle π.

These considerations motivate the gradient-based control strategy

$$\dot{\theta}_i = \omega_0 - K \frac{\partial}{\partial \theta_i} \sum_{\{i,j\} \in E} \mathrm{U}_{ij}(\theta_i - \theta_j) = \omega_0 - K \sum_{j=1}^{n} a_{ij} \sin(\theta_i - \theta_j), \quad i \in \{1, \ldots, n\}. \tag{13.21}$$

to synchronize the heading angles of the particles for $K > 0$ (gradient descent), respectively, to disperse the heading angles for $K < 0$ (gradient ascent). The term ω_0 can induce additional rotations (for $\omega_0 \neq 0$) or translations (for $\omega_0 = 0$). A few representative trajectories are illustrated in Figure 13.9.

The controlled phase dynamics (13.21) give rise to elegant and useful coordination patterns that mimic animal flocking behavior (Leonard et al., 2012) and fish schools. Inspired by these biological phenomena, scientists have studied the controlled phase dynamics (13.21) and their variations in the context of tracking and formation controllers in swarms of autonomous vehicles (Paley et al., 2007).
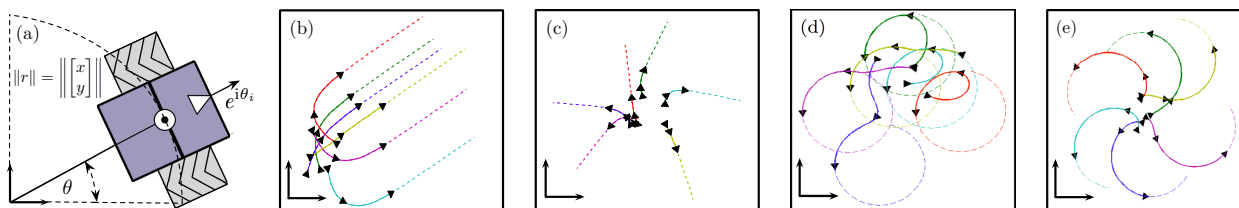
Figure 13.9: Figure (a) illustrates the particle kinematics (13.20). Figures (b)-(e) illustrate the controlled dynamics (13.20)-(13.21) with $n = 6$ particles, a complete interaction graph, and identical and constant natural frequencies: $\omega_0(t) = 0$ in figures (b) and (c) and $\omega_0(t) = 1$ in figures (d) and (e). The values of K are $K = +1$ in figures (b) and (d) and $K = -1$ in figure (c) and (e). The arrows depict the orientation, the dashed curves show the long-term position dynamics, and the solid curves show the initial transient position dynamics. As illustrated, the resulting motion displays synchronized or dispersed heading angles for $K = \pm 1$, and translational motion for $\omega_0 = 0$, respectively circular motion for $\omega_0 = 1$. Image reprinted from (Dörfler and Bullo, 2014) with permission from Elsevier.

Scientific questions of interest for coupled oscillator model include:

(i) When do the oscillators asymptotically achieve frequency synchronization, that is, when do they asymptotically reach an equal velocity?

(ii) When do they reach phase synchronization?

(iii) Are frequency (or phase) synchronized solutions stable and attractive in some sense?

13.4 Appendix: Stochastic propagation models

In this appendix, for readers with a background in probability theory, we discuss some models for stochastic virus propagation and show that their behavior in expectation is captured by the deterministic scalar models described above in Section 13.2.

We first consider a stochastic SI model. We consider a population of n individuals in which pairwise meetings between individuals take place. We assume the existence of a *meeting rate* $\beta_{\mathrm{m}} > 0$ such that, over the interval $(t, t + \Delta t)$, $n\beta_{\mathrm{m}}\Delta t$ individuals will meet on average other $n\beta_{\mathrm{m}}\Delta t$ individuals. Assuming meetings involve uniformly-selected individuals on average, over the interval $(t, t + \Delta t)$, there are $s(t)^2 n\beta_{\mathrm{m}}\Delta t$ meetings between a susceptible and another susceptible individual; these meetings, as well as meetings between infected individuals result in no epidemic propagation. However, there will also be $s(t)x(t)n\beta_{\mathrm{m}}\Delta t + x(t)s(t)n\beta_{\mathrm{m}}\Delta t$ meetings between a susceptible and an infected individual. We assume a fraction $\beta_{\mathrm{i}} \in [0, 1]$, called *transmission rate*, of these meetings results on average in the successful

transmission of the infection:

$$\beta_i \Big(s(t)x(t)n\beta_m\Delta t + x(t)s(t)n\beta_m\Delta t \Big) = 2\beta_i\beta_m x(t)s(t)n\Delta t.$$

In summary, based on these assumptions, the fraction of infected individuals satisfies on average

$$x(t + \Delta t) = x(t) + 2\beta_i\beta_m x(t)s(t)\Delta t.$$

Now it is immediate to see that the SI model (13.4) is the limit at $\Delta t \to 0^+$, where the infection parameter β is twice the product of meeting rate β_m and infection transmission fraction β_i.

The SIS and SIR models are also justified by showing that the constant recovery rate assumption corresponds to assuming a so-called Poisson recovery rate for the stochastic version of the SI model. This assumption is arguably not very realistic, but it leads to a simple analysis.

13.5   Exercises

E13.1   **Logistic ordinary differential equation.** Given a growth rate $r > 0$ and a carrying capacity $\kappa > 0$, consider the logistic equation (13.1) defined by

$$\dot{x} = rx(1 - x/\kappa),$$

with initial condition $x(0) \in \mathbb{R}_{\geq 0}$. Show that

   (i)   there are two equilibrium points 0 and κ,
   (ii)   the solution is

$$x(t) = \frac{\kappa x(0)\,\mathrm{e}^{rt}}{\kappa + x(0)(\mathrm{e}^{rt}-1)}, \tag{E13.1}$$

   and it takes value in $\mathbb{R}_{\geq 0}$,
   (iii)   all solutions with $0 < x(0) < \kappa$ are monotonically increasing and converge asymptotically to κ,
   (iv)   all solutions with $\kappa < x(0)$ are monotonically decreasing and converge asymptotically to κ, and
   (v)   if $x(0) < \kappa/2$, then the solution $x(t)$ has an inflection point when $x(t) = \kappa/2$.

E13.2   **Simulating coupled oscillators.** Simulate in your favorite programming language and software package the coupled Kuramoto oscillators in equation (13.16). Set $n = 10$, define a vector $\omega \in \mathbb{R}^{10}$ with entries deterministically uniformly-spaced between -1 and 1. Select random initial phases.

   (i)   Simulate the resulting differential equations for $K = 10$ and $K = 0.1$.
   (ii)   Find the approximate value of K at which the qualitative behavior of the system changes from asynchrony to synchrony.

   Turn in your code, a few printouts (as few as possible), and your written responses.

Stability Theory for Dynamical Systems

In this chapter we provide a brief self-contained review of stability theory for nonlinear dynamical systems. We review the key ideas and theorems in stability theory, including the Lyapunov Stability Criteria and the Krasovskiĭ-LaSalle Invariance Principle. We then apply these theoretical tools to a number of example systems, including linear and linearized systems, negative gradient systems, continuous-time averaging dynamics (i.e., the Laplacian flow) and positive linear systems described by Metzler matrices.

This chapter is not meant to provide a comprehensive treatment, e.g., we leave out matters of existence and uniqueness of solutions and we do not include proofs. Section 14.9 below provides numerous references for further reading. We start the chapter by introducing a running example with three prototypical dynamical systems.

Example 14.1 (Gradient and mechanical systems). We start by introducing a differentiable function $V : \mathbb{R} \to \mathbb{R}$; for example see Figure 14.1. Based on V and on two positive coefficients m and d, we define
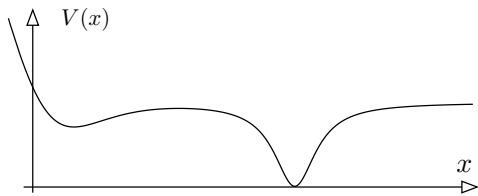
Figure 14.1: A differentiable function V playing the role of a potential energy function (i.e., a function describing the potential energy stored) in a negative gradient system, a conservative mechanical systems or a dissipative mechanical systems. Specifically, $V(x) = -x\,\mathrm{e}^{-x}/(1 + \mathrm{e}^{-x}) + (x - 10)^2/(1 + (x - 10)^2)$.

three instructive and prototypical dynamical systems:

$$\text{negative gradient system:} \qquad \dot{x} = -\frac{\partial V}{\partial x}(x), \tag{14.1}$$

$$\text{conservative mechanical system:} \qquad m\ddot{x} = -\frac{\partial V}{\partial x}(x), \tag{14.2}$$

$$\text{dissipative mechanical system:} \qquad m\ddot{x} = -\frac{\partial V}{\partial x}(x) - d\dot{x}. \tag{14.3}$$

In the study of physical systems, the parameter m is an inertia, d is a damping coefficient, and the function V is the potential energy function, describing the potential energy stored in the system.

These example are also know as a (first order, second order, or second order dissipative) particle on an energy landscape, or the "rolling ball on a hill" examples. According to Newton's law, the correct physical systems are models (14.2) and (14.3), but we will also see interesting examples of first-order negative gradient systems (14.1).   •

14.1   On sets and functions

Before proceeding we review some basic general properties of sets and functions. First, we recall that a set $W \subset \mathbb{R}^n$ is *bounded* if there exists a constant K that each $w \in W$ satisfies $\|w\| \leq K$, *closed* if it contains its boundary (or, equivalently, if it contains all its limit points), and *compact* if it is bounded and closed.

Second, given a differentiable function $V : \mathbb{R}^n \to \mathbb{R}$, a *critical point* of V is a point $x^* \in \mathbb{R}^n$ satisfying

$$\frac{\partial V}{\partial x}(x^*) = \mathbb{0}_n.$$

A critical point x^* is a *local minimum point* (resp. *local strict minimum point*) of V if there exists a distance $\varepsilon > 0$ such that $V(x^*) \leq V(x)$ (resp. $V(x^*) < V(x)$) for all $x \neq x^*$ within distance ε of x. The point x^* is a global minimum if $V(x^*) < V(x)$ for all $x \neq x^*$. Local and global maximum points are defined similarly.

Given a constant $\ell \in \mathbb{R}$, we define the ℓ-*level set* of V and the ℓ-*sublevel set* of V by

$$V^{-1}(\ell) = \{y \in \mathbb{R}^n \mid V(y) = \ell\}, \quad \text{and} \quad V_{\leq}^{-1}(\ell) = \{y \in \mathbb{R}^n \mid V(y) \leq \ell\}.$$

These notions are illustrated in Figure 14.2.

$$V_{\leq}^{-1}(\ell_2) = \{x \mid V(x) \leq \ell_2\}$$
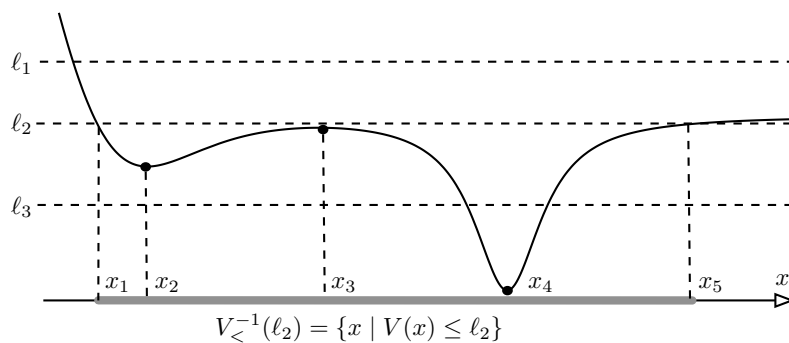
Figure 14.2: A differentiable function, its sublevel set and its critical points. The sublevel set $V_{\leq}^{-1}(\ell_1) = \{x \mid V(x) \leq \ell_1\}$ is unbounded. The sublevel set $V_{\leq}^{-1}(\ell_2) = [x_1, x_5]$ is compact and contains three critical points (x_2 and x_4 are local minima and x_3 is a local maximum). Finally, the sublevel set $V_{\leq}^{-1}(\ell_3)$ is compact and contains a single critical point, the global minimum x_4.

Third, given a point $x_0 \in \mathbb{R}^n$, a function $V : \mathbb{R}^n \to \mathbb{R}$ is

(i) *locally positive-definite* (resp. *positive-semidefinite*) *about* x_0 if $V(x_0) = 0$ and if there exists a neighborhood U of x_0 such that $V(x) > 0$ (resp. $V(x) \geq 0$) for all $x \in U \setminus \{x_0\}$,

(ii) *globally positive-definite about* x_0 if $V(x_0) = 0$ and $V(x) > 0$ for all $x \in \mathbb{R}^n \setminus \{x_0\}$, and

(iii) *locally* (resp. *globally*) *negative-definite* if $-V$ is *locally* (resp. *globally*) *positive-definite*; and *negative-semidefinite* if $-V$ is *positive-semidefinite*.

Note: Assume a differentiable V is locally positive-definite about x_0. Pick $\alpha > V(x_0)$. One can show that the sublevel set $V_{\leq}^{-1}(\alpha)$ contains a neighborhood of x_0. Indeed, in Figure 14.2, V is locally positive-definite about x_4 and $V_{\leq}^{-1}(\ell_2)$ and $V_{\leq}^{-1}(\ell_3)$ are both compact intervals containing x_4.

Fourth and finally, a non-negative continuous function $V : X \to \mathbb{R}_{\geq 0}$ is

(i) *radially unbounded* if $X = \mathbb{R}^n$ and $V(x) \to \infty$ along any trajectory such that $\|x\| \to \infty$, i.e., for any sequence $\{x_n\}_{n \in \mathbb{N}}$ satisfying $\lim_{n \to \infty} \|x_n\| = \infty$ we have $\lim_{n \to \infty} V(x_n) = \infty$, and

(ii) *proper* if, for all $\ell \in \mathbb{R}$, the ℓ-sublevel set of V is compact.

We illustrate these concepts in Figure 14.3 and state a useful equivalence without proof.

Lemma 14.2. *A continuous function $V : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ is proper if and only if it is radially unbounded.*

(a) This function $V : \mathbb{R} \to \mathbb{R}$ is not radially unbounded because $\lim_{x \to +\infty} V(x) = 1$.

(b) The function $V_{\text{lin-log}} : \mathbb{R}_{>0} \to \mathbb{R}$ is proper on $X = \mathbb{R}_{>0}$ since each sublevel set is a compact interval.
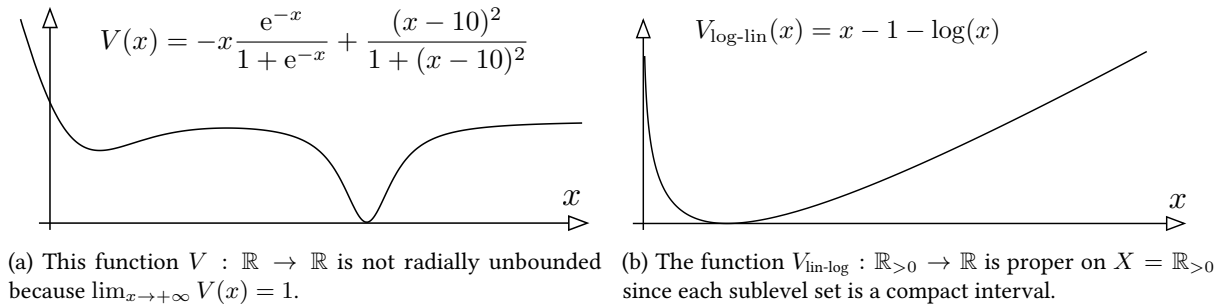
Figure 14.3: Example proper and not proper functions

14.2 Dynamical systems and stability notions

Dynamical systems

A *(continuous-time) dynamical system* is a pair (X, f) where X, called the *state space*, is a subset of $\mathbb{R}^n$ and f, called the *vector field*, is a map from X to $\mathbb{R}^n$. Given an initial state $x_0 \in X$, the *solution* (also called *trajectory* or *evolution*) of the dynamical system is a curve $t \mapsto x(t) \in X$ satisfying the differential equation

$$\dot{x}(t) = f(x(t)), \quad x(0) = x_0.$$

A dynamical system (X, f) is *linear* if $x \mapsto f(x) = Ax$ for some square matrix A.

Typically, the map f is assumed to have some continuity properties so that the solution exists and is unique for at least small times. Moreover, some of our examples are defined on closed submanifolds of $\mathbb{R}^n$ (e.g., the Lotka-Volterra model (13.3) is defined over the positive orthant $\mathbb{R}_{\geq 0}^n$, the network SIS model (13.10) is defined over the hypercube $[0, 1]^n$, and the coupled oscillator model (13.15) is defined over the set of n angles) and additional assumptions are required to ensure that the solution exists for all times in X. We do

not discuss these topics in great detail here, we simply assume the systems admit solutions inside X for all time, and refer to the references in Section 14.9 below.

Equilibrium points and their stability

An *equilibrium point* for the dynamical systems (X, f) is a point $x^* \in X$ such that $f(x^*) = \mathbb{0}_n$. If the initial state is $x(0) = x^*$, then the solution exists unique for all time and is constant: $x(t) = x^*$ for all $t \in \mathbb{R}_{\geq 0}$.

An equilibrium point x^* for the dynamical system (X, f) is

(i) *stable* (or *Lyapunov stable*) if, for each $\varepsilon > 0$, there exists $\delta = \delta(\varepsilon) > 0$ so that if $\|x(0) - x^*\| < \delta$, then $\|x(t) - x^*\| < \varepsilon$ for all $t \geq 0$,

(ii) *unstable* if it is not stable, and

(iii) *locally asymptotically stable* if it is stable and if there exists $\delta > 0$ such that $\lim_{t \to \infty} x(t) = x^*$ for all trajectories satisfying $\|x(0) - x^*\| < \delta$.

These three concepts are illustrated in Figure 14.4.
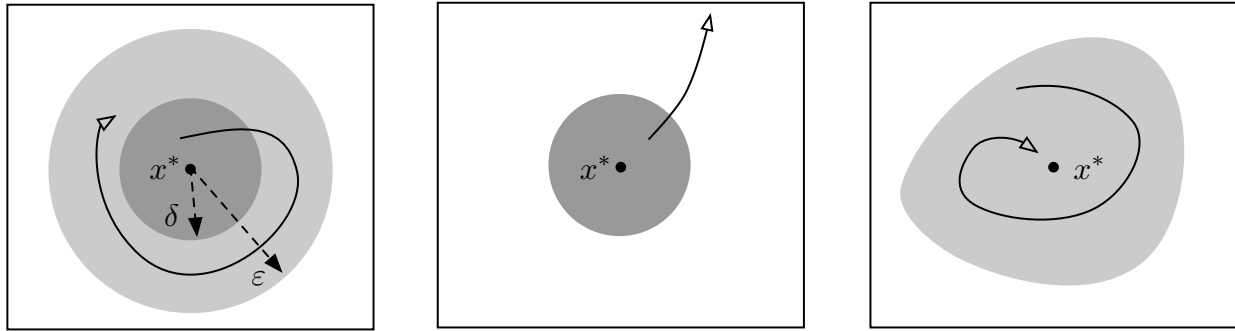
(a) Stable equilibrium: for all ε, each solution inside a sufficiently small δ-disk remains inside the ε-disk.

(b) Unstable equilibrium: no matter how small δ is, at least one solution starting inside the δ-disk diverges.

(c) Asymptotically stable equilibrium: solutions starting in a sufficiently small δ-disk converge asymptotically to the equilibrium.

Figure 14.4: Illustrations of a stable, an unstable and an asymptotically stable equilibrium.

These first three notions are local in nature. To characterize global properties of a dynamical system (X, f), we introduce the following notions. Given a locally asymptotically stable equilibrium point x^*,

(i) the set of initial conditions $x_0 \in X$ whose corresponding solution $x(t)$ converges to x^* is called the *region of attraction* of x^*,

(ii) x^* is said to be *globally asymptotically stable* if its region of attraction is the whole space X, and

(iii) x^* is said to be *globally* (respectively, *locally*) *exponentially stable* if it is globally (respectively, locally) asymptotically stable and there exist positive constants c_1 and c_2 such that all trajectories starting in the region of attraction satisfy

$$\|x(t) - x^*\| \leq c_1 \|x(0) - x^*\| \, e^{-c_2 t} .$$

Example 14.3 (Gradient and mechanical systems: Example 14.1 continued). It is instructive to report some numerical simulations of the three dynamical systems and state some conjectures about their equilibria and stability properties. These conjectures will be established in the next section.

(a) Conjecture: solutions converge to one of the two local minima.

(b) Sketch of the motion on the potential energy surface.

(c) Conjecture: solutions oscillate around a local minimum or diverge.

(d) Sketch of the motion on the potential energy surface.

(e) Conjecture: solutions converge to one of the two local minima.

(f) Sketch of the motion on the potential energy surface.
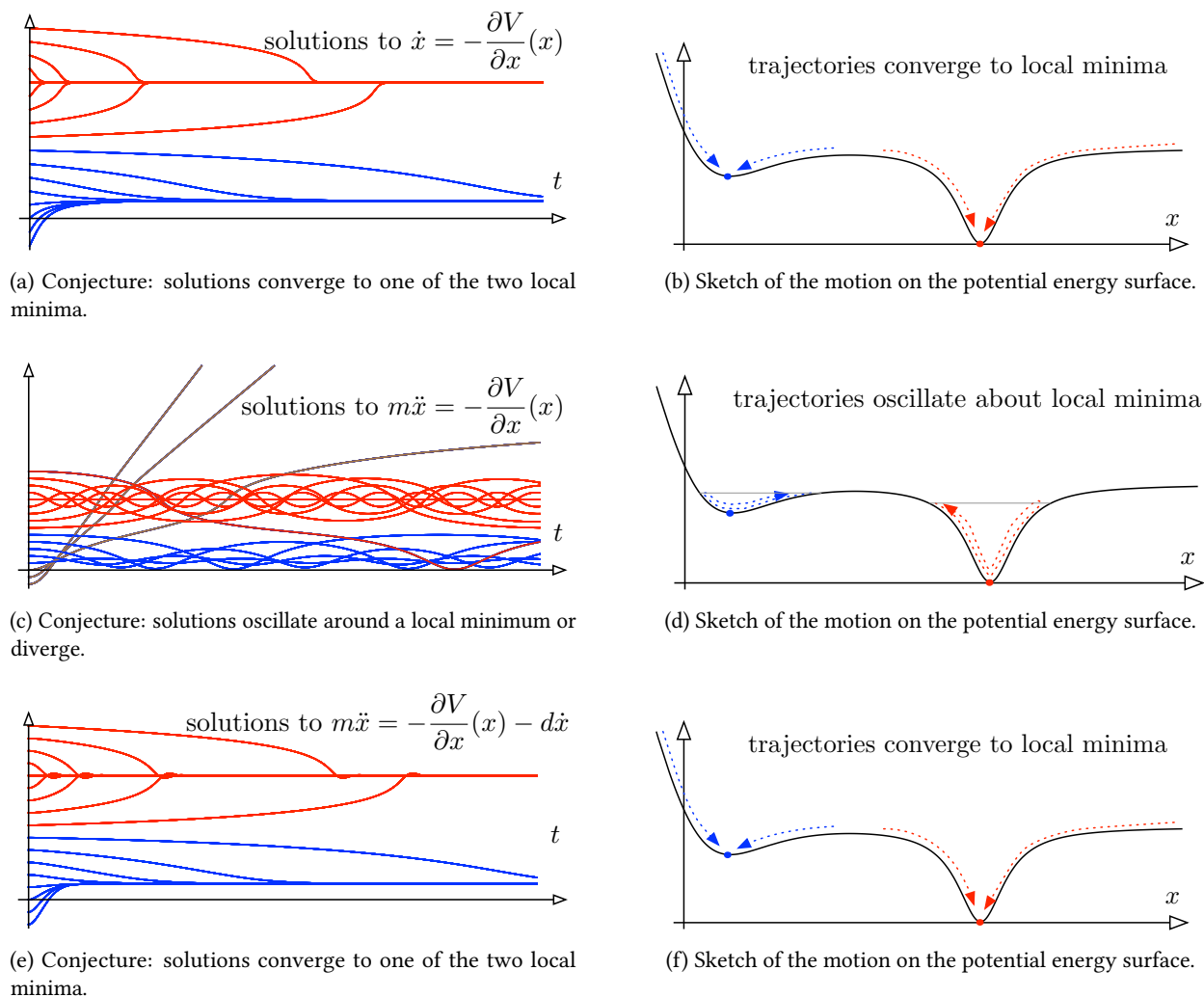
Figure 14.5: Numerically computed solutions (left) and graphical visualization of the solutions (right) for the three example systems with potential energy function V. Parameters are $x(0) \in \{-2, -1, \ldots, 14\}$ and $m = d = 1$.

•

14.3 First main convergence tool: the Lyapunov Stability Criteria

We are now ready to provide a critical tool in the study of the stability and convergence properties of a dynamical system. Roughly speaking, Lyapunov's idea is to use the concept of an energy function with a local/global minimum that is non-increasing along the system's solution.

Before proceeding, we require one final useful notion. The *Lie derivative* (also called the *directional derivative*) of a differentiable function $V : \mathbb{R}^n \to \mathbb{R}$ with respect to a vector field $f : \mathbb{R}^n \to \mathbb{R}^n$ is the function $\mathscr{L}_f V : \mathbb{R}^n \to \mathbb{R}$ defined by

$$\mathscr{L}_f V(x) = \frac{\partial V}{\partial x}(x)f(x) = \sum_{i=1}^{n} \frac{\partial V}{\partial x_i}(x)f_i(x). \tag{14.4}$$

Note that, along the flow of a dynamical system (X, f), we have $\dot{V}(x(t)) = \mathscr{L}_f V(x(t))$. Therefore, $V : \mathbb{R}^n \to \mathbb{R}$ is non-increasing along every trajectory of (X, f) if each solution $x : \mathbb{R}_{\geq 0} \to X$ satisfies

$$\dot{V}(x(t)) = \mathscr{L}_f V(x(t)) \leq 0,$$

or, equivalently, if each point $x \in X$ satisfies $\mathscr{L}_f V(x) \leq 0$. Because of this last inequality, when the vector field f is clear from the context, it is customary to adopt a slight abuse of notation and write $\dot{V}(x) = \mathscr{L}_f V(x)$.

We are now ready to present the main result of this section.

Theorem 14.4 (Lyapunov Stability Criteria). *Consider a dynamical system $(\mathbb{R}^n, f)$ with differentiable vector field f and with an equilibrium point $x^* \in \mathbb{R}^n$. The equilibrium point x^* is*

stable *if there exists a continuously-differentiable function $V : \mathbb{R}^n \to \mathbb{R}$, called a* weak Lyapunov function, *satisfying*

    (L1) *V is locally positive-definite about x^*,*

    (L2) *$\mathscr{L}_f V$ is locally negative-semidefinite about x^*;*

locally asymptotically stable *if there exists a continuously-differentiable function $V : \mathbb{R}^n \to \mathbb{R}$, called a* local Lyapunov function, *satisfying Assumption* (L1) *and*

    (L3) *$\mathscr{L}_f V$ is locally negative-definite about x^*;*

globally asymptotically stable *if there exists a continuously-differentiable function $V : \mathbb{R}^n \to \mathbb{R}$, called a* global Lyapunov function, *satisfying*

    (L4) *V is globally positive-definite about x^*,*

    (L5) *$\mathscr{L}_f V$ is globally negative-definite about x^*,*

    (L6) *V is proper.*

Note the immediate implications: (L4) $\implies$ (L1) and (L5) $\implies$ (L3) $\implies$ (L2).

Note: Theorem 14.4 assumes the existence of a Lyapunov function with certain properties, but does not provide constructive methods to design or compute one. In what follows we will see that Lyapunov functions can be designed for certain classes of systems. But, in general, the design of Lyapunov function is challenging. A common procedure is based on trial-and-error: one selects a so-called *candidate Lyapunov function* and verifies which, if any, of the properties (L1)–(L6) is satisfied.

Example 14.5 (Gradient and mechanical systems: Example 14.3 continued). We now apply the Lyapunov Stability Criteria in Theorem 14.4 to the example dynamical systems in Exercise 14.1. Based on the properties of the function V in Figures 14.2 with local minimum points x_2 and x_4, we establish most of the conjectures from Exercise 14.3. Note that the vector fields and the Lyapunov functions we adopt in what follows are all continuously differentiable.

Negative gradient systems: For the dynamics $\dot{x} = -\partial V/\partial x(x)$, as candidate Lyapunov function about x_2, we consider the function $V(x) - V(x_2)$. We compute

$$\dot{V}(x) = -\|\partial V/\partial x\|^2 \leq 0.$$

Note that $V - V(x_2)$ is locally positive definite about x_2 (Assumption (L1)) and $\dot{V}$ is locally negative definite about x_2 (Assumption (L3)); hence $V - V(x_2)$ is a local Lyapunov function for the equilibrium point x_2. An identical argument applies to x_4. Hence, both local minima x_2 and x_4 are locally asymptotically stable;

Conservative and dissipative mechanical systems: Given a positive inertia m and a non-negative damping coefficient $d \geq 0$, we write the conservative and the dissipative mechanical systems in first order form as:

$$\dot{x} = v, \quad m\dot{v} = -dv - \frac{\partial V}{\partial x}(x),$$

where $(x, v) \in \mathbb{R}^2$ are the position and velocity coordinates. As candidate Lyapunov function about the equilibrium point $(x_2, 0)$, we consider the *mechanical energy* $\mathrm{E} : \mathbb{R} \times \mathbb{R} \to \mathbb{R}_{\geq 0}$ given by the sum of kinetic and potential energy:

$$\mathrm{E}(x, v) = \frac{1}{2}mv^2 + V(x).$$

We compute its derivative along trajectories of the considered mechanical system as follows:

$$\dot{\mathrm{E}}(x, v) = mv\dot{v} + \frac{\partial V}{\partial x}(x)\dot{x} = v\left(-dv - \frac{\partial V}{\partial x}(x)\right) + \frac{\partial V}{\partial x}(x)v = -dv^2 \leq 0.$$

This calculation, and x_2 being a local minimum of V, together establish that, for $d \geq 0$, the function $\mathrm{E} - V(x_2)$ is locally positive definite about x_2 (Assumption (L1)) and $\dot{\mathrm{E}}$ is locally negative semidefinite about x_2 (Assumption (L2)). Hence, the function $\mathrm{E} - V(x_2)$ is a weak Lyapunov function for the equilibrium point $(x_2, 0)$ and, therefore, the point $(x_2, 0)$ is stable for both the conservative and the dissipative mechanical system. An identical argument applies to the point $(x_4, 0)$.

Note that we obtain the correct properties, i.e., consistent with the simulations in the previous exercise, for negative gradient system and for the conservative mechanical system. But more work is required to show that the local minima are locally asymptotically stable for the dissipative mechanical system. •

Example 14.6 (The logistic equation). As second example, we consider the logistic equation (13.1):

$$\dot{x}(t) = rx(t)\left(1 - \frac{x(t)}{\kappa}\right) =: f_{\text{logistic}}(x),$$
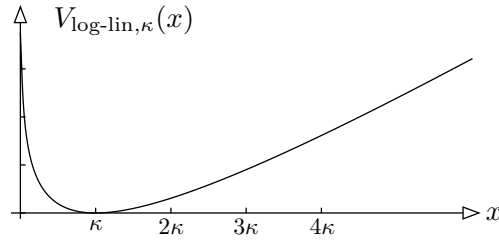
Figure 14.6: The function $V_{\text{log-lin},\kappa}(x) = x - \kappa - \kappa \log(x/\kappa)$, with $\kappa > 0$.

with growth rate r and carrying capacity κ. We neglect the possible initial condition $x(0) = 0$ (with subsequent equilibrium solution $x(t) = 0$ for all $t \geq 0$) and restrict out attention to solutions in $X = \mathbb{R}_{>0}$.

For $\kappa > 0$, define the *logarithmic-linear function* $V_{\text{log-lin},\kappa} : \mathbb{R}_{>0} \to \mathbb{R}$, illustrated in Figure 14.6, by

$$V_{\text{log-lin},\kappa}(x) = x - \kappa - \kappa \log\left(\frac{x}{\kappa}\right).$$

In Exercise E14.1 we ask the reader to verify that

(i) $V_{\text{log-lin},\kappa}$ is continuously differentiable with $\frac{d}{dx}V_{\text{log-lin},\kappa}(x) = (x - \kappa)/x$,

(ii) $V_{\text{log-lin},\kappa}(x) \geq 0$ for all $x > 0$ and $V_{\text{log-lin},\kappa}(x) = 0$ if and only if $x = \kappa$, and

(iii) $\lim_{x \to 0^+} V_{\text{log-lin},\kappa}(x) = \lim_{x \to \infty} V_{\text{log-lin},\kappa}(x) = +\infty$.

Next we compute

$$\mathscr{L}_{f_{\text{logistic}}}V_{\text{log-lin},\kappa}(x) = \frac{x - \kappa}{x} \cdot rx\left(1 - \frac{x}{\kappa}\right) = -\frac{r}{\kappa}(x - \kappa)^2.$$

In summary, we have established that f_{logistic} is a differentiable vector field, $x^* = \kappa$ is an equilibrium point, $V_{\text{log-lin},\kappa}$ is globally positive definite about κ, $\mathscr{L}_{f_{\text{logistic}}}V_{\text{log-lin},\kappa}$ is globally negative definite about κ, and $V_{\text{log-lin},\kappa}$ is proper. Hence, $V_{\text{log-lin},\kappa}$ is a global Lyapunov function and $x^* = \kappa$ is globally asymptotically stable. (This result is consistent with the behavior characterized in Exercise E13.1.)   •

14.4   Second main convergence tool: the Krasovskiĭ-LaSalle Invariance Principle

While the Lyapunov Stability Criteria are very useful, it is sometimes difficult to find a Lyapunov function with a negative-definite Lie derivative. To overcome this obstacle, in this section we introduce a powerful tool for the convergence analysis, namely the Krasovskiĭ-LaSalle Invariance Principle.

Before stating the main result, we introduce two useful concepts:

(i) A curve $t \mapsto x(t)$ *approaches* a set $S \subset \mathbb{R}^n$ as $t \to +\infty$ if the distance[1] from $x(t)$ to the set S converges to 0 as $t \to +\infty$.

If the set S consists of a single point s and $t \mapsto x(t)$ approaches S, then $t \mapsto x(t)$ converges to s in the usual sense: $\lim_{t \to +\infty} x(t) = s$. If the set S consists of multiple disconnected components and

---

[1]Here we define the distance from a point y to a set Z to be $\inf_{z \in Z} \|y - z\|$.

$t \mapsto x(t)$ approaches S, then $t \mapsto x(t)$ must approach one of the disconnected components of S. Specifically, if the set S is composed of a finite number of points, then $t \mapsto x(t)$ must converge to one of the points.

(ii) Given a dynamical system (X, f), a set $W \subset X$ is *invariant* if each solution starting in W remains in W, that is, if $x(0) \in W$ implies $x(t) \in W$ for all $t \geq 0$.

For example, any sublevel set of a function is invariant for the corresponding negative gradient flow. We are now ready to present the main result of this section.

Theorem 14.7 (Krasovskiĭ-LaSalle Invariance Principle). *For a dynamical system (X, f) with differentiable f, assume that*

(KL1) *all trajectories of (X, f) are bounded,*

(KL2) *there exists a closed set $W \subset X$ that is invariant for (X, f), and*

(KL3) *there exists a continuously-differentiable function $V : X \to \mathbb{R}$ satisfying $\mathscr{L}_f V(x) \leq 0$ for all $x \in X$.*

Then for each solution $t \mapsto x(t)$ starting in W there exists $c \in \mathbb{R}$ such that x converges to the largest invariant set contained in

$$\big\{ x \in W \mid \mathscr{L}_f V(x) = 0 \big\} \cap V^{-1}(c).$$

Note: if the closed invariant set $W \subset X$ in Assumption (KL2) is also bounded, then Assumption (KL1) is automatically satisfied.

Note: unlike in the Lyapunov Stability Criteria, the Krasovskiĭ-LaSalle Invariance Principle does not require the function V to be locally positive definite and establishes certain asymptotic convergence properties without requiring the Lie derivative of V to be locally negative definite.

Note: in some examples it is sometimes sufficient for one's purposes to show that $x(t) \to \big\{ x \in W \mid \mathscr{L}_f V(x) = 0 \big\}$. In other cases, however, one really needs to analyze the largest invariant set inside $\big\{ x \in W \mid \mathscr{L}_f V(x) = 0 \big\}$.

Note: If the largest invariant set is the union of multiple disjoint non-empty sets, then the solution to the negative gradient flow must converge to one of these disjoint sets.

Example 14.8 (Gradient and mechanical systems: Example 14.5 continued). We continue the analysis of the example dynamical systems in Exercises 14.1 and 14.5. Specifically, we sharpen here our results about the dissipative mechanical system about a local minimum point x_2 (or x_4) based on Krasovskiĭ-LaSalle Invariance Principle.

First, we note that the assumptions of the Krasovskiĭ-LaSalle Invariance Principle in Theorem 14.7 are satisfied:

(i) the function E and the vector field (the right-hand side of the mechanical system) are continuously differentiable;

(ii) the derivative $\dot{\text{E}}$ is locally negative semidefinite; and

(iii) for any initial condition $(x_0, v_0) \in \mathbb{R}^2$ sufficiently close to $(x_2, 0)$ the sublevel set $\{(x, v) \in \mathbb{R}^2 \mid \text{E}(x, v) \leq \text{E}(x_0, v_0)\}$ is compact due to the local positive definiteness of V at x_2.

It follows that $(x(t), v(t))$ converges to largest invariant set contained in

$$C = \{(x, v) \in \mathbb{R}^2 \mid \mathrm{E}(x, v) \leq \mathrm{E}(x_0, v_0), v = 0\} = \{(x, 0) \in \mathbb{R}^2 \mid \mathrm{E}(x, 0) \leq \mathrm{E}(x_0, v_0)\}.$$

A subset of C is invariant if any trajectory initiating in the subset remains in it. But this is only true if the starting position $\bar{x}$ satisfies $\frac{\partial}{\partial x} V(\bar{x}) = 0$, because otherwise the resulting trajectory would experience a strictly non-zero $\dot{v}(0)$ and hence leave C. In other words, the largest invariant set inside C is $\{(x, 0) \in \mathbb{R}^2 \mid \mathrm{E}(x, 0) \leq \mathrm{E}(x_0, v_0), \frac{\partial}{\partial x} V(x) = 0\}$. But the local minimum point x_2 is the unique critical point in the sublevel set and, therefore,

$$\lim_{t \to +\infty} (x(t), v(t)) = (x_2, 0).$$ ●

14.5 Application #1: Linear and linearized systems

It is interesting to study the convergence properties of a linear system. Recall that a symmetric matrix is positive definite if all its eigenvalues are strictly positive.

Theorem 14.9 (Convergence of linear systems). *For a matrix $A \in \mathbb{R}^{n \times n}$, the following properties are equivalent:*

(i) *each solution to the differential equation $\dot{x} = Ax$ satisfies $\lim_{t \to +\infty} x(t) = \mathbb{0}_n$,*

(ii) *A is Hurwitz, i.e., all the eigenvalues of A have strictly-negative real parts, and*

(iii) *for every positive-definite matrix Q, there exists a unique solution positive-definite matrix P to the so-called Lyapunov equation:*

$$A^\mathsf{T} P + PA = -Q.$$

Note: one can show that statement (iii) implies statement (i) using the Lyapunov Stability Criteria with function $V(x) = x^\mathsf{T} Px$, whose Lie derivative along the systems solutions is $\dot{V} = x^\mathsf{T}(A^\mathsf{T} P + PA)x = -x^\mathsf{T} Qx \leq 0$.

Next, we show a very useful way to apply linear stability methods to analyze the local stability of a nonlinear system.

The *linearization at the equilibrium point x^** of the dynamical system (X, f) is the linear dynamical system defined by the differential equation $\dot{x} = Ax$, where

$$A = \frac{\partial f}{\partial x}(x^*).$$

Theorem 14.10 (Convergence of nonlinear systems via linearization). *Consider a dynamical system (X, f) with an equilibrium point x^*, with twice differentiable vector field f, and with linearization A at x^*. The following statements hold:*

(i) *the equilibrium point x^* is locally exponentially stable if all the eigenvalues of A have strictly-negative real parts; and*

(ii) *the equilibrium point x^* is unstable if at least one eigenvalue of A has strictly-positive real part.*

Example 14.11 (Two coupled oscillators). For $\theta \in \mathbb{R}$, consider the dynamical system (13.17) arising from two coupled oscillators:

$$\dot{\theta} = f(\theta) = \omega - \sin(\theta).$$

If $\omega \in [0, 1[$, then there are two equilibrium points inside the range $\theta \in [0, 2\pi[$:

$$\theta_1^* = \arcsin(\omega) \in [0, \pi/2[, \qquad \text{and} \qquad \theta_2^* = \pi - \arcsin(\omega) \in]\pi/2, +\pi].$$

(Moreover, for $\theta \in \mathbb{R}$, the 2π-periodic set of equilibria are $\{\theta_1^* + 2k\pi \mid k \in \mathbb{Z}\}$ and $\{\theta_2^* + 2k\pi \mid k \in \mathbb{Z}\}$.) The linearization matrix $A(\theta_i^*) = \frac{\partial f}{\partial \theta}(\theta_i^*) = -\cos(\theta_i^*)$ for $i \in \{1, 2\}$ shows that θ_1^* is locally exponentially stable and θ_2^* is unstable. ●

Example 14.12 (A third order scalar system). Pick a scalar c and, for $x \in \mathbb{R}$, consider the dynamical system

$$\dot{x} = f(x) = c \cdot x^3.$$

The linearization at the equilibrium $x^* = 0$ is indefinite: $A(x^*) = 0$. Thus, Theorem 14.10 offers no conclusions other than the equilibrium cannot be exponentially stable. On the other hand, the Krasovskiĭ-LaSalle Invariance Principle shows that for $c < 0$ every trajectory converges to $x^* = 0$. Here, a non-increasing and differentiable function is given by $V(x) = x^2$ with Lie derivative $\mathscr{L}_f V(x) = -2cx^4 \leq 0$. Since $V(x(t))$ is non-increasing along the solution to the dynamical system, a compact invariant set is then readily given by any sublevel set $\{x \mid V(x) \leq \ell\}$ for $\ell \geq 0$. ●

14.6 Application #2: Negative gradient systems

We now summarize and extend the analysis given in Example 14.3 of the stability properties of negative gradient systems. Recall for convenience that, given a differentiable function $V : \mathbb{R}^n \to \mathbb{R}$, the *negative gradient flow* defined by V is the dynamical system

$$\dot{x}(t) = -\frac{\partial V}{\partial x}(x(t)). \tag{14.5}$$

We start by noting that, as in the Exercise, the Lie derivative of V along the negative gradient flow is

$$\mathscr{L}_{-\frac{\partial V}{\partial x}} V(x) = -\left\| \frac{\partial V}{\partial x}(x) \right\|^2 \leq 0,$$

and that, therefore, each sublevel set $V_{\leq}^{-1}(\ell)$, for $\ell \in \mathbb{R}$ is invariant (provided it is non-empty).

Given a twice differentiable function $V : \mathbb{R}^n \to \mathbb{R}$ and a point $x \in \mathbb{R}^n$, the *Hessian matrix of* V, denoted by $\text{Hess}\, V(x) \in \mathbb{R}^{n \times n}$, is the symmetric matrix of second order partial derivatives at x: $(\text{Hess}\, V)_{ij}(x) = \partial^2 V / \partial x_i \partial x_j(s)$. Given a critical point x^* of V, if the Hessian matrix $\text{Hess}\, V(x^*)$ is positive definite, then x^* is an isolate local minimum point of V. The converse is not true; as a counterexample, consider the function $V(x) = x^4$ and the critical point $x^* = 0$.

Theorem 14.13 (Convergence of negative gradient flows). *Let $V : \mathbb{R}^n \to \mathbb{R}$ be twice-differentiable and assume its sublevel set $V_{\leq}^{-1}(\ell) = \{x \in \mathbb{R}^n \mid V(x) \leq \ell\}$ is compact for some $\ell \in \mathbb{R}$. Then the negative gradient flow (14.5) has the following properties:*

(i) *each solution $t \mapsto x(t)$ starting in $V_{\leq}^{-1}(\ell)$ satisfies $\lim_{t \to +\infty} V(x(t)) = c$, for some $c \leq \ell$, and approaches the set of critical points of V:*

$$\left\{ x \in \mathbb{R}^n \mid \frac{\partial V}{\partial x}(x) = \mathbb{0}_n \right\},$$

(ii) *each local minimum point x^* is locally asymptotically stable and it is locally exponentially stable if and only if $\operatorname{Hess} V(x^*)$ is positive definite,*

(iii) *a critical point x^* is unstable if at least one eigenvalue of $\operatorname{Hess} V(x^*)$ is strictly negative,*

(iv) *if the function V is analytic, then every solution starting in a compact sublevel set has finite length (as a curve in $\mathbb{R}^n$) and converges to a single equilibrium point.*

Proof. To show statement (i), we verify that the assumptions of the Krasovskiĭ-LaSalle Invariance Principle are satisfied as follows. First, as set W we adopt the sublevel set $V_{\leq}^{-1}(\ell)$ which is compact by assumption and is invariant. Second we know the Lie derivative of V along the vector field is non-positive. Statement (i) is now an immediate consequence of the Krasovskiĭ-LaSalle Invariance Principle.

The statements (ii) and (iii) follow from observing that the linearization of the negative gradient system at the equilibrium x^* is the negative Hessian matrix evaluated at x^* and from applying Theorem 14.10.

Regarding statement (iv), we refer to the original source (Łojasiewicz, 1984). ∎

Note: If the function V has isolated critical points, then the negative gradient flow evolving in a compact set must converge to a single critical point. In such circumstances, it is also true that from almost all initial conditions the solution converges to a local minimum rather than a local maximum point or other critical points.

Note: If a twice-differentiable function V is strictly convex (as defined in Section 7.6), then its unique global minimum point x^* is globally exponentially stable; see Exercise E14.3.

14.7 Application #3: Continuous-time averaging systems and Laplacian matrices

In this section we revisit the continuous-time averaging system, i.e., the Laplacian flow, and study the evolution of the max-min function as a Lyapunov function.

As in Section 11.5, define the *max-min function* $V_{\text{max-min}} : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ by

$$V_{\text{max-min}}(x) = \max_{i \in \{1,\ldots,n\}} x_i - \min_{i \in \{1,\ldots,n\}} x_i, \tag{14.6}$$

and that $V_{\text{max-min}}(x) \geq 0$, and $V_{\text{max-min}}(x) = 0$ if an only if $x = \alpha \mathbb{1}_n$ for some $\alpha \in \mathbb{R}$.

Lemma 14.14 (The max-min function along the Laplacian flow). *Let $L \in \mathbb{R}^{n \times n}$ be the Laplacian matrix of a weighted digraph G. Let $x(t)$ be the solution to the Laplacian flow $\dot{x} = -Lx$. Then*

(i) *$t \mapsto V_{\text{max-min}}(x(t))$ is non-increasing,*

(ii) *if G has a globally reachable node, then $\lim_{t \to \infty} V_{\text{max-min}}(x(t)) = 0$ and $\lim_{t \to \infty} x(t) = \alpha \mathbb{1}_n$ for some $\alpha \in \mathbb{R}$.*

Numerous proofs for these results are possible (e.g., statement (ii) is established in Theorem 7.4). A second approach is to use the properties of the row-stochastic matrices $\exp(-Lt)$, $t \in \mathbb{R}_{\geq 0}$, as established in Theorem 7.2.

Here we pursue a strategy based on adopting $V_{\text{max-min}}$ as a weak Lyapunov function and, because $V_{\text{max-min}}$ is not continuously-differentiable, applying an appropriate generalization of the Krasovskiĭ-LaSalle Invariance Principle in Theorem 14.7. For our purposes here, it suffices to present the following concepts.

Definition 14.15. *The* upper right Dini derivative *and* upper left Dini derivative *of a continuous function* $f :]a, b[\to \mathbb{R}$ *at a point* $t \in]a, b[$ *are defined by, respectively,*

$$D^+ f(t) = \limsup_{\Delta t > 0, \Delta t \to 0} \frac{f(t + \Delta t) - f(t)}{\Delta t}, \quad and \quad D^- f(t) = \limsup_{\Delta t < 0, \Delta t \to 0} \frac{f(t + \Delta t) - f(t)}{\Delta t}.$$

Note that the sup operator is always defined (possibly equal to $+\infty$) and therefore so are the Dini derivatives.

Lemma 14.16 (Properties of the upper Dini derivatives). *Given a continuous function* $f :]a, b[\to \mathbb{R}$,

(i) *if f is differentiable at $t \in]a, b[$, then $D^+ f(t) = D^- f(t) = \frac{d}{dt} f(t)$ is the usual derivative of f at t,*

(ii) *if $D^+ f(t) \leq 0$ and $D^- f(t) \leq 0$ for all $t \in]a, b[$, then f is non-increasing on $]a, b[$, and*

(iii) *given differentiable functions $f_1, \ldots, f_m :]a, b[\to \mathbb{R}$, if*

$$f(t) = \max\{f_i(t) \mid i \in \{1, \ldots, m\}\},$$

then

$$D^+ f(t) = \max\left\{\frac{d}{dt} f_i(t) \mid i \in I(t)\right\}, \quad and \quad D^- f(t) = \min\left\{\frac{d}{dt} f_i(t) \mid i \in I(t)\right\},$$

where $I(t) = \{i \in \{1, \ldots, m\} \mid f_i(t) = f(t)\}$.

Note: statement (i) follows from the definition of derivative of a differentiable function. Statement (ii) is a consequence of Lemmas 1.3 and 1.4 in (Giorgi and Komlósi, 1992), where proofs are given. Statement (iii) is known as Danskin's Lemma. Given differentiable functions $f_1, \ldots, f_m$, a consequence of statements (ii) and (iii) is that the function $t \mapsto \max\{f_1(t), \ldots, f_m(t)\}$ is non-increasing on $]a, b[$ if $D^+ f(t) \leq 0$ for all $t \in]a, b[$.

Proof of Lemma 14.14. Let $x_{\max}(t) = \max(x(t))$ and $x_{\min}(t) = \min(x(t))$. For simplicity, let $\operatorname{argmax}(x(t)) = \{i \in \{1, \ldots, n\} \mid x_i(t) = x_{\max}(t)\}$ and $\operatorname{argmin}(x(t)) = \{i \in \{1, \ldots, n\} \mid x_i(t) = x_{\min}(t)\}$. Along the Laplacian flow $\dot{x}_i = \sum_{j=1}^{n} a_{ij}(x_j - x_i)$, Lemma 14.16(iii) (Danskin's Lemma) implies

$$D^+ V_{\text{max-min}}(x(t)) = \max\{\dot{x}_i(t) \mid i \in \operatorname{argmax}(x(t))\} - \min\{\dot{x}_i(t) \mid i \in \operatorname{argmin}(x(t))\}$$

$$= \max\left\{\sum_{j=1}^{n} a_{ij}(x_j - x_{\max}) \mid i \in \operatorname{argmax}(x(t))\right\}$$

$$- \min\left\{\sum_{j=1}^{n} a_{ij}(x_j - x_{\min}) \mid i \in \operatorname{argmin}(x(t))\right\},$$

where we have used $-\min(x) = \max(-x)$. Because $x_j - x_{\max} \leq 0$ and $x_j - x_{\min} \geq 0$ for all $j \in \{1, \ldots, n\}$, we have established that $D^+V_{\text{max-min}}(x(t))$ is the sum of two non-positive terms. This property, combined with Lemma 14.16(ii), implies that $t \mapsto V_{\text{max-min}}(x(t))$ is non-increasing, thereby completing the proof of statement (i).

To establish statement (ii) we invoke a generalized version of the Krasovskiĭ-LaSalle Invariance Principle 14.7. First, we note that statement (i) implies that any solution is bounded inside $[x_{\min}(0), x_{\max}(0)]^n$; this is a sufficient property (in lieu of the compactness of the set W). Second, we know the continuous function $V_{\text{max-min}}$ along the Laplacian flow is non-increasing (in lieu of the same property for a Lie derivative of a continuously-differentiable function). Therefore, we now know that there exists c such that the solution starting from $x(0)$ converges to the largest invariant set C contained in

$$\left\{ x \in [x_{\min}(0), x_{\max}(0)]^n \mid D^+V_{\text{max-min}}(x)\big|_{\dot{x}=-Lx} = 0 \right\} \cap V_{\text{max-min}}^{-1}(c).$$

Because $V_{\text{max-min}}$ is non-negative, we know $c \geq 0$. We now assume by absurd that $c > 0$, we let $y(t)$ be a trajectory originating in C, and we aim to show that $V_{\text{max-min}}(y(t))$ decreases along time (which is a contradiction because C is invariant).

Let k be a globally reachable node. Let i (resp. j) be an arbitrary index in $\operatorname{argmax}(y(0))$ (resp. $\operatorname{argmin}(y(0))$) so that $y_i(0) - y_j(0) = c > 0$. Without loss of generality we assume $y_k(0) < y_i(0)$. (Otherwise it would need to be $y_k(0) > y_j(0)$ and we would proceed similarly.) Recall we know $\dot{y}_i(0) \leq 0$. We now note that, if $\dot{y}_i(t) = 0$ for all $t \in (0, \varepsilon)$ for a positive ε, then the equation $\dot{y}_i = \sum_j a_{ij}(y_j - y_i)$ and the property $y_i(0) = \max y(0)$ together imply that $y_j(t) = y_i(t)$ for all $t \in (0, \varepsilon)$ and for all j such that $a_{ij} > 0$. Iterating this argument along the directed path from i to k, we get the contradiction that $y_k(t) = y_i(t)$ for all $t \in (0, \varepsilon)$. Therefore, we know that $\dot{y}_i(t) < 0$ for small times. Because i is an arbitrary index in $\operatorname{argmax}(y(0))$, we have proved that $t \mapsto \max y(t)$ is strictly decreasing for small times. This establishes that C is not invariant if $c > 0$ and completes the proof of statement (ii). ∎

14.8   Application #4: Positive linear systems and Metzler matrices

In this section we study the positive linear system $\dot{x} = Ax$, $x \in \mathbb{R}_{\geq 0}^n$, with equilibrium point $\mathbb{0}_n$, and with matrix A being Metzler matrix. To establish the stability properties of $\mathbb{0}_n$, we start by characterizing certain properties of Metzler matrices.

We recall from Section 9.2 the properties of Metzler matrices. For example the Perron-Frobenius Theorem 9.4 for Metzler matrices establishes the existence of a dominant eigenvalue. If the dominant eigenvalue is negative, then the Metzler matrix is Hurwitz; this case was studied in Theorem 9.5.

For the remainder of this section, given a symmetric matrix $A \in \mathbb{R}^{n \times n}$, we write $A \succ 0$ (resp. $A \prec 0$) if A is positive definite (resp. negative definite), that is, if all its eigenvalues are strictly positive (resp. strictly negative).

Theorem 14.17 (Properties of Hurwitz Metzler matrices: continued). *For a Metzler matrix $A \in \mathbb{R}^{n \times n}$, the following statements are equivalent:*

(i) *A is Hurwitz,*

(ii) *A is invertible and $-A^{-1} \geq 0$,*

(iii) *for all $b \geq \mathbb{0}_n$, there exists $x^* \geq \mathbb{0}_n$ solving $Ax^* + b = \mathbb{0}_n$,*

(iv) *there exists $\xi \in \mathbb{R}^n$ such that $\xi > \mathbb{0}_n$ and $A\xi < \mathbb{0}_n$,*

 (v) *there exists $\eta \in \mathbb{R}^n$ such that $\eta > \mathbb{0}_n$ and $\eta^{\mathsf{T}} A < \mathbb{0}^{\mathsf{T}}$, and*

(vi) *there exists a diagonal matrix $P \succ 0$ such that $A^{\mathsf{T}} P + PA \prec 0$.*

Note: if the vectors ξ and η satisfy the conditions of statements (iv) and (v) respectively, then the matrix $P = \mathrm{diag}(\eta_1/\xi_1, \ldots, \eta_n/\xi_n)$ satisfies the conditions of statement (vi).

Note: a matrix A with a diagonal matrix P as in statement (vi) is said to be *diagonally stable*.

Proof. The equivalence between statements (i), (ii), and (iii) is established in Theorem 9.5.

Statements (iv) and (v) are equivalent because of the following argument and its converse: if statement (iv) holds with $\xi = \xi(A)$, then statement (v) holds with $\eta = \xi(A^{\mathsf{T}})$.

We first prove that (ii) implies (iv). Set $\xi = -A^{-1}\mathbb{1}_n$. Because $-A^{-1} \geq 0$ is invertible, it can have no row identically equal to zero. Hence $\xi = -A^{-1}\mathbb{1}_n > \mathbb{0}_n$. Moreover $A\xi = -\mathbb{1}_n < \mathbb{0}_n$.

Next, we prove that (iv) implies (i). Let λ be an eigenvalue of A with eigenvector v. Define $w \in \mathbb{R}^n$ by $w_i = v_i/\xi_i$, for $i \in \{1, \ldots, n\}$, where ξ is as in statement (iv). We have therefore $\lambda \xi_i w_i = \sum_{j=1}^{n} a_{ij} \xi_j w_j$. If ℓ is the index satisfying $|w_\ell| = \max_i |w_i| > 0$, then

$$\lambda \xi_\ell = a_{\ell\ell} \xi_\ell + \sum_{j=1, j\neq \ell}^{n} a_{\ell j} \xi_j \frac{w_j}{w_\ell},$$

which, in turn, implies

$$|\lambda \xi_\ell - a_{\ell\ell} \xi_\ell| \leq \sum_{j=1, j\neq \ell}^{n} a_{\ell j} \xi_j \left| \frac{w_j}{w_\ell} \right| \leq \sum_{j=1, j\neq \ell}^{n} a_{\ell j} \xi_j < -a_{\ell\ell} \xi_\ell,$$

where the last equality follows from the ℓ-th row of the inequality $A\xi < \mathbb{0}_n$. Therefore, $|\lambda - a_{\ell\ell}| < -a_{\ell\ell}$. This inequality implies that the eigenvalue λ must belong to an open disc in the complex plan with center $a_{\ell\ell} < 0$ and radius $|a_{\ell\ell}|$. Hence, λ, together with all other eigenvalues of A, must have negative real part.

We now prove that (iv) implies (vi). From statement (iv) applied to A and A^{T}, let $\xi > \mathbb{0}_n$ satisfy $A\xi < \mathbb{0}_n$ and $\eta > \mathbb{0}_n$ satisfy $A^{\mathsf{T}}\eta < \mathbb{0}_n$. Define $P = \mathrm{diag}(\eta_1/\xi_1, \ldots, \eta_n/\xi_n)$ and consider the symmetric matrix $A^{\mathsf{T}} P + PA$. This matrix is Metzler and satisfies $(A^{\mathsf{T}} P + PA)\xi = A^{\mathsf{T}}\eta + PA\xi < \mathbb{0}_n$. Hence, $A^{\mathsf{T}} P + PA$ is negative diagonally dominant and, because (iv) $\implies$ (i), Hurwitz. In summary, $A^{\mathsf{T}} P + PA$ is symmetric and Hurwitz, hence, it is negative definite.

Finally, the implication (vi) $\implies$ (i) is established in Theorem 14.9. ∎

The following corollary illustrates how each of the conditions (iv), (v), and (vi) corresponds to a Lyapunov function of a specific form for a Hurwitz Metzler system.

Corollary 14.18 (Lyapunov functions for positive linear systems). *Let A be a Hurwitz Metzler matrix. The positive linear system $\dot{x} = Ax$, $x \in \mathbb{R}^n_{\geq 0}$, with equilibrium point $\mathbb{0}_n$, admits the following global Lyapunov functions:*

$$
\begin{aligned}
V_1(x) &= \max_{i \in \{1, \ldots, n\}} x_i/\xi_i, & &\text{for } \xi > 0 \text{ satisfying } A\xi < 0, \\
V_2(x) &= \eta^{\mathsf{T}} x, & &\eta > 0 \text{ satisfying } \eta^{\mathsf{T}} A < 0, \\
V_3(x) &= V(x) = x^{\mathsf{T}} P x, & &\text{for a diagonal matrix } P \succ 0 \text{ satisfying } A^{\mathsf{T}} P + PA \prec 0.
\end{aligned}
$$

We illustrate the level sets of these three Lyapunov functions in Figure 14.7.

(a) Rectangular level set   (b) Linear level set   (c) Quadratic level set
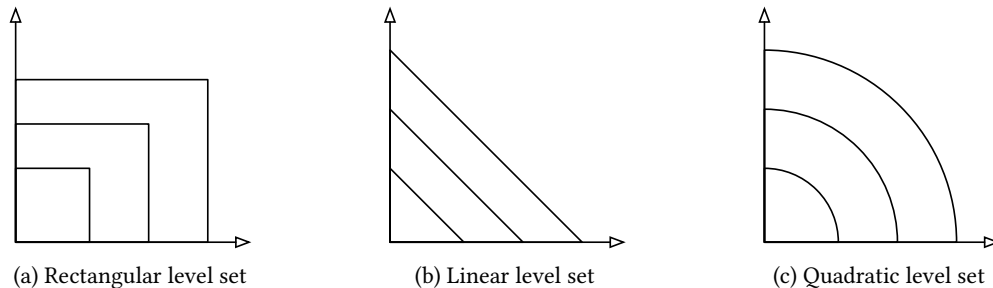
Figure 14.7: Level curves of Lyapunov functions for positive linear systems defined by Hurwitz Lyapunov matrix

14.9 Historical notes and further reading

Classic historical works on stability properties of physical systems include (Lagrange, 1788; Maxwell, 1868; Thomson and Tait, 1867). Modern stability theory started with the work by Lyapunov (1892), who proposed the key ideas towards a general treatment of stability notions and tests for nonlinear dynamical systems. Lyapunov's ideas were extended by Barbashin and Krasovskiĭ (1952); Krasovskiĭ (1963) and LaSalle (1960, 1968, 1976) through their work on invariance principles. Other influential works include (Chetaev, 1961; Hahn, 1967).

For comprehensive treatments, we refer the reader to the numerous excellent texts in this area, e.g., including the classic control texts (Sontag, 1998; Khalil, 2002; Vidyasagar, 2002), the classic dynamical systems texts (Hirsch and Smale, 1974; Arnol'd, 1992; Guckenheimer and Holmes, 1990), and the more recent works (Haddad and Chellaboina, 2008; Goebel et al., 2012; Blanchini and Miani, 2015).

This chapter has treated systems evolving in continuous time. Naturally, it is possible to develop a Lyapunov theory for discrete-time systems, even though remarkably there are only few references on this topic; see (LaSalle, 1976, Chapter 1).

Our treatment of Metzler matrices in Section 14.8 is inspired by the presentation in (Rantzer, 2015).

We refer to (Clarke et al., 1998; Cortés, 2008) for a comprehensive review of stability theory for nonsmooth systems and Lyapunov functions. Properties of the Dini derivatives are reviewed by Giorgi and Komlósi (1992). The usefulness of Dini derivatives in continuous-time averaging systems is highlighted for example by Lin et al. (2007); see also (Danskin, 1966) for Danskin's Lemma.

14.10 Exercises

E14.1 **The logarithmic-linear function.** For $\kappa > 0$, define the function $V_{\text{log-lin},\kappa} : \mathbb{R}_{>0} \to \mathbb{R}$ by

$$V_{\text{log-lin},\kappa}(x) = x - \kappa - \kappa \log\Big(\frac{x}{\kappa}\Big).$$

Show that

(i) $V_{\text{log-lin},\kappa}$ is continuously differentiable and $\frac{d}{dx} V_{\text{log-lin},\kappa}(x) = (x - \kappa)/x$,

(ii) $V_{\text{log-lin},\kappa}(x) \geq 0$ for all $x > 0$ and $V_{\text{log-lin},\kappa}(x) = 0$ if and only if $x = \kappa$, and

(iii) $\lim_{x \to 0^+} V_{\text{log-lin},\kappa}(x) = \lim_{x \to \infty} V_{\text{log-lin},\kappa}(x) = +\infty$.

E14.2 **Grönwall-Bellman Comparison Lemma.** Given a continuous function of time $t \mapsto a(t) \in \mathbb{R}$, suppose the signal $t \mapsto x(t)$ satisfies

$$\dot{x}(t) \leq a(t)x(t).$$

Define a new signal $t \mapsto y(t)$ satisfying $\dot{y}(t) = a(t)y(t)$. Show that

(i) $y(t) = y(0) \exp\big(\int_0^t a(\tau)d\tau\big)$, and

(ii) if $x(0) \leq y(0)$, then $x(t) \leq y(t)$.

E14.3 **The negative gradient flow of a strictly convex function.** Let $f : \mathbb{R}^n \to \mathbb{R}$ be a strictly convex and twice differentiable function. Show convergence of the associated negative gradient flow, $\dot{x} = -\frac{\partial}{\partial x} f(x)$, to the global minimizer x^* of f using the Lyapunov function candidate $V(x) = (x - x^*)^\mathsf{T}(x - x^*)$ and the Krasovskiĭ-LaSalle Invariance Principle in Theorem 14.7.
Hint: *Use the* global underestimate property *of a strictly convex function stated as follows:* $f(y) - f(x) > \frac{\partial}{\partial x} f(x)(y - x)$ *for all distinct x and y in the domain of f.*

E14.4 **Distributed optimization using the Laplacian flow.** Consider the saddle point dynamics (7.16) that solve the optimization problem (7.15) in a distributed fashion. Assume that the objective functions are strictly convex and twice differentiable and that the underlying communication graph among the distributed processors is connected and undirected. By using the Krasovskiĭ-LaSalle Invariance Principle show that all solutions of the saddle point dynamics converge to the set of saddle points.
Hint: *Use the following global underestimate property of a strictly convex function:* $f(y) - f(x) > \frac{\partial}{\partial x} f(x)(y - x)$ *for all distinct x and y in the domain of f; and the following global overestimate property of a concave function:* $g(y) - g(x) \leq \frac{\partial}{\partial x} g(x)(y - x)$ *for all distinct x and y in the domain of g. Finally, note that the overestimate property holds with equality $g(y) - g(x) = \frac{\partial}{\partial x} g(x)(y - x)$ if $g(x)$ is affine.*

E14.5 **Region of attraction for an example nonlinear systems.** Consider the nonlinear system

$$\dot{x}_1 = -2x_1 - 2x_2 - 4x_1^3 x_2^2,$$
$$\dot{x}_2 = -2x_1 - 2x_2 - 2x_1^4 x_2.$$

Is the origin locally asymptotically stable? What is the region of attraction?

E14.6 **A useful corollary by Barbashin and Krasovskiĭ (1952).** Consider a dynamical system $(\mathbb{R}^n, f)$ with differentiable vector field f and with an equilibrium point $x^* \in \mathbb{R}^n$.

Assume the continuously-differentiable $V : \mathbb{R}^n \to \mathbb{R}$ is a weak Lyapunov function, but not a local Lyapunov function (as defined in Theorem 14.4). In other words, assume V is locally positive-definite about x^* (Assumption (L1)) and $\mathscr{L}_f V$ is locally negative-semidefinite about x^* (Assumption (L2)), but $\mathscr{L}_f V$ is not locally negative-definite about x^* (Assumption (L3)). Then Lyapunov Theorem 14.4 implies that x^* is stable but not necessarily locally asymptotically stable.

Now, assume:

(L7) $\{x^*\}$ is the only positively invariant set in $\{x \in W \mid \mathscr{L}_f V(x) = 0\}$, where W be a neighborhood of x^* on which V is positive-definite and $\mathscr{L}_f V$ is negative-semidefinite.

Prove that Assumptions (L1), (L2) and (L7) imply the equilibrium point x^* is locally asymptotically stable.

E14.7 **Limit sets of dynamical systems.** Consider the following nonlinear dynamical system

$$\dot{x}_1 = 4x_1^2 x_2 - f_1(x_1)(x_1^2 + 2x_2^2 - 4), \tag{E14.1a}$$
$$\dot{x}_2 = -2x_1^3 - f_2(x_2)(x_1^2 + 2x_2^2 - 4), \tag{E14.1b}$$

where the differentiable functions $f_1(x)$, $f_2(x)$ have the same sign as their arguments, i.e., $x_i f_i(x_i) > 0$ if $x_i \neq 0$, $f_i(0) = 0$, and $f_i'(0) > 0$. This vector field exhibit some very unconventional limit sets. In what follows you will investigate this vector field and show that each trajectory converge to an equilibrium, but that none of the equilibria is Lyapunov stable.

(i) Show that $\mathcal{E} = \{x \in \mathbb{R}^2 \mid x_1^2 + 2x_2^2 = 4\}$ is an invariant set. Calculate the equilibria on the set $\mathcal{E}$.

(ii) Show that all trajectories converge either to the invariant set $\mathcal{E}$ or to the origin $(0, 0)$.

(iii) Determine the largest invariant set inside $\mathcal{E}$, such that all trajectories originating in $\mathcal{E}$ converge to that set.

(iv) Show that the origin $(0, 0)$ and all equilibria on $\mathcal{E}$ are unstable, i.e., not stable in the sense of Lyapunov. Sketch the vector field.

Lotka-Volterra Population Dynamics

In this chapter we study the behavior of the Lotka-Volterra population model, that was introduced in Section 13.1. First we illustrate the behavior of the 2-dimensional model via simple phase portraits. Then, using Lyapunov stability theory from Chapter 13 we provide sufficient conditions for the general n-dimensional model to have a globally asymptotically stable point. As a special case, we study the case of cooperative models.

Recall that the Lotka-Volterra vector field for $n \geq 2$ interacting species, as given in equation (13.3), is

$$\dot{x} = \operatorname{diag}(x)\big(Ax + r\big) =: f_{\mathrm{LV}}(x), \tag{15.1}$$

where the matrix $A = [a_{ij}]$ is called the interaction matrix, and the vector r is called the intrisic growth rate.

15.1 Two-species model and analysis

In this section we consider the *two-species Lotka-Volterra system*

$$\begin{aligned}
\dot{x}_1 &= x_1(r_1 + a_{11}x_1 + a_{12}x_2), \\
\dot{x}_2 &= x_2(r_2 + a_{21}x_1 + a_{22}x_2),
\end{aligned} \tag{15.2}$$

with scalar parameters (r_1, r_2) and $(a_{11}, a_{12}, a_{21}, a_{22})$. It is possible to fully characterize the dynamics behavior of this system as a function of the six scalar parameters. As explained in Section 13.1, to model bounded resources, our standing assumptions are:

$$r_i > 0, \quad \text{and } a_{ii} < 0, \text{ for } i \in \{1, 2\}.$$

We study various cases depending upon the sign of a_{12} and a_{21}.

To study the phase portrait of this two-dimensional system, it is establish the following details:

(i) along the axis $x_2 = 0$, there exists a unique non-trivial equilibrium point $x_1^* = -r_1/a_{11}$;

(ii) similarly, along the axis $x_1 = 0$, there exists a unique non-trivial equilibrium point $x_2^* = -r_2/a_{22}$;

(iii) the x_1-*null-line* is the set of points (x_1, x_2) where $\dot{x}_1 = 0$, that is, the line in the (x_1, x_2) plane defined by $r_1 + a_{11}x_1 + a_{12}x_2 = 0$;

(iv) similarly, the x_2-*null-line* is the (x_1, x_2) plane defined by $r_2 + a_{21}x_1 + a_{22}x_2 = 0$.

Clearly, the x_1-null-line (respectively the x_1-null-line) passes through the equilibrium point x_1^* (respectively x_2^*).

In what follows we study the cases of mutualistic interactions and competitive interactions. We refer to Exercise E15.2 for a specially-interesting case of predator-prey interactions.

15.1.1   Mutualism

Here we assume inter-species mutualism, that is, we assume both inter-species coefficients a_{12} and a_{21} are positive. We identify two distinct parameter ranges corresponding to distinct dynamic behavior and illustrate them in Figure 15.1.

Lemma 15.1 (Two-species mutualism). *Consider the two-species Lotka-Volterra system* (15.2) *with scalar parameters* (r_1, r_2) *and* $(a_{11}, a_{12}, a_{21}, a_{22})$. *Assume the interaction is mutualistic, i.e., assume* $a_{12} > 0$ *and* $a_{21} > 0$. *The following statements hold:*

Case I: *if* $a_{12}a_{21} < a_{11}a_{22}$, *then there exists a unique positive equilibrium point* (x_1^*, x_2^*), *solution to*

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1^* \\ x_2^* \end{bmatrix} = - \begin{bmatrix} r_1 \\ r_2 \end{bmatrix},$$

and all trajectories starting in $\mathbb{R}_{>0}^2$ *converge to it;*

Case II: *otherwise, if* $a_{12}a_{21} > a_{11}a_{22}$, *then there exists no positive equilibrium point and all trajectories starting in* $\mathbb{R}_{>0}^2$ *diverge.*
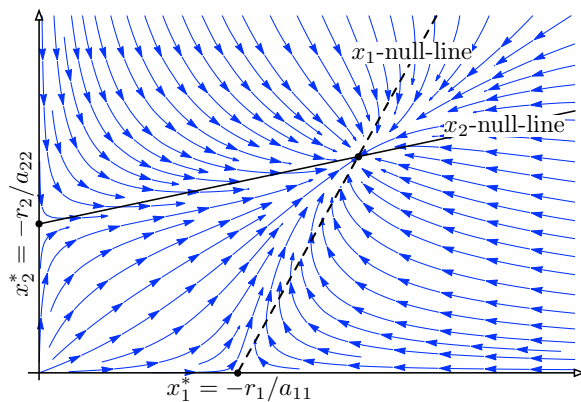
Case I: $a_{12} > 0$, $a_{21} > 0$, $a_{12}a_{21} < a_{11}a_{22}$. There exists a unique positive equilibrium point. All trajectories starting in $\mathbb{R}_{>0}^2$ converge to the equilibrium point.
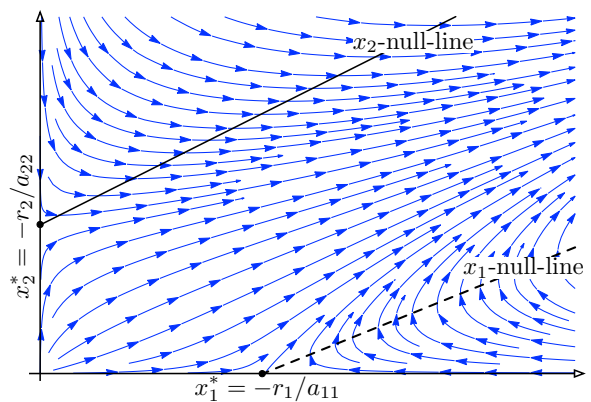
Case II: $a_{12} > 0$, $a_{21} > 0$, $a_{12}a_{21} > a_{11}a_{22}$. There exists no positive equilibrium point. All trajectories starting in $\mathbb{R}_{>0}^2$ diverge.

Figure 15.1: Two possible cases of mutualism in the two-species Lotka-Volterra system

15.1.2   Competition

Here we assume inter-species competition, that is, we assume both inter-species coefficients a_{12} and a_{21} are negative. We identify four (two sets of two) distinct parameter ranges corresponding to distinct dynamic behavior and illustrate them in Figures 15.2 and 15.3.

Lemma 15.2 (Two-species competition with a positive equilibrium). *Consider the two-species Lotka-Volterra system* (15.2) *with scalar parameters* (r_1, r_2) *and* $(a_{11}, a_{12}, a_{21}, a_{22})$. *Assume the interaction is competitive, i.e., assume* $a_{12} < 0$ *and* $a_{21} < 0$. *The following statements hold:*

Case III: *if* $r_2/|a_{22}| < r_1/|a_{12}|$ *and* $r_1/|a_{11}| < r_2/|a_{21}|$, *then there exists a unique positive equilibrium, which attracts all trajectories starting in* $\mathbb{R}^2_{>0}$;

Case IV: *if* $r_1/|a_{12}| < r_2/|a_{22}|$ *and* $r_2/|a_{21}| < r_1/|a_{11}|$, *then the equilibrium in* $\mathbb{R}^2_{>0}$ *is unstable; all trajectories (except the equilibrium solution) converge either to the equilibrium* $(-r_1/a_{11}, 0)$ *or to the equilibrium* $(0, -r_2/a_{22})$.

As for Case I, for Cases III and IV, it is easy to compute the unique positive equilibrium point (x_1^*, x_2^*) as the solution to $\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1^* \\ x_2^* \end{bmatrix} = - \begin{bmatrix} r_1 \\ r_2 \end{bmatrix}$.
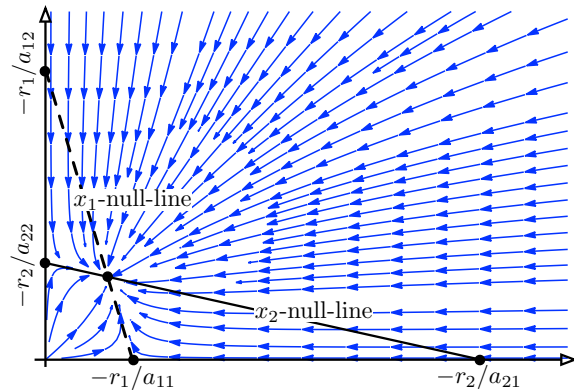
Case III: $a_{12} < 0$, $a_{21} < 0$, $r_2/|a_{22}| < r_1/|a_{12}|$, and $r_1/|a_{11}| < r_2/|a_{21}|$. There exists a unique positive equilibrium, which attracts all trajectories starting in $\mathbb{R}^2_{>0}$.
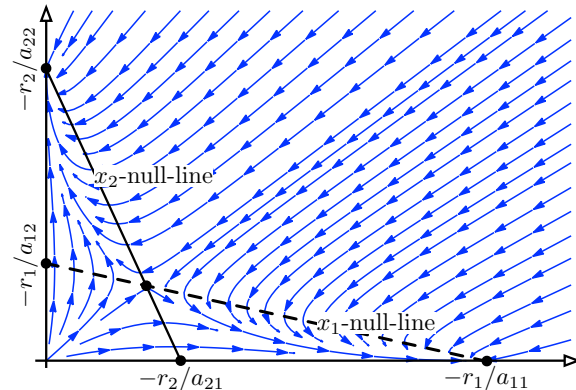
Case IV: $a_{12} < 0$, $a_{21} < 0$, $r_1/|a_{12}| < r_2/|a_{22}|$, and $r_2/|a_{21}| < r_1/|a_{11}|$. The equilibrium in $\mathbb{R}^2_{>0}$ is unstable; all trajectories (except the equilibrium solution) converge either to the equilibrium $(-r_1/a_{11}, 0)$ or to the equilibrium $(0, -r_2/a_{22})$.

Figure 15.2: Two competition cases with an equilibrium in the two-species Lotka-Volterra system

Lemma 15.3 (Two-species competition without positive equilibria). *Consider the two-species Lotka-Volterra system* (15.2) *with scalar parameters* (r_1, r_2) *and* $(a_{11}, a_{12}, a_{21}, a_{22})$. *Assume the interaction is competitive, i.e., assume* $a_{12} < 0$ *and* $a_{21} < 0$. *The following statements hold:*

Case V: *if* $r_2/|a_{22}| < r_1/|a_{12}|$ *and* $r_2/|a_{21}| < r_1/|a_{11}|$, *then there exists no equilibrium in* $\mathbb{R}^2_{>0}$ *and all trajectories starting in* $\mathbb{R}^2_{>0}$ *converge to the equilibrium* $(-r_1/a_{11}, 0)$;

Case VI: *if* $r_1/|a_{12}| < r_2/|a_{22}|$ *and* $r_1/|a_{11}| < r_2/|a_{21}|$, *then there exists no equilibrium in* $\mathbb{R}^2_{>0}$ *and all trajectories starting in* $\mathbb{R}^2_{>0}$ *converge to the equilibrium* $(0, -r_2/a_{22})$.
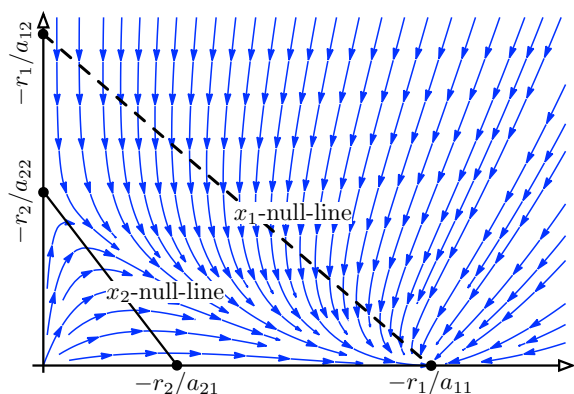
Case V: $a_{12} < 0$, $a_{21} < 0$, $r_2/|a_{22}| < r_1/|a_{12}|$, and $r_2/|a_{21}| < r_1/|a_{11}|$. There exists no equilibrium in $\mathbb{R}^2_{>0}$. All trajectories starting in $\mathbb{R}^2_{>0}$ converge to the equilibrium $(-r_1/a_{11}, 0)$.
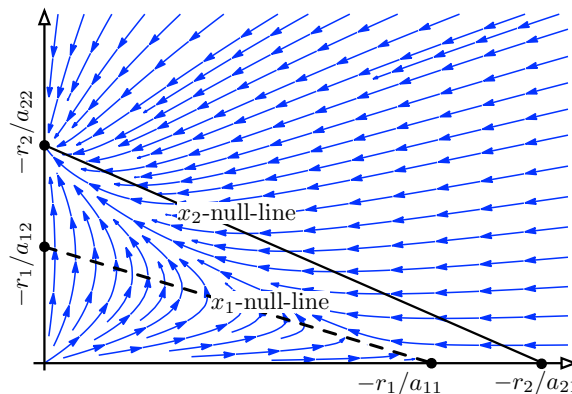
Case VI: $a_{12} < 0$, $a_{21} < 0$, $r_1/|a_{12}| < r_2/|a_{22}|$, and $r_1/|a_{11}| < r_2/|a_{21}|$. There exists no equilibrium in $\mathbb{R}^2_{>0}$. All trajectories starting in $\mathbb{R}^2_{>0}$ converge to the equilibrium $(0, -r_2/a_{22})$.

Figure 15.3: Two competition cases without equilibria in the two-species Lotka-Volterra system

15.2 General results for Lotka-Volterra models

We have seen some variety of behavior in the 2-species Lotka-Volterra model (15.2). Much richer dynamical behavior is possible in the n-species Lotka-Volterra model (13.3), including persistence, extinction, equilibria, periodic orbits, and chaotic evolution. In what follows we focus on sufficient conditions for the existence and stability of equilibrium points.

Lemma 15.4 (Lotka-Volterra is a positive system). *For* $n \geq 2$, *the Lotka-Volterra system* (15.1) *is a positive system, i.e.,* $x(0) \geq 0$ *implies* $x(t) \geq 0$ *for all subsequent* t. *Moverover, if* $x_i(0) = 0$, *then* $x_i(t) = 0$ *for all subsequent* t.

Therefore, without loss of generality, we can assume that all initial conditions are positive vectors in $\mathbb{R}^n_{>0}$. In other words, if a locally-asymptotically stable positive equilibrium exists, the best we can hope for is to establish that its region of attraction is $\mathbb{R}^n_{>0}$. We are now ready to state the main result of this section.

Theorem 15.5 (Sufficient conditions for global asymptotic stability). *For the Lotka-Volterra system* (15.1) *with interaction matrix* A *and intrinsic growth rate* r, *assume*

(A1) *A is diagonally stable, i.e., there exists a positive vector* $p \in \mathbb{R}^n$ *such that* $\mathrm{diag}(p)A + A^\mathsf{T}\mathrm{diag}(p)$ *is negative definite, and*

(A2) *the unique equilibrium point* $x^* = -A^{-1}r$ *is positive.*

Then x^* is globally asymptotically stable on $\mathbb{R}^n_{>0}$.

Proof. Note that A diagonally stable implies A Hurwitz and invertible. For $\kappa > 0$, recall the *logarithmic-linear function* $V_{\text{log-lin},\kappa} : \mathbb{R}_{>0} \to \mathbb{R}$ illustrated in Figure 14.6 and defined by

$$V_{\text{log-lin},\kappa}(x) = x - \kappa - \kappa \log\left(\frac{x}{\kappa}\right).$$

Assumption (A2) allows us to define $V : \mathbb{R}^n_{>0} \to \mathbb{R}_{\geq 0}$ by

$$V(x) = \sum_{i=1}^{n} p_i V_{\text{log-lin},x_i^*}(x_i) = \sum_{i=1}^{n} p_i\big(x_i - x_i^* - x_i^* \log(x_i/x_i^*)\big).$$

From Exercise E14.1 we know that the function $V_{\text{log-lin},\kappa}$ is continuously differentiable, takes non-negative values and satisfies $V_{\text{log-lin},\kappa}(x_i) = 0$ if and only if $x_i = \kappa$. Moreover, this function is unbounded in the limits as $x_i \to \infty$ and $x_i \to 0^+$. Therefore, V is globally positive-definite about x^* and proper.

Next, we compute the Lie derivative of V along the flow of the Lotka-Volterra vector field $f_{\text{LV}}(x) = \text{diag}(x)(Ax + r)$. First, compute $\frac{d}{dx_i} V_{\text{log-lin},x_i^*}(x_i)(x_i) = (x_i - x_i^*)/x_i$, so that

$$\mathscr{L}_{f_{\text{LV}}} V(x) = \sum_{i=1}^{n} p_i \frac{x_i - x_i^*}{x_i}(f_{\text{LV}}(x))_i.$$

Because A is invertible and $x^* = -A^{-1}r$, we write $Ax + r = A(x - x^*)$ and obtain

$$\begin{aligned}
\mathscr{L}_{f_{\text{LV}}} V(x) &= \sum_{i=1}^{n} p_i(x_i - x_i^*)(A(x - x^*))_i \\
&= (x - x^*)^{\mathsf{T}} A^{\mathsf{T}} \text{diag}(p)(x - x^*) \\
&= \frac{1}{2}(x - x^*)^{\mathsf{T}}(A^{\mathsf{T}} \text{diag}(p) + \text{diag}(p)A)(x - x^*).
\end{aligned}$$

where we use the equality $y^{\mathsf{T}} By = y^{\mathsf{T}}(B + B^{\mathsf{T}})y/2$ for all $y \in \mathbb{R}^n$ and $B \in \mathbb{R}^{n \times n}$. Assumption (A1) now implies that $\mathscr{L}_{f_{\text{LV}}} V(x) \leq 0$ with equality if and only if $x = x^*$. Therefore, $\mathscr{L}_{f_{\text{LV}}} V$ is globally negative-definite about x^*. According to the Lyapunov Stability Criteria in Theorem 14.4, x^* is globally asymptotically stable on $\mathbb{R}^n_{>0}$. ∎

Note: Assumption (A2) is not critical and, via a more complex treatment, a more general theorem can be obtained. Under the diagonal stability Assumption (A1), (Takeuchi, 1996, Theorem 3.2.1) shows the existence of a unique non-negative and globally stable equilibrium point x^* for each $r \in \mathbb{R}^n$; the existence and uniqueness of x^* is established via a linear complementarity problem.

15.3 Cooperative Lotka-Volterra models

In this section we focus on the case of Lotka-Volterra systems with only mutualistic interactions. In other words, we consider systems whose interaction terms satisfy $a_{ij} \geq 0$ for all i and j. For such systems, whenever $i \neq j$ we know

$$\frac{\partial}{\partial x_j}(f_{\text{LV}})_i(x) = a_{ij}x_j \geq 0,$$

so that the Jacobian matrix of such systems is Metzler everywhere in $\mathbb{R}_{\geq 0}$. Such systems are called *cooperative*.

We recall from Section 9.2 the properties of Metzler matrices. For example the Perron-Frobenius Theorem 9.4 for Metzler matrices establishes the existence of a dominant eigenvalue. Metzler matrices have so much structure that we are able to provide the following fairly comprehensive characterization: (1) Metzler matrices with a positive dominant eigenvalue have unbounded solutions of the Lotka-Volterra model (see Lemma 15.6 below), and (2) Metzler matrices with a negative dominant eigenvalue (and positive intrinsic growth parameter) have a globally asymptotically-stable equilibrium point (see Theorem 15.7 below).

We start with a sufficient condition for unbounded evolutions.

Lemma 15.6 (Unbounded evolutions for unstable Metzler matrices). *Consider the Lotka-Volterra system* (15.1) *with interaction matrix A and intrinsic growth rate r. If A is an irreducible Metzler matrix with a positive dominant eigenvalue, then*

(i) *there exist unbounded solutions starting from $\mathbb{R}_{>0}$, and*

(ii) *if $r > 0$, then all solutions starting from $\mathbb{R}_{>0}$ are unbounded.*

Proof. Let $\lambda > 0$ and $w > 0$ with $\mathbb{1}_n^\mathsf{T} w = 1$ be the dominant eigenvalue and left eigenvector of A, whose existence and properties are established by the Perron-Frobenius Theorem 9.4 for Metzler matrices. Define $V : \mathbb{R}_{>0}^n \to \mathbb{R}_{>0}$ as the following weighted geometric average:

$$V(x) = \Pi_{i=1}^n x_i^{w_i}.$$

Along the flow of the Lotka-Volterra system, simple calculations show

$$\frac{\partial V(x)}{\partial x_i} = w_i \frac{1}{x_i} V(x) \quad \implies \quad \frac{\mathscr{L}_{f_{\mathrm{LV}}} V(x)}{V(x)} = \sum_{i=1}^n w_i \frac{1}{x_i} (f_{\mathrm{LV}}(x))_i = w^\mathsf{T}(Ax + r) = w^\mathsf{T}(\lambda x + r).$$

Generalizing the classic inequality $(a+b)/2 \geq (ab)^{1/2}$ for any $a, b \in \mathbb{R}_{>0}$, we recall from (Lohwater, 1982) the *weighted arithmetic-geometric mean inequality*: $w^\mathsf{T} x \geq \Pi_{i=1}^n x_i^{w_i}$ for any $x, w \in \mathbb{R}_{>0}^n$. Therefore, we have

$$\frac{\mathscr{L}_{f_{\mathrm{LV}}} V(x)}{V(x)} = w^\mathsf{T}(\lambda x + r) \geq \lambda \Pi_{i=1}^n x_i^{w_i} + w^\mathsf{T} r = \lambda V(x) + w^\mathsf{T} r,$$

so that

$$\mathscr{L}_{f_{\mathrm{LV}}} V(x) \geq V(x)(\lambda V(x) + w^\mathsf{T} r).$$

This inequality implies that, for any $x(0)$ such that $V(x(0)) > -w^\mathsf{T} r/\lambda$, the function $t \mapsto V(x(t))$, and therefore the state $x(t)$, goes to infinity in finite time. This concludes the proof of statement (i).

Statement (ii) follows by noting that $r > 0$ implies $V(x(0)) > -w^\mathsf{T} r/\lambda$ for all $x(0) \in \mathbb{R}_{>0}^n$. ∎

Note: this lemma is true for any interaction matrix A that has a positive left eigenvector with positive eigenvalue.

We next provide a sufficient condition for global convergence to a unique equation* point.

Theorem 15.7 (Global convergence for cooperative Lotka-Volterra). *For the Lotka-Volterra system* (15.1) *with interaction matrix A and intrinsic growth rate r, assume*

(A3) *the interaction matrix A is Metzler and Hurwitz, and*

(A4) *the intrinsic growth parameter r is positive.*

Then there exists a unique interior equilibrium point x^* and x^* is globally attractive on $\mathbb{R}^n_{>0}$.

Proof. We leave it to the reader to verify that, based on Assumptions (A3) and (A4), the Assumptions (A1) and (A2) of Theorem 15.5 are satisfied so that its consequences hold. ∎

Note: In (Baigent, 2010, Chapter 4), Theorem 15.7 is established via the Lyapunov function $V(x) = \max_{i \in \{1,\dots,n\}} \frac{|x_i - x_i^*|}{\xi_i}$, where x^* is the equilibrium point and $\mathrm{diag}(\xi_1, \dots, \xi_n)$ is the diagonal Lyapunov matrix (as in Theorem 14.17(vi)) for the Metzler Hurwitz matrix A.

15.4 Historical notes and further reading

The Lotka-Volterra population models are one the simplest and most widely adopted frameworks for modeling the dynamics of interacting populations in mathematical ecology. These equations were originally developed in (Lotka, 1920; Volterra, 1928).

An early reference for the analysis of the 2-species model is (Goh, 1976). Early references for the key stability result in Theorem 15.5 are (Takeuchi et al., 1978; Goh, 1979).

Textbook treatment include (Goh, 1980; Takeuchi, 1996; Baigent, 2010). For a more complete treatment of the n-special model, we refer the interested reader to (Takeuchi, 1996; Baigent, 2010). For example, Baigent (2010) discusses conservative Lotka-Volterra models (Hamiltonian structure and existence of periodic orbits), competitive and monotone models.

We refer to the texts (Hofbauer and Sigmund, 1998; Sandholm, 2010) for comprehensive discussions about the connection with between Lotka-Volterra models and evolutionary game dynamics.

15.5   Exercises

E15.1 **Proofs for 2-species behavior.** Provide proofs for Lemmas 15.1, 15.2, and 15.3.

E15.2 **The 2-dimensional Lotka-Volterra predator/prey dynamics.** In this exercise we study a 2-dimensional predator/prey model. We specialize the general Lotka-Volterra population model to the following set of equations:

$$\begin{aligned}
\dot{x}(t) &= \alpha x(t) - \beta x(t) y(t), \\
\dot{y}(t) &= -\gamma y(t) + \delta x(t) y(t),
\end{aligned} \tag{E15.1}$$

where x is the non-negative number of preys, y is the non-negative number of predators individuals, and α, β, and γ are fixed positive systems parameters.

  (i) Compute the unique non-zero equilibrium point (x^*, y^*) of the system.
 (ii) Determine, if possible, the stability properties of the equilibrium points $(0, 0)$ and (x^*, y^*) via linearization (Theorem 14.10).
(iii) Define the function $V(x, y) = -\delta x - \beta y + \gamma \ln(x) + \alpha \ln(y)$ and note its level sets as illustrated in Figure (E15.1).

      a) Compute the Lie derivative of $V(x, y)$ with respect to the Lotka-Volterra vector field.
      b) What can you say about the stability properties of (x^*, y^*)?
      c) Sketch the trajectories of the system for some initial conditions in the x-y positive orthant.
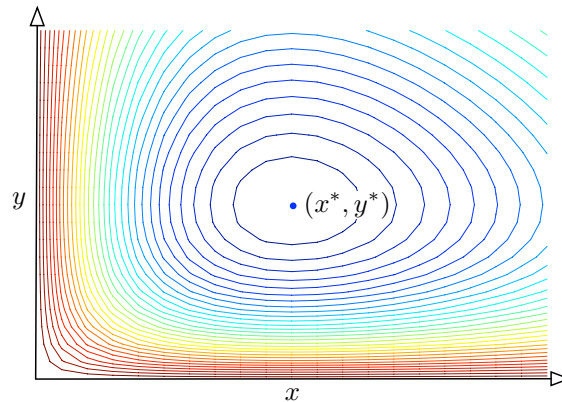
Figure E15.1: Level sets of the function $V(x, y)$ for unit parameter values

Virus Propagation in Contact Networks

In this chapter we continue our discussion about the diffusion and propagation of infectious diseases. Starting from the scalar lumped models discussed in Section 13.2, we presents deterministic nonlinear models over strongly-connected contact networks. We consider network models for susceptible-infected (SI), susceptible-infected-susceptible (SIS), and susceptible-infected-recovered (SIR) settings. In each setting, we provide a comprehensive nonlinear analysis of equilibria, stability properties, convergence, monotonicity, positivity, and threshold conditions; in all three cases the network results are appropriate generalizations of the respective scalar models.

As in previous chapters, for an irreducible nonnegative matrix A, we let $\lambda_{\max} = \rho(A)$, $w_{\max}$, and $v_{\max}$ denote the dominant eigenvalue of A and the corresponding positive left and right eigenvectors associated with $\lambda_{\max}$, normalized to satisfy $w_{\max}^\mathsf{T} v_{\max} = 1$.

16.1 Susceptible-Infected Model

Recall that the scalar SI model is discussed in Section 13.2, including model (13.4), solution (13.5), and sample evolution in Figure 13.4. Given an adjacency matrix $A \in \mathbb{R}_{\geq 0}^{n \times n}$, the network SI model, illustrated in Figure 16.1, is given by

$$\dot{x}_i(t) = \beta\big(1 - x_i(t)\big) \sum_{j=1}^n a_{ij} x_j(t), \tag{16.1}$$

or, in equivalent vector form,

$$\dot{x}(t) = \beta\big(I_n - \operatorname{diag}\big(x(t)\big)\big) A x(t), \tag{16.2}$$

where $\beta > 0$ is the infection rate. Alternatively, in terms of the fractions of susceptibile individuals $s(t) = \mathbb{1}_n - x(t)$, the network SI model is

$$\dot{s}(t) = -\beta \operatorname{diag}(s(t)) A(\mathbb{1}_n - s(t)). \tag{16.3}$$

Theorem 16.1 (Dynamical behavior of network SI model). *Consider the network SI model* (16.2) *with* $\beta > 0$. *For strongly connected graph with adjacency matrix A, the following statements hold:*
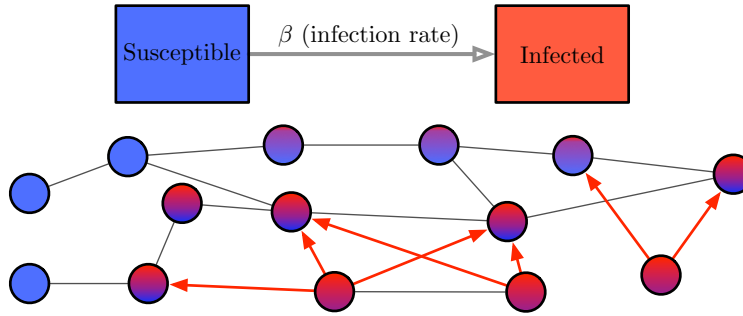
Figure 16.1: In the network SI model, each node is described by a fraction of infected individuals taking value between 0 (blue) and 1 (red). Individuals become increasingly infected with infection rate β.

(i) if $x(0), s(0) \in [0,1]^n$, then $x(t), s(t) \in [0,1]^n$ for all $t > 0$. Moreover, $x(t)$ is monotonically non-decreasing (here by monotonically non-decreasing we mean $x(t_1) \le x(t_2)$ for all $t_1 \le t_2$). Finally, if $x(0) \ge \mathbb{0}_n$ and $x(0) \ne \mathbb{0}_n$, then $x(t) > \mathbb{0}_n$ for all $t > 0$;

(ii) the model (16.2) has two equilibrium points: $\mathbb{0}_n$ (no epidemic), and $\mathbb{1}_n$ (full contagion);

   a) the linearization of model (16.2) about the equilibrium point $\mathbb{0}_n$ is $\dot{x} = \beta A x$ and it is exponentially unstable;

   b) let $D = \operatorname{diag}(A\mathbb{1}_n)$ be the out-degree matrix. The linearization of model (16.3) about the equilibrium $\mathbb{0}_n$ is $\dot{s} = -\beta D s$ and it is exponentially stable;

(iii) each trajectory with initial condition $x(0) \ne \mathbb{0}_n$ converges asymptotically to $\mathbb{1}_n$, that is, the epidemic spreads monotonically to the entire network.

Proof. Regarding statement (i), the fact that, if $x(0), s(0) \in [0,1]^n$, then $x(t), s(t) \in [0,1]^n$ for all $t > 0$ means that $[0,1]^n$ is an invariant set for the differential equation (16.2). This is the consequence of Nagumo's Theorem (see (Blanchini and Miani, 2015, Theorem 4.7)), since for any x belonging on the boundary of the set $[0,1]^n$, the vector $\beta(I_n - \operatorname{diag}(x))Ax$ is either tangent or points inside the set $[0,1]^n$.

Observe that the invariance of the set $[0,1]^n$ implies that $\dot{x}(t) \ge \mathbb{0}_n$ and so $x(t_1) \le x(t_2)$ for all $t_1 \le t_2$. We now want to prove that, if $x(0) \ge \mathbb{0}_n$ and $x(0) \ne \mathbb{0}_n$, then $x(t) > \mathbb{0}_n$ for all $t > 0$. If, by contradiction, there is $i \in \{1, \dots, n\}$ and $T > 0$ such that $x_i(T) = 0$, then the monotonicity of $x_i(t) = 0$ implies that $x_i(t) = 0$ for all $t \in [0,T]$, which yields $\dot{x}_i(t) = 0$ for all $t \in [0,T]$. By equation (16.1) this implies that $x_j(t) = 0$ for all $t \in [0,T]$ for all j such that $a_{ij} > 0$. We iterate this argument and using the irreducibility of A we get the contradiction that $x(t) = 0$ for all $t \in [0,T]$. This concludes the proof of (i).

Regarding statement (ii), note that $\mathbb{0}_n$ and $\mathbb{1}_n$ are clearly equilibrium points. Let $\bar{x} \in [0,1]^n$ be an equilibrium and assume that $\bar{x} \ne \mathbb{1}_n$. Then there is i such that $\bar{x}_i \ne 1$. Since $\beta(1 - \bar{x}_i) \sum_{j=1}^n a_{ij} \bar{x}_j = 0$, then $\sum_{j=1}^n a_{ij} \bar{x}_j = 0$ which implies that $\bar{x}_j = 0$ for all j such that $a_{ij} > 0$. By iterating this argument and using the irreducibility of A we get that $\bar{x} = 0$ concluding only $\mathbb{0}_n$ and $\mathbb{1}_n$ are equilibrium points. Statements (ii)a and (ii)b are obvious. Exponential stability of the linearization $\dot{s} = -\beta D s$ is obvious, and the Perron-Frobenius Theorem implies the existence of the unstable positive eigenvalue $\rho(A) > 0$ for the linearization $\dot{x} = \beta A x$.

Regarding statement (iii), consider the function $V(x) = \mathbb{1}_n^\mathsf{T}(\mathbb{1}_n - x)$; this is a smooth function defined over the compact and forward invariant set $[0,1]^n$ (see statement (i)). Since $\dot{V} = -\beta\mathbb{1}_n^\mathsf{T}(I_n - \mathrm{diag}(x))Ax$, we know that $\dot{V} \leq 0$ for all x and $\dot{V}(x) = 0$ if and only if $x \in \{\mathbb{0}_n, \mathbb{1}_n\}$. The Krasovskiĭ-LaSalle Invariance Principle implies that all trajectories with $x(0)$ converge asymptotically to either $\mathbb{1}_n$ or $\mathbb{0}_n$. Additionally, note that $0 \leq V(x) \leq n$ for all $x \in [0,1]^n$, that $V(x) = 0$ if and only if $x = \mathbb{1}_n$ and that $V(x) = n$ if and only if $x = \mathbb{0}_n$. Therefore, all trajectories with $x(0) \neq \mathbb{0}_n$ converge asymptotically to $\mathbb{1}_n$. ∎

For the adjacency matrix A, there exists a non-singular matrix T such that $A = TJT^{-1}$, where J is the Jordan normal form of A. Since A is non-negative and irreducible, the first Jordan block $J_1 = (\lambda_{\max})_{1\times 1}$ and $\lambda_{\max} > \Re(\lambda_i)$ for any other eigenvalue λ_i of A. Consider now the onset of an epidemic in a large population characterized by a small initial infection $x(0) = x_0$ much smaller than $\mathbb{1}_n$. The system evolution is approximated by $\dot{x} = \beta Ax$. This "initial-times" linear evolution satisfies

$$x(t) = e^{\beta At} x(0) = T e^{\beta Jt} T^{-1} x(0)$$
$$= e^{\beta\lambda_{\max}t} \left(T\mathbb{e}_1\mathbb{e}_1^\mathsf{T} T^{-1} x(0) + o(1) \right),$$

where $\mathbb{e}_1$ is the first standard basis vector in $\mathbb{R}^n$ and $o(1)$ denotes a time-varying vector that vanishes as $t \to +\infty$. Let u_1 denote the first column of T and let v_1^T denote the first row of T^{-1}. Since $AT = TJ$ and $T^{-1}A = JT^{-1}$, one can check that u_1 (v_1 resp.) is the right (left resp.) eigenvector of A associated with the eigenvalue $\lambda_{\max}$. Since $T^{-1}T = I_n$, we have $v_1^\mathsf{T} u_1 = 1$. Therefore,

$$x(t) = e^{\beta\lambda_{\max}t} \left(u_1 v_1^\mathsf{T} x(0) + o(1) \right)$$
$$= e^{\beta\lambda_{\max}t} \left(\frac{w_{\max}^\mathsf{T} x(0)}{w_{\max}^\mathsf{T} v_{\max}} v_{\max} + o(1) \right). \tag{16.4}$$

That is, the epidemic initially experiences exponential growth with rate $\beta\lambda_{\max}$ and with distribution among the nodes given by the eigenvector $v_{\max}$.

Now suppose that at some time T, for all i we have that $x_i(T) = 1 - \varepsilon_i$, where each ε_i is much smaller than 1. Then, for time $t > T$, the approximated system for $s(t)$ is given by:

$$\dot{s}_i(t) = -\beta d_i s_i(t) \quad \implies \quad s_i(t) = \varepsilon_i\, e^{-\beta d_i(t-T)}.$$

From the discussion above, we conclude that the initial infection rate is proportional to the eigenvector centrality, and the final infection speed is proportional to the degree centrality.

16.2 Susceptible-Infected-Susceptible model

Recall that the scalar SIS model is discussed in Section 13.2, including model (13.6), solution (13.7), and sample evolution in Figure 13.5. The network SIS model with infection rate β and recovery rate γ is given by:

$$\dot{x}_i(t) = \beta(1 - x_i(t)) \sum_{j=1}^{n} a_{ij} x_j(t) - \gamma x_i(t), \tag{16.5}$$

or, in equivalent vector form,

$$\dot{x}(t) = \beta\left(I_n - \mathrm{diag}(x(t))\right) Ax(t) - \gamma x(t). \tag{16.6}$$

In the rest of this section we study the dynamical properties of this model. We start by defining the monotonically-increasing functions

$$f_+(y) = y/(1+y), \quad \text{and} \quad f_-(z) = z/(1-z),$$

for $y \in \mathbb{R}_{\geq 0}$ and $z \in [0, 1[$. Note that $f_+(f_-(z)) = z$ for all $z \in [0, 1)$. For vector variables $y \in \mathbb{R}_{\geq 0}^n$ and $z \in [0, 1)^n$, we write $F_+(y) = (f_+(y_1), \ldots, f_+(y_n))$, and $F_-(z) = (f_-(z_1), \ldots, f_-(z_n))$.

We first characterize the behavior of the network SIS model in a regime we describe as "below the threshold."

Theorem 16.2 (Dynamical behavior of the network SIS model: Below the threshold). *Consider the network SIS model* (16.5), *with $\beta > 0$ and $\gamma > 0$, over a strongly connected digraph with adjacency matrix A. Let $\lambda_{\max}$ and $w_{\max}$ be the dominant eigenvalue of A and the corresponding normalized left eigenvector respectively. If $\beta\lambda_{\max}/\gamma < 1$, then*

   (i) *if $x(0), s(0) \in [0, 1]^n$, then $x(t), s(t) \in [0, 1]^n$ for all $t > 0$. Moreover, if $x(0) \geq \mathbb{0}_n$ and $x(0) \neq \mathbb{0}_n$, then $x(t) > \mathbb{0}_n$ for all $t > 0$;*

  (ii) *there exists a unique equilibrium point $\mathbb{0}_n$, the linearization of (16.5) about $\mathbb{0}_n$ is $\dot{x} = (\beta A - \gamma I_n)x$ and it is exponentially stable;*

 (iii) *from any $x(0) \neq \mathbb{0}_n$, the weighted average $t \mapsto w_{\max}^\mathsf{T} x(t)$ is monotonically and exponentially decreasing, and all the trajectories converge to $\mathbb{0}_n$.*

Proof. Regarding statement (i), as in Theorem 16.1 the first part is the consequence of Nagumo's Theorem. Then define $y(t) := e^{\gamma t} x(t)$. Notice that this variable satisfies the differential equation $\dot{y}(t) = \beta \operatorname{diag}(s(t)) A y(t)$. From the same arguments used in the proof of the statement (i) in Theorem 16.1 we argue that $y(t) > \mathbb{0}_n$ for all $t > 0$. From this it follows that also $x(t) > \mathbb{0}_n$ for all $t > 0$.

Regarding statement (ii), assume that x^* is an equilibrium point. It is easy to se that $x^* < \mathbb{1}_n$. Observe moreover that x^* is an equilibrium point if and only if $\hat{A}x^* = F_-(x^*)$ or, equivalently, if and only if $F_+\left(\frac{\beta}{\gamma} A x^*\right) = x^*$. This means that x^* is an equilibrium if and only if it is a fixed point of $\mathcal{F}$, where $\mathcal{F}(x) := F_+\left(\frac{\beta}{\gamma} A x\right)$. Let $\hat{A} = \beta A/\gamma$. For $x \in [0, 1]^n$, note $F_+(\hat{A}x) \leq \hat{A}x$ because $f_+(z) \leq z$. Moreover, $\mathbb{0}_n \leq x \leq y$ implies that $\mathbb{0}_n \leq \mathcal{F}(x) \leq \hat{A}y$. Therefore, if $\mathbb{0}_n \leq x$, then $\mathcal{F}^k(x) \leq \hat{A}^k x$, for all k. Since $\hat{A}$ is Schur stable, then $\lim_{k\to\infty} \mathcal{F}^k(x) = 0$. This shows that the only fixed point of $\mathcal{F}$ is zero.

Next, the linearization of equation (16.6) is verified by dropping the second-order terms. The linearized system is exponentially stable at $\mathbb{0}_n$ for $\beta\lambda_{\max} - \gamma < 0$ because $\lambda_{\max}$ is larger, in real part, than any other eigenvalue of A by the Perron-Frobenius Theorem for irreducible matrices.

Finally, regarding statement (iii), define $y(t) = w_{\max}^\mathsf{T} x(t)$ and note that $\left(I_n - \operatorname{diag}(z)\right) w_{\max} \leq w_{\max}$ for any $z \in [0, 1]^n$. Therefore,

$$\dot{y}(t) \leq \beta w_{\max}^\mathsf{T} A x(t) - \gamma w_{\max}^\mathsf{T} x(t) = (\beta\lambda_{\max} - \gamma)y(t) < 0.$$

By the Grönwall-Bellman Comparison Lemma in Exercise E14.2, $y(t)$ is monotonically decreasing and satisfies $y(t) \leq y(0) \exp\left((\beta\lambda_{\max} - \gamma)t\right)$ from all initial conditions $y(0)$. This concludes our proof of statement (iii). ∎

We next present the dynamical behavior of the network SIS model "above the threshold" as follows.

Theorem 16.3 (Dynamical behavior of the network SIS model: Above the threshold). *Consider the network SIS model (16.5), with $\beta > 0$ and $\gamma > 0$, over a strongly connected digraph with adjacency matrix A. Let $\lambda_{\max}$ be the dominant eigenvalue of A and let $w_{\max}$ and $v_{\max}$ be the corresponding normalized left and right eigenvectors respectively. Let $d = A\mathbb{1}_n$. If $\beta\lambda_{\max}/\gamma > 1$, then*

(i) *if $x(0), s(0) \in [0,1]^n$, then $x(t), s(t) \in [0,1]^n$ for all $t > 0$. Moreover, if $x(0) \geq \mathbb{0}_n$ and $x(0) \neq \mathbb{0}_n$, then $x(t) > \mathbb{0}_n$ for all $t > 0$;*

(ii) *$\mathbb{0}_n$ is an equilibrium point, the linearization of system (16.6) at $\mathbb{0}_n$ is unstable due to the unstable eigenvalue $\beta\lambda_{\max} - \gamma$ (i.e., there will be an epidemic outbreak);*

(iii) *besides the equilibrium $\mathbb{0}_n$, there exists a unique equilibrium point x^*, called the* endemic state, *such that*

    a) $x^* > \mathbb{0}_n$,

    b) $x^* = \delta a v_{\max} + \mathcal{O}(\delta^2)$ as $\delta \to 0^+$, where $\delta := \beta\lambda_{\max}/\gamma - 1$ and

$$a = \frac{w_{\max}^T v_{\max}}{w_{\max}^T \operatorname{diag}(v_{\max}) v_{\max}},$$

    c) $x^* = \mathbb{1}_n - (\gamma/\beta)\operatorname{diag}(d)^{-1}\mathbb{1}_n + \mathcal{O}(\gamma^2/\beta^2)$, at fixed A, as $\gamma/\beta \to 0^+$,

    d) define a sequence $\{y(k)\}_{k\in\mathbb{N}} \subset \mathbb{R}^n$ by

$$y(k+1) := F_+\left(\frac{\beta}{\gamma}Ay(k)\right). \tag{16.7}$$

If $y(0) \geq 0$ is a scalar multiple of $v_{\max}$ and satisfies either $0 < \max_i y_i(0) \leq 1 - \gamma/(\beta\lambda_{\max})$ or $\min_i y_i(0) \geq 1 - \gamma/(\beta\lambda_{\max})$, then

$$\lim_{k\to\infty} y(k) = x^*.$$

Moreover, if $\max_i y_i(0) \leq 1 - \gamma/(\beta\lambda_{\max})$, then $y(k)$ is monotonically non-decreasing; if $\min_i y_i(0) \geq 1 - \gamma/(\beta\lambda_{\max})$, then $y(k)$ is monotonically non-increasing.

(iv) *the endemic state x^* is locally exponentially stable and its domain of attraction is $[0,1]^n \setminus \mathbb{0}_n$.*

Note: statement (ii) means that, near the onset of an epidemic outbreak, the exponential growth rate is $\beta\lambda_{\max} - \gamma$ and the outbreak tends to align with the dominant eigenvector $v_{\max}$; for more details see the discussion leading up to the approximate evolution (16.4).

Proof of selected statements in Theorem 16.3. Statement (i) can be proved as done for statement (i) of Theorem 16.1.

Statement (ii) follows from the same analysis of the linearized system as in the proof of Theorem 16.2(ii).

Regarding statement (iii), we begin by establishing two properties of the map $x \mapsto F_+(\hat{A}x)$, for $\hat{A} = \beta A/\gamma$. First, we claim that, $y > z \geq \mathbb{0}_n$ implies $F_+(\hat{A}y) > F_+(\hat{A}z)$. Indeed, note that G being connected implies that the adjacency matrix A has at least one strictly positive entry in each row. Hence,

$y - z > \mathbb{0}_n$ implies $\hat{A}(y - z) > \mathbb{0}_n$ and, since f_+ is monotonically increasing, $\hat{A}y > \hat{A}z$ implies $F_+(\hat{A}y) > F_+(\hat{A}z)$.

Second, we observe that, for any $0 < \alpha < 1$ and $z > 0$, we have $f_+(\alpha z) \geq z$ if and only if $z \leq 1 - 1/\alpha$. Suppose $y(0)$ is a scalar multiple of $v_{\max}$ and $0 < \max_i y_i(0) \leq 1 - \gamma/(\beta\lambda_{\max})$. We have

$$F_+(\hat{A}y(0))_i = f_+\Big(\frac{\beta\lambda_{\max}}{\gamma}y_i(0)\Big) \geq y_i(0).$$

Therefore, the sequence $\{y(k)\}_{k\in\mathbb{N}}$ defined by equation (16.7) satisfies $y(1) \geq y(0)$, which in turn leads to $y(2) = F_+(\hat{A}y(1)) \geq F_+(\hat{A}y(0)) = y(1)$, and by induction, $y(k+1) = F_+(\hat{A}y(k)) \geq y(k)$ for any $k \in \mathbb{N}$. Such sequence $\{y(t)\}$ is monotonically non-decreasing and entry-wise upper bounded by $\mathbb{1}_n$. Therefore, as k diverges, $y(k)$ converges to some $x^* > \mathbb{0}_n$ such that $F_+(\hat{A}x^*) = x^*$. This proves the existence of an equilibrium $x^* = \lim_{k\to\infty} y(k) > \mathbb{0}_n$ as claimed in statements (iii)a and (iii)d.

Similarly, for any $0 < \alpha < 1$ and $z > 0$, $f_+(\alpha z) \leq z$ if and only if $z \geq 1 - 1/\alpha$. Following the same line of argument in the previous paragraph, one can check that the $\{y(k)\}_{k\in\mathbb{N}}$ defined by equation (16.7) is monotonically non-increasing and converges to some x^*, if $y(0)$ is a scalar multiple of $v_{\max}$ and satisfies $\min_i y_i(0) \geq 1 - \gamma/(\beta\lambda_{\max})$.

We now establish the uniqueness of the equilibrium $x^* \in [0,1]^n \setminus \{\mathbb{0}_n\}$. First, we claim that an equilibrium point with an entry equal to 0 must be $\mathbb{0}_n$. Indeed, assume y^* is an equilibrium point and assume $y_i^* = 0$ for some $i \in \{1, \ldots, n\}$. The equality $y_i^* = f_+(\sum_{j=1}^n a_{ij}y_j^*)$ implies that also any node j with $a_{ij} > 0$ must satisfy $y_j^* = 0$. Because G is connected, all entries of y^* must be zero. Second, by contradiction, we assume there exists another equilibrium point $y^* > \mathbb{0}_n$ distinct from x^*. Let $\alpha := \min_j\{y_j^*/x_j^*\}$ and let i satisfy $\alpha = y_i^*/x_i^*$. Then $y^* \geq \alpha x^* > \mathbb{0}_n$ and $y_i^* = \alpha x_i^*$. Notice that we can assume with no loss of generality that $\alpha < 1$, otherwise we exchange x^* and y^*. Observe now that

$$
\begin{aligned}
\big(F_+(\hat{A}y^*) - y^*\big)_i &= f_+\big((\hat{A}y^*)_i\big) - \alpha x_i^* \\
&\geq f_+\big(\alpha(\hat{A}x^*)_i\big) - \alpha x_i^* &&(\hat{A} \geq \mathbb{0}_{n\times n}) \\
&> \alpha f_+\big((\hat{A}x^*)_i\big) - \alpha x_i^* &&(0 < \alpha < 1 \text{ and } z > 0) \\
&= \alpha\big(F_+(\hat{A}x^*) - x^*\big)_i = 0. &&(x^* \text{ is an equilibrium})
\end{aligned}
$$

Therefore, $\big(F_+(\hat{A}y^*) - y^*\big)_i > 0$, which contradicts the fact that y^* is an equilibrium.

Now we prove statement (iii)b. Observe first that, if we take

$$y(0) = \Big(1 - \frac{\gamma}{\beta\lambda_{\max}}\Big)\frac{v_{\max}}{\max_i\{v_{\max,i}\}} = \frac{\delta}{\delta+1}\frac{v_{\max}}{\max_i\{v_{\max,i}\}},$$

then $y(k)$ is monotonically non-decreasing and converges to x^*, and if we take instead

$$y(0) = \Big(1 - \frac{\gamma}{\beta\lambda_{\max}}\Big)\frac{v_{\max}}{\min_i\{v_{\max,i}\}} = \frac{\delta}{\delta+1}\frac{v_{\max}}{\min_i\{v_{\max,i}\}},$$

then $y(k)$ is monotonically non-increasing and converges to x^*. These two statements together imply argue that

$$\frac{\delta}{\delta+1}\frac{v_{\max}}{\max_i\{v_{\max,i}\}} \leq x^* \leq \frac{\delta}{\delta+1}\frac{v_{\max}}{\min_i\{v_{\max,i}\}},$$

and, in turn, that x^* is infinitesimal as a function of δ. Consider now the Taylor expansion $x^*(\delta) = x_1\delta + x_2\delta^2 + \mathcal{O}(\delta^3)$. Since the equilibrium x^* satisfies the equation

$$(\delta + 1)\big(I_n - \operatorname{diag}(x^*)\big)Ax^* - \lambda_{\max}x^* = 0,$$

by substituting the expansion and equating to zero the coefficient of the term δ, we obtain the equation $Ax_1 - \lambda_{\max}x_1 = 0$. This proves that x_1 is a multiple of $v_{\max}$, namely $x_1 = av_{\max}$ for some constant a. By equating to zero the coefficient of the term δ^2, we obtain instead the equation

$$Ax_1 + Ax_2 - \operatorname{diag}(x_1)Ax_1 - \lambda_{\max}x_2 = 0.$$

Using the fact that $x_1 = av_{\max}$ we argue that

$$a\lambda_{\max}v_{\max} + Ax_2 - a^2\lambda_{\max}\operatorname{diag}(v_{\max})v_{\max} - \lambda_{\max}x_2 = 0.$$

By multiplying on the left by $w_{\max}^T$ we obtain

$$a\lambda_{\max}w_{\max}^Tv_{\max} - a^2\lambda_{\max}w_{\max}^T\operatorname{diag}(v_{\max})v_{\max} = 0,$$

which proves that

$$a = \frac{w_{\max}^Tv_{\max}}{w_{\max}^T\operatorname{diag}(v_{\max})v_{\max}}.$$

Point (iii)c can be proved in a similar way; we refer the reader to (Mei et al., 2017) for the details.

Regarding statement (iv) we refer the reader to (Lajmanovich and Yorke, 1976; Fall et al., 2007) or (Khanafer et al., 2016, Theorems 1 and 2) in the interest of brevity. ∎

16.3 Network Susceptible-Infected-Recovered Model

In this section we review the Susceptible-Infected-Susceptible (SIR) epidemic model.

Scalar SIR model   In this model individuals who recover from infection are assumed not susceptible to the epidemic any more. In this case, the population is divided into three distinct groups: $s(t)$, $x(t)$, and $r(t)$, denoting the fraction of susceptible, infected, and recovered individuals, respectively, with $s(t) + x(t) + r(t) = 1$. We write the (Susceptible–Infected–Recovered) SIR model as:

$$\begin{aligned}
\dot{s}(t) &= -\beta s(t)x(t),\\
\dot{x}(t) &= \beta s(t)x(t) - \gamma x(t), \qquad\qquad (16.8)\\
\dot{r}(t) &= \gamma x(t),
\end{aligned}$$

and present its dynamical behavior as follows.

Lemma 16.4 (Dynamical behavior of the SIR model). *Consider the SIR model* (16.8). *From each initial condition* $s(0) + x(0) + r(0) = 1$ *with* $s(0) > 0$, $x(0) > 0$ *and* $r(0) \geq 0$, *the resulting trajectory* $t \mapsto (s(t), x(t), r(t))$ *has the following properties:*

(i) $s(t) > 0$, $x(t) > 0$, $r(t) \geq 0$, and $s(t) + x(t) + r(t) = 1$ for all $t \geq 0$;

(ii) $t \mapsto s(t)$ is monotonically decreasing and $t \mapsto r(t)$ is monotonically increasing;

(iii) $\lim_{t \to \infty}(s(t), x(t), r(t)) = (s_\infty, 0, r_\infty)$, where r_∞ is the unique solution to the equality

$$1 - r_\infty = s(0)\, \mathrm{e}^{-\frac{\beta}{\gamma}\left(r_\infty - r(0)\right)}; \tag{16.9}$$

(iv) if $\beta s(0)/\gamma < 1$, then $t \mapsto x(t)$ monotonically and exponentially decreases to zero as $t \to \infty$;

(v) if $\beta s(0)/\gamma > 1$, then $t \mapsto x(t)$ first monotonically increases to a maximum value and then monotonically decreases to 0 as $t \to \infty$; the maximum fraction of infected individuals is given by:

$$x_{\max} = x(0) + s(0) - \frac{\gamma}{\beta}\Big(\log(s(0)) + 1 - \log\Big(\frac{\gamma}{\beta}\Big)\Big).$$

As mentioned before, we describe the behavior in statement (v) as an epidemic outbreak, an exponential growth of $t \mapsto x(t)$ for small times.

Network SIR model   The network SIR model on a graph with adjacency matrix A is given by

$$\dot{s}_i(t) = -\beta s_i(t) \sum_{j=1}^{n} a_{ij} x_j(t),$$
$$\dot{x}_i(t) = \beta s_i(t) \sum_{j=1}^{n} a_{ij} x_j(t) - \gamma x_i(t),$$
$$\dot{r}_i(t) = \gamma x_i(t),$$

where $\beta > 0$ is the infection rate and $\gamma > 0$ is the recovery rate. Note that the third equation is redundant because of the constraint $s_i(t) + x_i(t) + r_i(t) = 1$. Therefore, we regard the dynamical system in vector form as:

$$\dot{s}(t) = -\beta \operatorname{diag}(s(t)) A x(t), \tag{16.10a}$$
$$\dot{x}(t) = \beta \operatorname{diag}(s(t)) A x(t) - \gamma x(t). \tag{16.10b}$$

Theorem 16.5 (Dynamical behavior of the network SIR model). *Consider the network SIR model* (16.10)*, with $\beta > 0$ and $\gamma > 0$, over a strongly connected digraph with adjacency matrix A. For $t \geq 0$, let $\lambda_{\max}(t)$ and $w_{\max}(t)$ be the dominant eigenvalue of the non-negative matrix $\operatorname{diag}(s(t))A$ and the corresponding normalized left eigenvector, respectively. The following statements hold:*

(i) *if $x(0) \geq \mathbb{0}_n$ and $s(0) > \mathbb{0}_n$, then*

   a) *$t \mapsto s(t)$ and $t \mapsto x(t)$ are strictly positive for all $t > 0$,*

   b) *$t \mapsto s(t)$ is monotonically decreasing, and*

   c) *$t \mapsto \lambda_{\max}(t)$ is monotonically decreasing;*

(ii) *the set of equilibrium points is the set of pairs $(s^*, \mathbb{0}_n)$, for any $s^* \in [0,1]^n$, and the linearization of model* (16.10) *about $(s^*, \mathbb{0}_n)$ is*

$$\dot{s}(t) = -\beta \operatorname{diag}\left(s^*\right) A x,$$
$$\dot{x}(t) = \beta \operatorname{diag}\left(s^*\right) A x - \gamma x; \tag{16.11}$$

(iii) *(behavior below the threshold) let the time $\tau \geq 0$ satisfy $\beta \lambda_{\max}(\tau) < \gamma$. Then the weighted average $t \mapsto w_{\max}(\tau)^{\mathsf{T}} x(t)$, for $t \geq \tau$, is monotonically and exponentially decreasing to zero;*

(iv) *(behavior above the threshold) if $\beta \lambda_{\max}(0) > \gamma$ and $x(0) > \mathbb{0}_n$, then,*

    a) *(epidemic outbreak) for small time, the weighted average $t \mapsto w_{\max}(0)^{\mathsf{T}} x(t)$ grows exponentially fast with rate $\beta \lambda_{\max}(0) - \gamma$, and*

    b) *there exists $\tau > 0$ such that $\beta \lambda_{\max}(\tau) < \gamma$;*

(v) *each trajectory converges asymptotically to an equilibrium point, that is, $\lim_{t \to \infty} x(t) = \mathbb{0}_n$ so that the epidemic asymptotically disappears.*

In other words, when $\beta \lambda_{\max}(0)/\gamma > 1$, we have an epidemic outbreak, i.e., an exponential growth of infected individual for short time. In any case, the theorem guarantees that, after at most finite time, $\beta \lambda_{\max}(t)/\gamma < 1$ and the infected population decreases exponentially fast to zero.

Proof. Regarding statement (i)a, $s(t) > \mathbb{0}_n$ is due to the fact that Ax is bounded and $s(t)$ is continuously differentiable to t. The statement that $x(t) > \mathbb{0}_n$ for all $t > 0$ is proved in the same way as Theorem 16.2 (i). Statement (i)b is the immediate consequence of $\dot{s}_i(t)$ being strictly negative. From statement (i)a we know that each $s_i(t)$ is positive, and from A being irreducible and $x(0) \neq \mathbb{0}_n$ we know that $\sum_{j=1}^{n} a_{ij} x_j$ is positive. Therefore, $\dot{s}_i(t) = -\beta s_i(t) \sum_{j=1}^{n} a_{ij} x_j(t) < 0$ for all $i \in \{1, \dots, n\}$ and $t \geq 0$.

For statement (i)c, we start by recalling the following property from (Meyer, 2001, Example 7.10.2): for B and C nonnegative square matrices, if $B \leq C$, then $\rho(B) \leq \rho(C)$. Now, pick two time instances t_1 and t_2 with $0 < t_1 < t_2$. Let $\alpha = \max_i s_i(t_2)/s_i(t_1)$ and note $0 < \alpha < 1$ because $s(t)$ is strictly positive and monotonically decreasing. Now note that,

$$\operatorname{diag}(s(t_1))A > \alpha \operatorname{diag}(s(t_1))A \geq \operatorname{diag}(s(t_2))A,$$

so that, using the property above, we know

$$\rho(\operatorname{diag}(s(t_1))A) > \alpha\rho(\operatorname{diag}(s(t_1))A) \geq \rho(\operatorname{diag}(s(t_2))A).$$

This concludes the proof of statement (i)c.

Regarding statement (ii), note that a point (s^*, x^*) is an equilibrium if and only if:

$$\mathbb{0}_n = -\beta \operatorname{diag}(s^*)Ax^*, \quad \text{and}$$
$$\mathbb{0}_n = \beta \operatorname{diag}(s^*)Ax^* - \gamma x^*.$$

Therefore, each point of the form $(s^*, \mathbb{0}_n)$ is an equilibrium. On the other hand, summing the last two equalities we obtain $\mathbb{0}_n = \gamma x^*$ and thus x^* must be $\mathbb{0}_n$. As a straightforward result, the linearization of model (16.10) about any equilibrium point $(s^*, \mathbb{0}_n, \mathbb{1}_n - s^*)$ is given by equation (16.11).

Regarding statement (iii), left multiplying $w_{\max}(\tau)^{\mathsf{T}}$ on both sides of equation (16.10b) we obtain:

$$\frac{d}{dt}\big(w_{\max}(\tau)^{\mathsf{T}} x(t)\big) = w_{\max}(\tau)^{\mathsf{T}} \Big(\beta \operatorname{diag}\big(s(t)\big)Ax(t) - \gamma x(t)\Big),$$

$$\leq w_{\max}(\tau)^{\mathsf{T}} \Big(\beta \operatorname{diag}\big(s(\tau)\big)Ax(t) - \gamma x(t)\Big) = (\beta \lambda_{\max}(\tau) - \gamma)w_{\max}(\tau)^{\mathsf{T}} x(t).$$

Therefore, we obtain

$$w_{\max}(\tau)^\mathsf{T} x(t) \le (w_{\max}(\tau)^\mathsf{T} x(0))\, \mathrm{e}^{(\beta\lambda_{\max}(\tau)-\gamma)t} .$$

The right-hand side exponentially decays to zero when $\beta\lambda_{\max}(\tau) < \gamma$. Therefore, $w_{\max}(\tau)^\mathsf{T} x(t)$ also decreases monotonically and exponentially to zero for all $t > \tau$.

Regarding statement (iv)a, note that based on the argument in (i)a, we only need to consider the case when $x(0) > \mathbb{0}_n$. Left-multiplying $w_{\max}(0)^\mathsf{T}$ on both sides of equation (16.10b), we obtain:

$$\frac{d}{dt}\Big(w_{\max}(0)^\mathsf{T} x(t)\Big)\Big|_{t=0} = w_{\max}(0)^\mathsf{T}\Big(\beta\,\mathrm{diag}\big(s(t)\big)Ax(t) - \gamma x(t)\Big)\Big|_{t=0}$$
$$= (\beta\lambda_{\max}(0) - \gamma)w_{\max}(0)^\mathsf{T} x(0).$$

Since $\beta\lambda_{\max}(0) - \gamma > 0$, the initial time derivative of $w_{\max}(0)^\mathsf{T} x(t)$ is positive. Since $t \mapsto w_{\max}(0)^\mathsf{T} x(t)$ is a continuously differentiable function, there exists $\tau' > 0$ such that $\frac{d}{dt}\big(w_{\max}(0)^\mathsf{T} x(t)\big) > 0$ for any $t \in [0, \tau']$.

Regarding statement (iv)b, since $\dot{s}(t) \le \mathbb{0}_n$ and is lower bounded by $\mathbb{0}_n$, we conclude that the limit $\lim_{t\to+\infty} s(t)$ exists. Moreover, since $s(t)$ is monotonically non-increasing, we have $\lim_{t\to+\infty} \dot{s}(t) = 0$, which implies either $\lim_{t\to+\infty} s(t) = \mathbb{0}_n$ or $\lim_{t\to+\infty} x(t) = \mathbb{0}_n$. If $s(t)$ converges to $\mathbb{0}_n$, then $\dot{x}(t)$ converges to $-\gamma x(t)$. Therefore, there exists $T > 0$ such that $\beta\lambda_{\max}(T) < \gamma$, which leads to $x(t) \to \mathbb{0}_n$ as $t \to +\infty$. If $s(t)$ converges to some $s^* > \mathbb{0}_n$, then $x(t)$ still converges to $\mathbb{0}_n$. Therefore, for any $\big(s(0), x(0)\big)$, the trajectory $\big(s(t), x(t)\big)$ converges to some equilibria with the form $(s^*, \mathbb{0}_n)$, where $s^* \ge \mathbb{0}_n$. Let

$$s(t) = s^* + \delta_s(t), \quad \text{and} \quad x(t) = \mathbb{0}_n + \delta_x(t).$$

We know that $\delta_s(t) \ge 0$ and $\delta_x(t) \ge 0$ for all $t \ge 0$. Moreover, $\delta_s(t)$ is monotonically non-increasing and converges to $\mathbb{0}_n$, and there exists $\tilde{T} > 0$ such that, for any $t \ge T$, $\delta_x(t)$ is monotonically non-increasing and converges to $\mathbb{0}_n$.

Let λ^* and v^* denote the dominant eigenvalue and the corresponding normalized left eigenvector of matrix $\mathrm{diag}(s^*)A$, respectively, that is, $v^{*\mathsf{T}}\,\mathrm{diag}(s^*)A = \lambda^* v^{*\mathsf{T}}$. First let us suppose $\beta\lambda^* - \gamma > 0$, then the linearized system of (16.8) around $(s^*, \mathbb{0}_n)$ is written as

$$\dot{\delta}_s = -\beta\,\mathrm{diag}(s^*)A\delta_x,$$
$$\dot{\delta}_x = \beta\,\mathrm{diag}(s^*)A\delta_x - \gamma\delta_x.$$

Since $\beta\lambda^* - \gamma > 0$, the linearized system is exponentially unstable, which contradicts the fact that $\big(\delta_s(t), \delta_x(t)\big) \to (\mathbb{0}_n, \mathbb{0}_n)$ as $t \to +\infty$. Alternatively, suppose $\beta\lambda^* - \gamma = 0$. By left multiplying $v^{*\mathsf{T}}$ on both sides of the equation for $\dot{x}(t)$ in (16.8), we obtain

$$v^{*\mathsf{T}}\dot{\delta}_x = (\beta\lambda^* - \gamma)(v^{*\mathsf{T}}\delta_x) + \beta v^{*\mathsf{T}}\,\mathrm{diag}(\delta_s)A\delta_x$$
$$= \beta v^{*\mathsf{T}}\,\mathrm{diag}(\delta_s)A\delta_x \ge \mathbb{0}_n,$$

which contradicts $\delta_x(t) \to \mathbb{0}_n$ as $t \to +\infty$. Therefore, we conclude that $\beta\lambda^* - \gamma < 0$. Since $\lambda_{\max}(t)$ is continuous on t, we conclude that there exists $\tau < +\infty$ such that $\beta\lambda_{\max}(t) - \gamma < 0$.

∎

In the rest of this section, we present some numerical results for the network SIR model on the undirected unweighted graph illustrated in Figure 16.2. The adjacency matrix A is binary. Unless otherwise stated, the system parameters are $\beta = 0.5$ and $\gamma = 0.4$. As initial condition, we select one node fully infected (the dark-gray node in Figure 16.2, say, with index 1), 19 fully healthy individuals, and zero recovered fraction — corresponding to $x(0) = \mathbb{e}_1$, $r(0) = \mathbb{0}_n$, and $s(0) = \mathbb{1}_n - x(0)$.
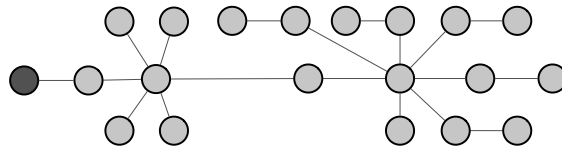
Figure 16.2: Sample undirected unweighted graph with 20 nodes

The left image in Figure 16.3 illustrates the time evolution of $(\beta/\gamma)\lambda_{\max}(t)$ with varying network parameters. Note that each evolution starts above the threshold, reaches the threshold value 1 in finite time, and converges to a final value below 1. The right image in Figure 16.3 illustrates the behavior of the average susceptible, average infected and average recovered quantities in populations starting from a small initial infection fraction and with $\beta\lambda_{\max}(t)/\gamma$ above 1 at time 0. Note that the evolution of the infected fraction of the population displays a unimodal dependence on time, like in the scalar model.

(a) The spectral radius of $(\beta/\gamma)\operatorname{diag}(s(t))A$; the parameter γ takes value in $.1, .2, \ldots, .9$ corresponding, respectively, to the curves from top to bottom.

(b) Fraction of susceptible, infected, and recovered individuals. At initial time we have $\beta\lambda_{\max}(0)/\gamma = 3.57$.
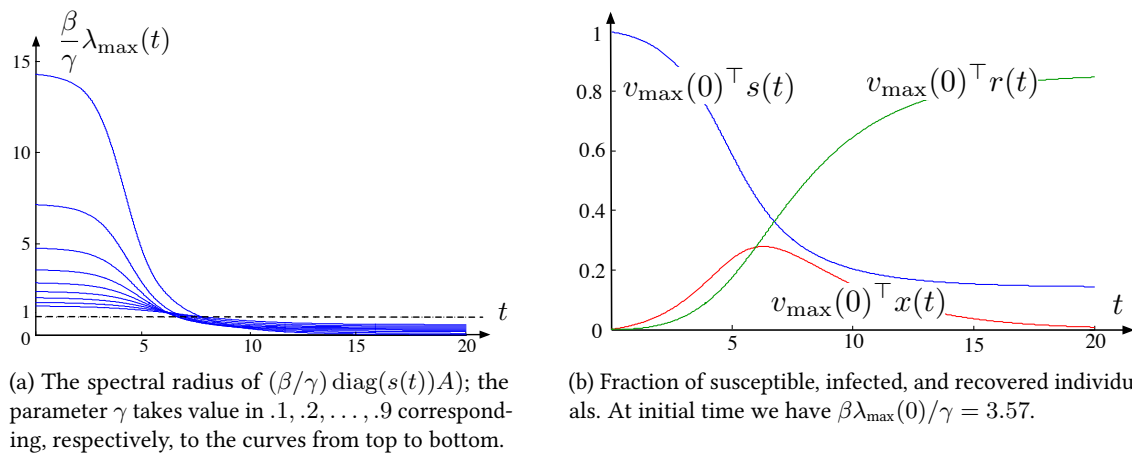
Figure 16.3: Evolution of the network SIR model from initial condition consisting of one node fully infected individual (the dark-gray node in Figure 16.2), 19 fully healthy individuals, and zero recovered fraction.

16.4 Appendix: The stochastic network SI model

Building on Appendix 13.4, we now present and study a stochastic model of the propagation phenomenon over a contact network.

The stochastic model   The *stochastic network SI model*, illustrated in Figure 16.4, is defined as follows:

(i) We consider a group of n individuals. The state of each individual is either S for susceptible or I for infected.

(ii) The n individuals are in pairwise contact, as specified by an undirected graph G with adjacency matrix A (without self-loops). The edge weights represent the frequency of contact among two individuals.

(iii) Each individual in susceptible status can transition to infected as follows: given an *infection rate* $\beta > 0$, if a susceptible individual i is in contact with an infected individual j for time Δt, the probability of infection is $a_{ij}\beta\Delta t$. Each individual can be infected by any neighboring individual: these random events are independent.
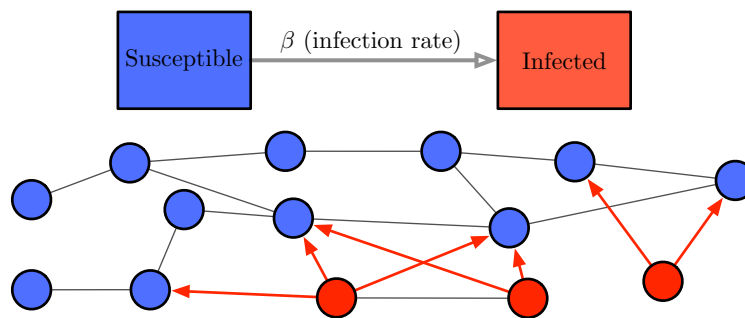
Figure 16.4: In the stochastic network SI model, each susceptible individual (blue) becomes infected by contact with infected individuals (red) in its neighborhood according to an infection rate β.

An approximate deterministic model   We define the *infection variable* at time t for individual i by

$$Y_i(t) = \begin{cases} 1, & \text{if node } i \text{ is in state } I \text{ at time } t, \\ 0, & \text{if node } i \text{ is in state } S \text{ at time } t, \end{cases}$$

and the *expected infection*, which turns out to be equal to the *probability of infection*, of individual i by

$$x_i(t) = \mathbb{E}[Y_i(t)] = 1 \cdot \mathbb{P}[Y_i(t) = 1] + 0 \cdot \mathbb{P}[Y_i(t) = 0]$$
$$= \mathbb{P}[Y_i(t) = 1].$$

In what follows it will be useful to approximate $\mathbb{P}[Y_i(t) = 0 \,|\, Y_j(t) = 1]$ with $\mathbb{P}[Y_i(t) = 0]$, that is, to require Y_i and Y_j to be independent for arbitrary i and j. We claim this approximation is acceptable over certain graphs with large numbers n of individuals. The final model, which we obtain below based on the Independence Approximation, is an upper bound on the true model because $\mathbb{P}[Y_i(t) = 0] \geq \mathbb{P}[Y_i(t) = 0 \,|\, Y_j(t) = 1]$.

Definition 16.6 (Independence Approximation). *For any two individuals i and j, the infection variables Y_i and Y_j are independent.*

Theorem 16.7 (From the stochastic to the deterministic network SI model). *Consider the stochastic network SI model with infection rate β over a contact graph with adjacency matrix A. The probabilities of infection satisfy*

$$\frac{d}{dt}\mathbb{P}[Y_i(t) = 1] = \beta \sum_{j=1}^{n} a_{ij}\mathbb{P}[Y_i(t) = 0, Y_j(t) = 1].$$

Moreover, under the Independence Approximation 16.6, the probabilities of infection $x_i(t) = \mathbb{P}[Y_i(t) = 1]$, $i \in \{1, \dots, n\}$, satisfy (deterministic) network SI model *defined by*

$$\dot{x}_i(t) = \beta(1 - x_i(t)) \sum_{j=1}^{n} a_{ij}x_j(t).$$

We study the deterministic network SI model in the next section.

Proof. We start by defining the random variables

$$Y_{-i}(t) = (Y_1(t), \dots, Y_{i-1}(t), Y_{i+1}(t), \dots, Y_n(t)),$$

and, similarly, $Y_{-i-j}(t)$, for $i, j \in \{1, \dots, n\}$. We are interested in the event that a susceptible individual remains susceptible (or, vice versa, the event that susceptible individual becomes infected) over the interval of time $[t, t + \Delta t]$, for a short interval duration Δt. We start by computing the probability of non-infection for a duration Δt, conditioned upon $Y_{-i}(t)$:

$$\mathbb{P}[Y_i(t + \Delta t) = 0 \,|\, Y_i(t) = 0, Y_{-i}(t)] = \prod_{j=1}^{n} \big(1 - a_{ij}Y_j(t)\beta\Delta t\big) = 1 - \sum_{j=1}^{n} a_{ij}Y_j(t)\beta\Delta t + \mathcal{O}(\Delta t^2).$$

The complementary probability, i.e., the probability of infection for duration Δt is:

$$\mathbb{P}[Y_i(t + \Delta t) = 1 \,|\, Y_i(t) = 0, Y_{-i}(t)] = \sum_{j=1}^{n} a_{ij}Y_j(t)\beta\Delta t + \mathcal{O}(\Delta t^2).$$

We are now ready to study the random variable $Y_i(t + \Delta t) - Y_i(t)$, given $Y_{-i}(t)$:

$$\begin{aligned}
\mathbb{E}[Y_i(t+\Delta t) &- Y_i(t) \,|\, Y_{-i}(t)] \\
&= 1 \cdot \mathbb{P}[Y_i(t + \Delta t) = 1, Y_i(t) = 0 \,|\, Y_{-i}(t)] \\
&\quad + 0 \cdot \mathbb{P}\big[(Y_i(t + \Delta t) = Y_i(t) = 0) \text{ or } (Y_i(t + \Delta t) = Y_i(t) = 1) \,|\, Y_{-i}(t)\big] \quad \text{(by def. expectation)} \\
&= \mathbb{P}[Y_i(t + \Delta t) = 1 \,|\, Y_i(t) = 0, Y_{-i}(t)] \cdot \mathbb{P}[Y_i(t) = 0 \,|\, Y_{-i}(t)] \quad \text{(by conditional prob.)} \\
&= \Big(\sum_{j=1}^{n} a_{ij}Y_j(t)\beta\Delta t + \mathcal{O}(\Delta t^2) \Big) \cdot \mathbb{P}[Y_i(t) = 0 \,|\, Y_{-i}(t)].
\end{aligned}$$

We now remove the conditioning upon $Y_{-i}(t)$ and compute

$$\begin{aligned}
\mathbb{E}[Y_i(t + \Delta t) - Y_i(t)] &= \mathbb{E}\big[\, \mathbb{E}[Y_i(t + \Delta t) - Y_i(t) \,|\, Y_{-i}(t)]\big] \\
&= \Big(\sum_{j=1}^{n} a_{ij}\beta\Delta t \Big) \cdot \mathbb{E}\big[Y_j(t) \cdot \mathbb{P}[Y_i(t) = 0 \,|\, Y_{-i}(t)]\big] + \mathcal{O}(\Delta t^2),
\end{aligned}$$

Next, assuming y is an arbitrary realization of the random variable Y, we have

$$\mathbb{E}\big[Y_j(t) \cdot \mathbb{P}[Y_i(t) = 0 \,|\, Y_{-i}(t)]\big]$$

$$= \sum\nolimits_{y_{-i}} y_j \cdot \mathbb{P}[Y_i(t) = 0 \,|\, Y_{-i}(t) = y_{-i}] \cdot \mathbb{P}[Y_{-i}(t) = y_{-i}] \qquad \text{(by def. expectation)}$$

$$= \sum\nolimits_{y_{-i-j}} 1 \cdot \mathbb{P}[Y_i(t) = 0 \,|\, Y_{-i-j}(t) = y_{-i-j}, Y_j(t) = 1]$$

$$\qquad\qquad \times \mathbb{P}[Y_{-i-j}(t) = y_{-i-j}, Y_j(t) = 1] \qquad \text{(because } y_j \in \{0,1\})$$

$$= \sum_{y_{-i-j}} \mathbb{P}[Y_i(t) = 0, Y_{-i-j}(t) = y_{-i-j}, Y_j(t) = 1] \qquad \text{(by conditional prob.)}$$

$$= \mathbb{P}[Y_i(t) = 0, Y_j(t) = 1],$$

where the first summation is taken over all possible values y_{-i} that the variable $Y_{-i}(t)$ takes and a similar convention applies to the other summations. In summary, we know

$$\mathbb{E}[Y_i(t + \Delta t) - Y_i(t)] = \sum_{j=1}^{n} a_{ij}\beta\Delta t \cdot \mathbb{P}[Y_i(t) = 0, Y_j(t) = 1] + \mathcal{O}(\Delta t^2),$$

so that, also recalling $\mathbb{P}[Y_i(t) = 1] = \mathbb{E}[Y_i(t)]$,

$$\frac{d}{dt}\mathbb{P}[Y_i(t) = 1] = \lim_{\Delta t \to 0^+} \frac{\mathbb{E}[Y_i(t + \Delta t) - Y_i(t)]}{\Delta t} = \beta \sum_{j=1}^{n} a_{ij}\mathbb{P}[Y_i(t) = 0, Y_j(t) = 1].$$

The final step is an immediate consequence of the Independence Approximation: $\mathbb{P}[Y_i(t) = 0, Y_j(t) = 1] = \mathbb{P}[Y_i(t) = 0 \,|\, Y_j(t) = 1] \cdot \mathbb{P}[Y_j(t) = 1] \approx (1 - \mathbb{P}[Y_i(t) = 1]) \cdot \mathbb{P}[Y_j(t) = 1]$. ∎

16.5 Historical notes and further reading

The dynamics of several classic scalar epidemic models are surveyed by Hethcote (2000).

The earliest work on the network SIS model is (Lajmanovich and Yorke, 1976); this article proposes a rigorous analysis of the threshold for the epidemic outbreak, which depends on both the disease parameters and the spectral radius of the contact network. For the case when the basic reproduction number is above the epidemic threshold, this paper establishes the existence and uniqueness of a nonzero steady-state infection probability, called the endemic state. Lajmanovich and Yorke (1976) refer to the model and the multi-group or multi-population SIS model. Numerous extensions and variations on these basic results have appeared over the years.

Allen (1994) proposes and analyzes a discrete-time network SIS model. This work appears to be the first to revisit and formally reproduce for the discrete-time case the earlier results by (Lajmanovich and Yorke, 1976); see also the later (Wang et al., 2003). Further recent results on the discrete-time model are obtained by Ahn and Hassibi (2013) and by Azizan Ruhi and Hassibi (2015).

Van Mieghem et al. (2009) rediscoveres this model, refers to it as the intertwined SIS model, and argue that the (continuous-time) network SIS model is in fact the mean-field approximation of the original Markov-chain SIS model of exponential dimension; this claim is rigorously proven by Sahneh et al. (2013).

Fall et al. (2007) and Khanafer et al. (2016) discuss the continuous-time network SIS model in a modern language. Fall et al. (2007) apply Lyapunov theory and Metzler matrix theory to establish existence, uniqueness, and stability of the equilibrium points below and above the epidemic threshold. Khanafer et al. (2016) use positive system theory in their analysis and extend the existence, uniqueness, and stability results to the setting of weakly connected digraphs. Mei et al. (2017) propose an algorithm and Taylor expansions for the endemic state above the epidemic threshold.

An early work by Hethcote (1978) proposes a general multi-group SIR model with birth, death, immunization, and de-immunization. The epidemic threshold and the equilibria below/above the threshold are characterized. For the simplified model without birth/death and de-immunization, Hethcote (1978) proves that the system converges asymptotically to an all-healthy state. Guo et al. (2008) consider a generalized network SIR model with vital dynamics, that is, with birth and death. Youssef and Scoglio (2011) study a special case of the network SIR model under the name of individual-based SIR model over undirected networks.

We conclude by mentioning other surveys and textbook treatments. In (Mesbahi and Egerstedt, 2010), the stability of equilibria for a generalized SIR model is reviewed through Lyapunov and graph theory. (Newman, 2010, Chapter 17), (Easley and Kleinberg, 2010, Chapter 21), and (Barrat et al., 2008, Chapter 9) review various heterogeneous epidemic models. The recent survey by Nowzari et al. (2016) presents various epidemic models and discusses related control problems.

16.6   Exercises

E16.1   **Initial evolution of network SIS model.** Consider the network SIS model with initial fraction $x(0) = \varepsilon x_0$, where we take $x_0 \ll \mathbb{1}_n$ and $\varepsilon \ll 1$. Show that in the time scale $t(\varepsilon) = \ln(1/\varepsilon)/(\beta \lambda_{\max})$, the linearized evolution satisfies

$$\lim_{\varepsilon \to 0^+} x\big(t(\varepsilon)\big) = \big(w_{\max}^{\mathsf{T}} x_0\big) w_{\max}.$$

Networks of Coupled Oscillators

In this chapter we continue our discussion about coupled-oscillator models and their behavior. Starting from the basic models discussed in Section 13.3, we here focus on characterizing synchronization and other dynamic phenomena.

Recall that the coupled-oscillators model, as given in equation (13.15), is

$$\dot{\theta}_i = \omega_i - \sum_{j=1}^n a_{ij} \sin(\theta_i - \theta_j), \qquad i \in \{1, \ldots, n\}, \tag{17.1}$$

and its homogeneous counterpart, the Kuramoto model (13.16), is

$$\dot{\theta}_i = \omega_i - \frac{K}{n} \sum_{j=1}^n \sin(\theta_i - \theta_j), \qquad i \in \{1, \ldots, n\}. \tag{17.2}$$

17.1 Preliminary notation and analysis

17.1.1 The geometry of the circle and the torus

Parametrization　The unit circle is $\mathbb{S}^1$. The torus $\mathbb{T}^n$ is the set consisting of n-copies of the circle. We parametrize the circle $\mathbb{S}^1$ by assuming (i) angles are measured counterclockwise, (ii) the 0 angle is the intersection of the unit circle with the positive horizontal axis, and (iii) angles take value in $[-\pi, \pi[$.

Geodesic distance　The *clockwise arc-length from θ_i to θ_j* is the length of the clockwise arc from θ_i to θ_j. The counterclockwise arc-length is defined analogously. The *geodesic distance between θ_i and θ_j* is the minimum between clockwise and counterclockwise arc-lengths and is denoted by $|\theta_i - \theta_j|$. In the parametrization:

$$\text{dist}_{\textsf{cc}}(\theta_1, \theta_2) = \text{mod}((\theta_2 - \theta_1), 2\pi), \qquad \text{dist}_{\textsf{c}}(\theta_1, \theta_2) = \text{mod}((\theta_1 - \theta_2), 2\pi)$$
$$|\theta_1 - \theta_2| = \min\{\text{dist}_{\textsf{c}}(\theta_1, \theta_2), \text{dist}_{\textsf{cc}}(\theta_1, \theta_2)\}.$$

Rotations   Given the angle $\alpha \in [-\pi, \pi[$, the rotation of the n-tuple $\theta = (\theta_1, \ldots, \theta_n) \in \mathbb{T}^n$ by α, denoted by $\mathrm{rot}_\alpha(\theta)$, is the counterclockwise rotation of each entry $(\theta_1, \ldots, \theta_n)$ by α. For $\theta = \in \mathbb{T}^n$, we also define its *rotation set* to be

$$[\theta] = \{\mathrm{rot}_\alpha(\theta) \in \mathbb{T}^n \mid \alpha \in [-\pi, \pi[\}.$$

The coupled oscillator model (17.1) is *invariant* under rotations, that is, given a solution $\theta : \mathbb{R}_{\geq 0} \to \mathbb{T}^n$ to the coupled oscillator model, a rotation of $\mathrm{rot}_\alpha(\theta(t))$ by any angle α is again a solution.

Arc subsets of the n-torus   Given a length $\gamma \in [0, 2\pi[$, the *arc subset* $\overline{\Gamma}_{\mathrm{arc}}(\gamma) \subset \mathbb{T}^n$ is the set of n-tuples $(\theta_1, \ldots, \theta_n)$ such that there exists an arc of length γ containing all $\theta_1, \ldots, \theta_n$. The set $\Gamma_{\mathrm{arc}}(\gamma)$ is the interior of $\overline{\Gamma}_{\mathrm{arc}}(\gamma)$. For example, $\theta \in \overline{\Gamma}_{\mathrm{arc}}(\pi)$ implies all angles $\theta_1, \ldots, \theta_n$ belong to a closed half circle. Note:

(i) If $(\theta_1, \ldots, \theta_n) \in \overline{\Gamma}_{\mathrm{arc}}(\gamma)$, then $|\theta_i - \theta_j| \leq \gamma$ for all i and j. The converse is not true in general. For example, $\{\theta \in \mathbb{T}^n \mid |\theta_i - \theta_j| \leq \pi \text{ for all } i, j\}$ is equal to the entire $\mathbb{T}^n$. However, the converse statement is true in the following form (see also Exercise E17.1): if $|\theta_i - \theta_j| \leq \gamma$ for all i and j and $(\theta_1, \ldots, \theta_n) \in \overline{\Gamma}_{\mathrm{arc}}(\pi)$, then $(\theta_1, \ldots, \theta_n) \in \overline{\Gamma}_{\mathrm{arc}}(\gamma)$.

(ii) If $\theta = (\theta_1, \ldots, \theta_n) \in \Gamma_{\mathrm{arc}}(\pi)$, then $\mathrm{average}(\theta)$ is well posed. (The average of n angles is ill-posed in general. For example, there is no reasonable definition of the average of two diametrically-opposed points.)

17.1.2 Synchronization notions

Consider the following notions of synchronization for a solution $\theta : \mathbb{R}_{\geq 0} \to \mathbb{T}^n$:

Frequency synchrony: A solution $\theta : \mathbb{R}_{\geq 0} \to \mathbb{T}^n$ is *frequency synchronized* if $\dot{\theta}_i(t) = \dot{\theta}_j(t)$ for all time t and for all i and j.

Phase synchrony: A solution $\theta : \mathbb{R}_{\geq 0} \to \mathbb{T}^n$ is *phase synchronized* if $\theta_i(t) = \theta_j(t)$ for all time t and for all i and j.

Phase cohesiveness: A solution $\theta : \mathbb{R}_{\geq 0} \to \mathbb{T}^n$ is *phase cohesive* with respect to $\gamma > 0$ if one of the following conditions holds for all time t:

  (i) $\theta(t) \in \Gamma_{\mathrm{arc}}(\gamma)$;
  (ii) $|\theta_i(t) - \theta_j(t)| \leq \gamma$ for all edges (i, j) of a graph of interest; or
  (iii) $\sqrt{\sum_{i,j=1}^n |\theta_i(t) - \theta_j(t)|^2 / 2} < \gamma$.

Asymptotic notions: We will also talk about solutions that *asymptotically achieve* certain synchronization properties. For example, a solution $\theta : \mathbb{R}_{\geq 0} \to \mathbb{T}^n$ *achieves phase synchronization* if $\lim_{t \to \infty} |\theta_i(t) - \theta_j(t)| = 0$. Analogous definitions can be given for asymptotic frequency synchronization and asymptotic phase cohesiveness.

Finally, notice that phase synchrony is the extreme case of all phase cohesiveness notions with $\gamma = 0$.

17.1.3   Preliminary results

We have the following result on the synchronization frequency.

Lemma 17.1 (Synchronization frequency). *If a solution of the coupled oscillator model* (17.1) *achieves frequency synchronization, then it does so with a constant synchronization frequency equal to*

$$\omega_{\text{sync}} \triangleq \frac{1}{n} \sum_{i=1}^{n} \omega_i = \text{average}(\omega).$$

Proof. This fact is obtained by summing all equations (17.1) for $i \in \{1, \dots, n\}$. ∎

Lemma 17.1 implies that, by expressing each angle with respect to a rotating frame with frequency ω_{sync} and by replacing ω_i by $\omega_i - \omega_{\text{sync}}$, we obtain $\omega_{\text{sync}} = 0$ or, equivalently, $\omega \in \mathbb{1}_n^\perp$. In this rotating frame a frequency-synchronized solution is an equilibrium. Due to the rotational invariance of the coupled oscillator model (17.1), it follows that if $\theta^* \in \mathbb{T}^n$ is an equilibrium point, then every point in the rotation set

$$[\theta^*] = \{\theta \in \mathbb{T}^n \mid \text{rot}_\alpha(\theta^*), \alpha \in [-\pi, \pi[\}$$

is also an equilibrium. Notice that the set $[\theta^*]$ is a connected circle in $\mathbb{T}^n$, and we refer to it as an *equilibrium set*.

We have the following important result on local stability properties of equilibria.

Lemma 17.2 (Linearization). *Assume the frequencies satisfy $\omega \in \mathbb{1}_n^\perp$ and G is connected with incidence matrix B. The following statements hold:*

(i) *Jacobian: the Jacobian of the coupled oscillator model* (17.1) *at $\theta \in \mathbb{T}^n$ is*

$$J(\theta) = -B \, \text{diag}(\{a_{ij} \cos(\theta_i - \theta_j)\}_{\{i,j\} \in E}) B^\mathsf{T},$$

(ii) *Local stability: if there exists an equilibrium θ^* such that $|\theta_i^* - \theta_j^*| < \pi/2$ for all $\{i, j\} \in E$, then*

   a) *$-J(\theta^*)$ is a Laplacian matrix; and*
   b) *the equilibrium set $[\theta^*]$ is locally exponentially stable.*

Proof. We start with statements (i) and (ii)a. Given $\theta \in \mathbb{T}^n$, we define the undirected graph $G_{\text{cosine}}(\theta)$ with the same nodes and edges as G and with edge weights $a_{ij} \cos(\theta_i - \theta_j)$. Next, we compute

$$\frac{\partial}{\partial \theta_i}\big(\omega_i - \sum\nolimits_{j=1}^{n} a_{ij} \sin(\theta_i - \theta_j)\big) = -\sum\nolimits_{j=1}^{n} a_{ij} \cos(\theta_i - \theta_j),$$

$$\frac{\partial}{\partial \theta_j}\big(\omega_i - \sum\nolimits_{k=1}^{n} a_{ik} \sin(\theta_i - \theta_k)\big) = a_{ij} \cos(\theta_i - \theta_j).$$

Therefore, the Jacobian is equal to minus the Laplacian matrix of the (possibly negatively weighted) graph $G_{\text{cosine}}(\theta)$ and statement (i) follows from Lemma 8.1. Regarding statement (ii)a, if $|\theta_i^* - \theta_j^*| < \pi/2$ for all $\{i, j\} \in E$, then $\cos(\theta_i^* - \theta_j^*) > 0$ for all $\{i, j\} \in E$, so that $G_{\text{cosine}}(\theta)$ has strictly non-negative weights and all usual properties of Laplacian matrices hold.

To prove statement (ii)b notice that $J(\theta^*)$ is negative semidefinite with the nullspace $\mathbb{1}_n$ arising from the rotational symmetry. All other eigenvectors are orthogonal to $\mathbb{1}_n$ and have negative eigenvalues. We now restrict our analysis to the orthogonal complement of $\mathbb{1}_n$: we define a coordinate transformation matrix $Q \in \mathbb{R}^{(n-1) \times n}$ with orthonormal rows orthogonal to $\mathbb{1}_n$,

$$Q\mathbb{1}_n = \mathbb{0}_{n-1} \quad \text{and} \quad QQ^\mathsf{T} = I_{n-1},$$

and we note that $QJ(\theta^*)Q^\mathsf{T}$ has negative eigenvalues. Therefore, in the original coordinates, the zero eigenspace $\mathbb{1}_n$ is exponentially stable. By Theorem 14.10, the corresponding equilibrium set $[\theta^*]$ is locally exponentially stable.   ∎

Corollary 17.3 (Frequency synchronization). *If a solution of the coupled oscillator model* (17.1) *satisfies the phase cohesiveness properties* $|\theta_i(t) - \theta_j(t)| \leq \gamma$ *for some* $\gamma \in [0, \pi/2[$ *and for all* $t \geq 0$, *then the coupled oscillator model* (17.1) *achieves exponential frequency synchronization.*

Proof. Let $x_i(t) = \dot{\theta}_i(t)$ be the frequency. Then $\dot{x}(t) = J(\theta(t))x(t)$ is a time-varying averaging system. The associated undirected graph has time-varying yet strictly positive weights $a_{ij} \cos(\theta_i(t) - \theta_j(t)) \geq a_{ij} \cos(\gamma) > 0$ for each $\{i, j\} \in E$. Hence, the weighted graph is connected for each $t \geq 0$. From the analysis of time-varying averaging systems in Theorem 11.11, the exponential convergence of $x(t)$ to $\text{average}(x(0))\mathbb{1}_n$ follows. Equivalently, the frequencies synchronize.   ∎

17.1.4 The order parameter and the mean field model

An alternative synchronization measure (besides phase cohesiveness) is the magnitude of the *order parameter*

$$re^{\mathrm{i}\psi} = \frac{1}{n} \sum_{j=1}^{n} e^{\mathrm{i}\theta_j} . \tag{17.3}$$

The order parameter (17.3) is the centroid of all oscillators represented as points on the unit circle in $\mathbb{C}^1$. The magnitude r of the order parameter is a synchronization measure:

- if the oscillators are phase-synchronized, then $r = 1$;
- if the oscillators are spaced equally on the unit circle, then $r = 0$; and
- for $r \in]0, 1[$ and oscillators contained in a semi-circle, the associated configuration of oscillators satisfy a certain level of phase cohesiveness; see Exercise E17.2.

By means of the order parameter $re^{\mathrm{i}\psi}$ the all-to-all Kuramoto model (17.2) can be rewritten in the insightful form

$$\dot{\theta}_i = \omega_i - Kr\sin(\theta_i - \psi), \quad i \in \{1, \ldots, n\} . \tag{17.4}$$

(We ask the reader to establish this identity in Exercise E17.3.) Equation (17.4) gives the intuition that the oscillators synchronize because of their coupling to a *mean field* represented by the order parameter $re^{\mathrm{i}\psi}$, which itself is a function of $\theta(t)$. Intuitively, for small coupling strength K each oscillator rotates with its distinct natural frequency ω_i, whereas for large coupling strength K all angles $\theta_i(t)$ will entrain to the mean field $re^{\mathrm{i}\psi}$, and the oscillators synchronize. The transition from incoherence to synchrony occurs at a critical threshold value of the coupling strength, denoted by K_{critical}.

17.2 Synchronization of identical oscillators

We start our discussion with the following insightful lemma.

Lemma 17.4. *Consider the coupled oscillator model* (17.1). *If $\omega_i \neq \omega_j$ for some distinct $i, j \in \{1, \ldots, n\}$, then the oscillators cannot achieve phase synchronization.*

Proof. We prove the lemma by contraposition. Assume that all oscillators are in phase synchrony $\theta_i(t) = \theta_j(t)$ for all $t \geq 0$ and all $i, j \in \{1, \ldots, n\}$. Then by equating the dynamics, $\dot{\theta}_i(t) = \dot{\theta}_j(t)$, it follows necessarily that $\omega_i = \omega_j$. ∎

Motivated by Lemma 17.4, we consider oscillators with identical natural frequencies, $\omega_i = \omega \in \mathbb{R}$ for all $i \in \{1, \ldots, n\}$. By working in a rotating frame with frequency ω, we have $\omega = 0$. Thus, we consider the model

$$\dot{\theta}_i = -\sum_{j=1}^{n} a_{ij} \sin(\theta_i - \theta_j), \qquad i \in \{1, \ldots, n\}. \tag{17.5}$$

Notice that phase synchronization is an equilibrium of the this model. Conversely, phase synchronization cannot be an equilibrium of the original coupled oscillator model (17.1) if $\omega_i \neq \omega_j$.

17.2.1 An averaging-based approach

Let us first analyze the coupled oscillator model (17.5) with initial conditions restricted to an open semi-circle, $\theta(0) \in \Gamma_{\mathrm{arc}}(\gamma)$ for some $\gamma \in [0, \pi[$. In this case, the oscillators remain in a semi-circle at least for small times $t > 0$ and the two coordinate transformations

$$x_i(t) = \tan(\theta_i(t)) \ \ (\text{with } x_i \in \mathbb{R}), \quad \text{and} \quad y_i(t) = \theta_i(t) \ \ (\text{with } y_i \in \mathbb{R})$$

are well-defined and bijective (at least for small times).

In the x_i-coordinates, the coupled oscillator model reads as the time-varying continuous-time averaging system

$$\dot{x}_i(t) = -\sum_{j=1}^{n} b_{ij}(t)(x_i(t) - x_j(t)), \tag{17.6}$$

where $b_{ij}(t) = a_{ij}\sqrt{(1 + x_i(t)^2)/(1 + x_j(t)^2)}$ and $b_{ij}(t) \geq a_{ij}\cos(\gamma/2)$; see Exercise E17.7 for a derivation. Similarly, in the y_i-coordinates, the coupled oscillator model reads as

$$\dot{y}_i(t) = -\sum_{j=1}^{n} c_{ij}(t)(y_i(t) - y_j(t)), \tag{17.7}$$

where $c_{ij}(t) = a_{ij}\operatorname{sinc}(y_i(t) - y_j(t))$ and $c_{ij}(t) \geq a_{ij}\operatorname{sinc}(\gamma)$. Notice that both averaging formulations (17.6) and (17.7) are well-defined as long as the the oscillators remain in a semi-circle $\Gamma_{\mathrm{arc}}(\gamma)$ for some $\gamma \in [0, \pi[$.

Theorem 17.5 (Phase cohesiveness and synchronization in open semicircle). *Consider the coupled oscillator model* (17.5) *with a connected, undirected, and weighted graph $G = (\{1, \ldots, n\}, E, A)$. The following statements hold:*

(i) phase cohesiveness: for each $\gamma \in [0, \pi[$ each solution orginating in $\Gamma_{\mathrm{arc}}(\gamma)$ remains in $\Gamma_{\mathrm{arc}}(\gamma)$ for all times;

(ii) *asymptotic phase synchronization: each trajectory originating in* $\Gamma_{\mathrm{arc}}(\gamma)$ *for* $\gamma \in [0, \pi[$ *achieves exponential phase synchronization, that is,*

$$\|\theta(t) - \mathrm{average}(\theta(0))\mathbb{1}_n\|_2 \leq \|\theta(0) - \mathrm{average}(\theta(0))\mathbb{1}_n\|_2 e^{\lambda_{\mathrm{ps}}t}, \qquad (17.8)$$

where $\lambda_{\mathrm{ps}} = -\lambda_2(L)\cos(\gamma/2)$.

Proof. Consider the averaging formulations (17.6) and (17.7) with initial conditions $\theta(0) \in \Gamma_{\mathrm{arc}}(\gamma)$ for some $\gamma \in [0, \pi[$. By continuity, for small positive times $t > 0$, the oscillators remain in a semi-circle, the time-varying weights $b_{ij}(t) \geq a_{ij}(\cos(\gamma/2)$ and $c_{ij}(t) \geq a_{ij}\mathrm{sinc}(\gamma)$ are strictly positive for each $\{i, j\} \in E$, the associated time-dependent graph is connected. As one establishes in the proof of Theorem 11.11, the max-min function $V_{\mathrm{max\text{-}min}}$, defined in equation (14.6), evaluated along the solutions to the time-varying consensus systems (17.6) and (17.7) are strictly decreasing for until consensus is reached.

Thus, the oscillators remain in $\Gamma_{\mathrm{arc}}(\gamma)$ phase synchronization exponentially fast. Since the graph is undirected, we can also conclude convergence to the average phase. Finally, the explicit convergence estimate (17.8) follows, for example, by analyzing (17.6) with the disagreement Lyapunov function and using $b_{ij}(t) \geq a_{ij}\cos(\gamma/2)$. $\blacksquare$

17.2.2   The potential landscape, convergence and phase synchronization

The consensus analysis in Theorem 17.5 leads to a powerful result but is inherently restricted to a semi-circle. To overcome this limitation, we use potential functions as an analysis tool. Inspired by Examples #1 and #3 in Section 13.3, define the potential function $U : \mathbb{T}^n \to \mathbb{R}$ by

$$U(\theta) = \sum_{\{i,j\} \in E} a_{ij}\big(1 - \cos(\theta_i - \theta_j)\big). \qquad (17.9)$$

Then the coupled oscillator model (17.1) (with all $\omega_i = 0$) can be formulated as the negative gradient flow

$$\dot{\theta} = -\frac{\partial U(\theta)}{\partial \theta}^{\mathsf{T}}. \qquad (17.10)$$

Among the many critical points of the potential function U in equation (17.9), each point in the set of phase-synchronized angles is a global minimum of U. This fact can be easily seen since each summand in (17.9) is bounded in $[0, 2a_{ij}]$ and the lower bound is reached only if neighboring oscillators are phase-synchronized.

Theorem 17.6 (Phase synchronization). *Consider the coupled oscillator model* (17.5) *with a connected, undirected, and weighted graph* $G = (\{1, \dots, n\}, E, A)$. *Then*

(i) *Global convergence: For all initial conditions* $\theta(0) \in \mathbb{T}^n$, *the phases* $\theta_i(t)$ *converge to the set of critical points* $\{\theta \in \mathbb{T}^n \mid \partial U(\theta)/\partial \theta = \mathbb{0}_n^{\mathsf{T}}\}$; *and*

(ii) *Local stability: Phase synchronization is a locally exponentially stable equilibrium set.*

Proof. Statement (i) is an immediate consequence of the Krasovskiĭ-LaSalle Invariance Principle in Theorem 14.7. Statement (ii) follows from the Jacobian result in Lemma 17.2 and Theorem 14.10. $\blacksquare$

Theorem 17.6 together with Theorem 17.5 gives a fairly complete picture of the convergence and phase synchronization properties of the coupled oscillator model (17.5).

According to Theorem 17.6 phase synchronization is only locally stable. A stronger result can be made in case of an all-to-all homogeneous coupling graph, that is, for the Kuramoto model (17.2).

Corollary 17.7 (Almost global phase synchronization for the Kuramoto model). *Consider the Kuramoto model* (17.2) *with identical natural frequencies* $\omega_i = \omega_j$ *for all* $i, j \in \{1, \ldots, n\}$. *Then for almost all initial conditions in* $\mathbb{T}^n$, *the oscillators achieve phase synchronization.*

Proof. For identical natural frequencies, the Kuramoto model (17.2) can be put in rotating coordinates so that $\omega_i = 0$ for all $i \in \{1, \ldots, n\}$; see Section 17.2. The Kuramoto model reads in the order-parameter formulation (17.4) as

$$\dot{\theta}_i = -Kr\sin(\theta_i - \psi), \quad i \in \{1, \ldots, n\}. \tag{17.11}$$

The associated potential function reads as (see Exercise E17.5)

$$U(\theta) = \sum_{\{i,j\} \in E} a_{ij}\big(1 - \cos(\theta_i - \theta_j)\big) = \frac{Kn}{2}(1 - r^2), \tag{17.12}$$

and its unique global minimum is obtained for $r = 1$, that is, in the phase-synchronized state. By Theorem 17.6, all angles converge to the set of equilibria which are from (17.11) either (i) $r = 0$, (ii) $r > 0$ and in-phase with the order parameter $\theta_i = \psi$, or (iii) $r > 0$ and out-of-phase with the order parameter $\theta_i = \psi + k\pi$ for $k \in \mathbb{Z} \setminus \{0\}$ for all $i \in \{1, \ldots, n\}$. In the latter case, any infinitesimal deviation from an out-of-phase equilibrium causes the potential (17.12) to decrease, that is, the out-of-phase equilibria are unstable. Likewise, the equilibria with $r = 0$ correspond to the global maxima of the potential (17.12), and any infinitesimal deviation from these equilibria causes the potential (17.12) to decrease. It follows that, from almost all initial conditions[1], the oscillators converge to phase-synchronized equilibria $\theta_i = \psi$ for all $i \in \{1, \ldots, n\}$. ∎

17.2.3 Phase balancing

Applications in neuroscience, vehicle coordination, and central pattern generators for robotic locomotion motivate the study of coherent behaviors with synchronized frequencies where the phases are not synchronized, but rather dispersed in appropriate patterns. While the phase-synchronized state can be characterized by the order parameter r achieving its maximal (unit) magnitude, we say that a solution $\theta : \mathbb{R}_{\geq 0} \to \mathbb{T}^n$ to the coupled oscillator model (17.1) achieves *phase balancing* if all phases θ_i asymptotically converge to the set

$$\big\{\theta \in \mathbb{T}^n \mid r(\theta) = \big|\sum\nolimits_{j=1}^{n} e^{i\theta_j}/n\big| = 0\big\},$$

that is, asymptotically the oscillators are uniformly distributed over the unit circle $\mathbb{S}^1$ so that their centroid converges to the origin.

---

[1]To be precise further analysis is needed. A linearization of the Kuramoto model (17.11) at the unstable out-of-pase equilibria yields that these are exponentially unstable. The region of attraction (the so-called stable manifold) of such exponentially unstable equilibria is known to be a zero measure set (Potrie and Monzón, 2009, Proposition 4.1).

For a complete homogeneous graph with coupling strength $a_{ij} = K/n$, i.e., for the Kuramoto model (17.2), we have a remarkable identity between the magnitude of the order parameter r and the potential function $U(\theta)$

$$U(\theta) = \frac{Kn}{2}\left(1 - r^2\right). \tag{17.13}$$

(We ask the reader to establish this identity in Exercise E17.5.) For the complete graph, the correspondence (17.13) shows that the global minimum of the potential function $U(\theta) = 0$ (for $r = 1$) corresponds to phase-synchronization and the global maximum $U(\theta) = Kn/2$ (for $r = 0$) corresponds to phase balancing. This motivates the following gradient ascent dynamics to reach phase balancing:

$$\dot{\theta} = +\frac{\partial U(\theta)}{\partial \theta}^{\mathsf{T}}, \qquad \text{or, equivalently,} \qquad \dot{\theta}_i = \sum_{j=1}^{n} a_{ij}\sin(\theta_i - \theta_j). \tag{17.14}$$

Theorem 17.8 (Phase balancing). *Consider the coupled oscillator model* (17.14) *with a connected, undirected, and weighted graph* $G = (\{1, \ldots, n\}, E, A)$. *Then*

(i) *Global convergence: For all initial conditions* $\theta(0) \in \mathbb{T}^n$, *the phases* $\theta_i(t)$ *converge to the set of critical points* $\{\theta \in \mathbb{T}^n \mid \partial U(\theta)/\partial \theta = \mathbb{0}_n^{\mathsf{T}}\}$; *and*

(ii) *Local stability: For a complete graph with uniform weights* $a_{ij} = K/n$, *phase balancing is the global maximizer of the potential function* (17.13) *and is a locally asymptotically stable equilibrium set.*

Proof. The proof statement (i) is analogous to the proof of statement (i) in Theorem 17.6.

To prove statement (ii), notice that, for a complete graph, the phase balanced set characterized by $r = 0$ achieves the global maximum of the potential $U(\theta) = \frac{Kn}{2}\left(1 - r^2\right)$. By Theorem 14.13, local maxima of the potential are locally asymptotically stable for the gradient ascent dynamics (17.14). ∎

17.3 Synchronization of heterogeneous oscillators

In this section we analyze non-identical oscillators with $\omega_i \neq \omega_j$. As shown in Lemma 17.4, these oscillator networks cannot achieve phase synchronization. On the other hand frequency synchronization with a certain degree of phase cohesiveness can be achieved provided that the natural frequencies satisfy certain bounds relative to the network coupling. We start off with the following necessary conditions.

Lemma 17.9. *Necessary synchronization condition Consider the coupled oscillator model* (17.1) *with graph* $G = (\{1, \ldots, n\}, E, A)$, *frequencies* $\omega \in \mathbb{1}_n^{\perp}$, *and nodal degree* $\deg_i = \sum_{j=1}^{n} a_{ij}$ *for each node* $i \in \{1, \ldots, n\}$. *If there exists a frequency-synchronized solution satisfying the phase cohesiveness* $|\theta_i - \theta_j| \leq \gamma$ *for all* $\{i, j\} \in E$ *and for some* $\gamma \in [0, \pi/2]$, *then the following conditions hold:*

(i) *Absolute bound: For each node* $i \in \{1, \ldots, n\}$,

$$\deg_i \sin(\gamma) \geq |\omega_i|. \tag{17.15}$$

(ii) *Incremental bound: For distinct* $i, j \in \{1, \ldots, n\}$,

$$(\deg_i + \deg_j)\sin(\gamma) \geq |\omega_i - \omega_j|. \tag{17.16}$$

Proof. Statement (i) follows directly from the fact that synchronized solutions must satisfy the equilibrium equation $\dot{\theta}_i = 0$. Since the sinusoidal interaction terms in equation (17.1) are upper bounded by the nodal degree $\deg_i = \sum_{j=1}^{n} a_{ij}$, condition (17.15) is necessary for the existence of an equilibrium.

Statement (ii) follows from the fact that frequency-synchronized solutions must satisfy $\dot{\theta}_i - \dot{\theta}_j = 0$. By analogous arguments, we arrive at the necessary condition (17.16). ∎

17.3.1 Synchronization of heterogeneous oscillators over complete homogeneous graphs

Consider the Kuramoto model over a complete homogeneous graph:

$$\dot{\theta}_i = \omega_i - \frac{K}{n} \sum_{j=1}^{n} \sin(\theta_i - \theta_j), \qquad i \in \{1, \ldots, n\}. \tag{17.17}$$

As discussed in Subsection 17.1.4, the Kuramoto model synchronizes provided that the coupling gain K is larger than some critical value K_{critical}. The necessary condition (17.16) delivers a lower bound for K_{critical} given by

$$K \geq \frac{n}{2(n-1)} \Big(\max_i \omega_i - \min_i \omega_i \Big).$$

Here we evaluated the left-hand side of (17.16) for $a_{ij} = K/n$, for the maximum $\gamma = \pi/2$, and for all distinct $i, j \in \{1, \ldots, n\}$. Perhaps surprisingly, the lower necessary bound (17.3.1) is a factor $1/2$ away from the upper sufficient bound.

Theorem 17.10 (Synchronization test for all-to-all Kuramoto model). *Consider the Kuramoto model (17.17) with natural frequencies $\omega \in \mathbb{1}_n^\perp$ and coupling strength K. Assume*

$$K > K_{\text{critical}} \triangleq \max_i \omega_i - \min_i \omega_i, \tag{17.18}$$

and define the arc lengths $\gamma_{\min} \in [0, \pi/2[$ and $\gamma_{\max} \in]\pi/2, \pi]$ as the unique solutions to $\sin(\gamma_{\min}) = \sin(\gamma_{\max}) = K_{\text{critical}}/K$.

The following statements hold:

  (i) *phase cohesiveness: each solution starting in $\Gamma_{\text{arc}}(\gamma)$, for $\gamma \in [\gamma_{\min}, \gamma_{\max}]$, remains in $\Gamma_{\text{arc}}(\gamma)$ for all times;*

 (ii) *asymptotic phase cohesiveness: each solution starting in $\Gamma_{\text{arc}}(\gamma_{\max})$ asymptotically reaches the set $\overline{\Gamma}_{\text{arc}}(\gamma_{\min})$; and*

(iii) *asymptotic frequency synchronization: each solution starting in $\Gamma_{\text{arc}}(\gamma_{\max})$ achieves frequency synchronization.*

Moreover, the following converse statement is true: Given an interval $[\omega_{\min}, \omega_{\max}]$, the coupling strength K satisfies $K > \omega_{\max} - \omega_{\min}$ if, for all frequencies ω supported on $[\omega_{\min}, \omega_{\max}]$ and for the arc length $\gamma_{\max}$ computed as above, the set $\Gamma_{\text{arc}}(\gamma_{\max})$ is positively invariant.

We illustrate the definitions of $\gamma_{\min}$, $\gamma_{\max}$, and $\Gamma_{\text{arc}}(\gamma)$, for $\gamma \in [\gamma_{\min}, \gamma_{\max}]$ in Figure 17.1.
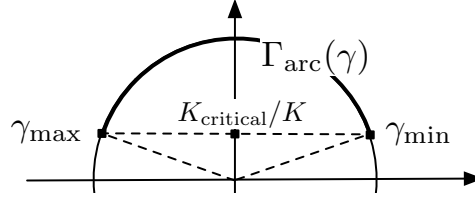
Figure 17.1: Illustrating the definitions of $\gamma_{\min}$, $\gamma_{\max}$, and $\Gamma_{\mathrm{arc}}(\gamma)$, for $\gamma \in [\gamma_{\min}, \gamma_{\max}]$.

Proof. We start with statement (i). Define the function $W : \Gamma_{\mathrm{arc}}(\pi) \to [0, \pi[$ by

$$W(\psi) = \max\{|\psi_i - \psi_j| \mid i, j \in \{1, \ldots, n\}\}.$$

The arc containing all angles ψ has two boundary points: a counterclockwise maximum and a counterclockwise minimum. If $U_{\max}(\psi)$ (resp. $U_{\min}(\psi)$) denotes the set indices of the angles $\psi_1, \ldots, \psi_n$ that are equal to the counterclockwise maximum (resp. the counterclockwise minimum), then

$$W(\psi) = |\psi_{m'} - \psi_{k'}|, \quad \text{for all } m' \in U_{\max}(\psi) \text{ and } k' \in U_{\min}(\psi).$$

We now assume $\theta(0) \in \overline{\Gamma}_{\mathrm{arc}}(\gamma)$, for $\gamma \in [\gamma_{\min}, \gamma_{\max}]$, and aim to show that $\theta(t) \in \overline{\Gamma}_{\mathrm{arc}}(\gamma)$ for all times $t > 0$. By continuity, $\overline{\Gamma}_{\mathrm{arc}}(\gamma)$ is positively invariant if and only if $W(\theta(t))$ does not increase at any time t such that $W(\theta(t)) = \gamma$.

In the next equation we compute the maximum possible amount of infinitesimal increase of $t \mapsto W(\theta(t))$ along system (17.17). Based on the notion of upper Dini derivative and the treatment in Section 14.7, we compute

$$D^+W(\theta(t)) := \limsup_{\Delta t \to 0^+} \frac{W(\theta(t + \Delta t)) - W(\theta(t))}{\Delta t} = \dot{\theta}_m(t) - \dot{\theta}_k(t),$$

where $m \in U_{\max}(\theta(t))$ and $k \in U_{\min}(\theta(t))$ have the property that $\dot{\theta}_m(t) = \max\{\dot{\theta}_{m'}(t) \mid m' \in U_{\max}(\theta(t))\}$ and $\dot{\theta}_k(t) = \min\{\dot{\theta}_{k'}(t) \mid k' \in U_{\min}(\theta(t))\}$. In components

$$D^+W(\theta(t)) = \omega_m - \omega_k - \frac{K}{n} \sum_{j=1}^{n} \Big(\sin(\theta_m(t) - \theta_j(t)) + \sin(\theta_j(t) - \theta_k(t)) \Big).$$

The trigonometric identity $\sin(x) + \sin(y) = 2\sin(\frac{x+y}{2})\cos(\frac{x-y}{2})$ leads to

$$D^+W(\theta(t)) = \omega_m - \omega_k - \frac{K}{n} \sum_{i=1}^{n} \left(2\sin\left(\frac{\theta_m(t) - \theta_k(t)}{2}\right) \cos\left(\frac{\theta_m(t) - \theta_i(t)}{2} - \frac{\theta_i(t) - \theta_k(t)}{2}\right) \right).$$

Measuring angles counterclockwise and modulo 2π, the equality $W(\theta(t)) = \gamma$ implies $\theta_m(t) - \theta_k(t) = \gamma$, $\theta_m(t) - \theta_i(t) \in [0, \gamma]$, and $\theta_i(t) - \theta_k(t) \in [0, \gamma]$. Moreover,

$$\min_\theta \cos\left(\frac{\theta_m - \theta_i}{2} - \frac{\theta_i - \theta_k}{2}\right) = \cos\left(\max_\theta \left|\frac{\theta_m - \theta_i}{2} - \frac{\theta_i - \theta_k}{2}\right|\right) = \cos(\gamma/2),$$

so that

$$D^+ W(\theta(t)) \leq \omega_m - \omega_k - \frac{K}{n} \sum_{i=1}^{n} \left(2 \sin\left(\frac{\gamma}{2}\right) \cos\left(\frac{\gamma}{2}\right)\right).$$

Applying the reverse identity $2 \sin(x) \cos(y) = \sin(x - y) + \sin(x + y)$, we obtain

$$D^+ W(\theta(t)) \leq \omega_m - \omega_k - \frac{K}{n} \sum_{i=1}^{n} \sin(\gamma) \leq \left(\max_i \omega_i - \min_i \omega_i\right) - K \sin(\gamma).$$

Hence, the $W(\theta(t))$ does not increase at all t such that $W(\theta(t)) = \gamma$ if $K \sin(\gamma) \geq K_{\text{critical}} = \max_i \omega_i - \min_i \omega_i$.

Given the structure of the level sets of $\gamma \mapsto K \sin(\gamma)$, there exists an open interval of arc lengths $\gamma \in [0, \pi]$ satisfying $K \sin(\gamma) \geq \max_i \omega_i - \min_i \omega_i$ if and only if equation (17.18) is true with the strict equality sign at $\gamma^* = \pi/2$, that is, if $K > K_{\text{critical}}$. Additionally, if $K > K_{\text{critical}}$, there exists a unique $\gamma_{\min} \in [0, \pi/2[$ and a unique $\gamma_{\max} \in]\pi/2, \pi]$ that satisfy equation (17.18) with the equality sign. In summary, for every $\gamma \in [\gamma_{\min}, \gamma_{\max}]$, if $W(\theta(t)) = \gamma$, then the arc-length $W(\theta(t))$ is non-increasing. This concludes the proof of statement (i).

Moreover, pick $\varepsilon \ll \gamma_{\max} - \gamma_{\min}$. For all $\gamma \in [\gamma_{\min} + \varepsilon, \gamma_{\max} - \varepsilon]$, there exists a positive $\delta(\varepsilon)$ with the property that, if $W(\theta(t)) = \gamma$, then $D^+ W(\theta(t)) \leq -\delta(\varepsilon)$. Hence, each solution $\theta : \mathbb{R}_{\geq 0} \to \mathbb{T}^n$ starting in $\Gamma_{\text{arc}}(\gamma_{\max} - \varepsilon)$ must satisfy $W(\theta(t)) \leq \gamma_{\min} - \varepsilon$ after time at most $(\gamma_{\max} - \gamma_{\min})/\delta(\varepsilon)$. This proves statement (ii).

Regarding statement (iii), we just proved that for every $\theta(0) \in \Gamma_{\text{arc}}(\gamma_{\max})$ and for all $\gamma \in]\gamma_{\min}, \gamma_{\max}]$ there exists a finite time $T \geq 0$ such that $\theta(t) \in \overline{\Gamma}_{\text{arc}}(\gamma)$ for all $t \geq T$ and for some $\gamma < \pi/2$. It follows that $|\theta_i(t) - \theta_j(t)| \leq \gamma < \pi/2$ for all $\{i, j\} \in E$ and for all $t \geq T$. We now invoke Corollary 17.3 to conclude the proof of statement (iii).

The converse statement can be established by noticing that all of the above inequalities and estimates are exact for a bipolar distribution of natural frequencies $\omega_i \in \{\underline{\omega}, \overline{\omega}\}$ for all $i \in \{1, \ldots, n\}$. We refer the reader for these details to the full proof in (Dörfler and Bullo, 2011). ∎

17.3.2 Synchronization of heterogeneous oscillators over weighted undirected graphs

Consider the coupled oscillator model over a weighted undirected graph:

$$\dot{\theta}_i = \omega_i - \sum_{j=1}^{n} a_{ij} \sin(\theta_i - \theta_j), \qquad i \in \{1, \ldots, n\}. \tag{17.19}$$

Adopt the following shorthands:

$$\|\omega\|_{2,\text{ pairs}} = \sqrt{\frac{1}{2} \sum_{i,j=1}^{n} (\omega_i - \omega_j)^2}, \qquad \text{and} \qquad \|\theta\|_{2,\text{ pairs}} = \sqrt{\frac{1}{2} \sum_{i,j=1}^{n} |\theta_i - \theta_j|^2}.$$

Theorem 17.11 (Synchronization test I). *Consider the coupled oscillator model* (17.19) *with frequencies* $\omega \in \mathbb{1}_n^{\perp}$ *defined over a weighted undirected graph with Laplacian matrix L. Assume*

$$\lambda_2(L) > \lambda_{\text{critical}} \triangleq \|\omega\|_{2, \text{ pairs}}, \tag{17.20}$$

and define $\gamma_{\max} \in \;]\pi/2, \pi]$ *and* $\gamma_{\min} \in [0, \pi/2[$ *as the solutions to* $(\pi/2) \cdot \text{sinc}(\gamma_{\max}) = \sin(\gamma_{\min}) = \lambda_{\text{critical}}/\lambda_2(L)$. *The following statements hold:*

(i) *phase cohesiveness: each solution starting in* $\{\theta \in \Gamma_{\text{arc}}(\pi) \mid \|\theta\|_{2, \text{ pairs}} \leq \gamma\}$, *for* $\gamma \in [\gamma_{\min}, \gamma_{\max}]$, *remains in* $\{\theta \in \Gamma_{\text{arc}}(\pi) \mid \|\theta\|_{2, \text{ pairs}} \leq \gamma\}$ *for all times,*

(ii) *asymptotic phase cohesiveness: each solution starting in* $\{\theta \in \Gamma_{\text{arc}}(\pi) \mid \|\theta\|_{2, \text{ pairs}} < \gamma_{\max}\}$ *asymptotically reaches the set* $\{\theta \in \Gamma_{\text{arc}}(\pi) \mid \|\theta\|_{2, \text{ pairs}} \leq \gamma_{\min}\}$; *and*

(iii) *asymptotic frequency synchronization: each solution starting in* $\{\theta \in \Gamma_{\text{arc}}(\pi) \mid \|\theta\|_{2, \text{ pairs}} < \gamma_{\max}\}$ *achieves frequency synchronization.*

The proof of Theorem 17.11 follows the reasoning of the proof of Theorem 17.10 using the quadratic Lyapunov function $\|\theta\|_{2, \text{ pairs}}^2$. The full proof is in (Dörfler and Bullo, 2012, Appendix B).

17.4   Historical notes and further reading

The scientific interest in synchronization of coupled oscillators can be traced back to the work by Huygens (1673) on "an odd kind of sympathy" between coupled pendulum clocks. The model of coupled oscillator which we study was originally proposed by Winfree (1967). For complete interaction graphs, this model is nowadays known as the Kuramoto model due to the work by Kuramoto (1975, 1984). An detailed historical account is given by Strogatz (2000).

The Kuramoto model and its variations appear in the study of biological synchronization phenomena such as pacemaker cells in the heart (Michaels et al., 1987), circadian rhythms (Liu et al., 1997), neuroscience (Varela et al., 2001; Brown et al., 2003; Crook et al., 1997), metabolic synchrony in yeast cell populations (Ghosh et al., 1971), flashing fireflies (Buck, 1988), chirping crickets (Walker, 1969), and rhythmic applause (Néda et al., 2000), among others. The Kuramoto model also appears in physics and chemistry in modeling and analysis of spin glass models (Daido, 1992; Jongen et al., 2001), flavor evolutions of neutrinos (Pantaleone, 1998), and in the analysis of chemical oscillations (Kiss et al., 2002). Some technological applications include deep brain stimulation (Tass, 2003), vehicle coordination (Paley et al., 2007; Sepulchre et al., 2007; Klein et al., 2008), semiconductor lasers (Kozyreff et al., 2000; Hoppensteadt and Izhikevich, 2000), microwave oscillators (York and Compton, 1991), clock synchronization in wireless networks (Simeone et al., 2008), and droop-controlled inverters in microgrids (Simpson-Porco et al., 2013).

Our treatment borrows ideas from (Dörfler and Bullo, 2011, 2014). Recent surveys include (Strogatz, 2000; Acebrón et al., 2005; Arenas et al., 2008; Mauroy et al., 2012; Dörfler and Bullo, 2014). We refer to (Mallada et al., 2016; Gushchin et al., 2016) for a more general treatment with odd-coupling functions and with varying coupling strengths.

17.5 Exercises

E17.1 **Phase cohesiveness and arc length.** Pick $\gamma < 2\pi/3$ and $n \geq 3$. Show the following statement: if $\theta \in \mathbb{T}^n$ satisfies $|\theta_i - \theta_j| \leq \gamma$ for all $i, j \in \{1, \ldots, n\}$, then there exists an arc of length γ containing all angles, that is, $\theta \in \overline{\Gamma}_{\mathrm{arc}}(\gamma)$.

E17.2 **Order parameter and arc length.** Given $n \geq 2$ and $\theta \in \mathbb{T}^n$, the *shortest arc length* $\gamma(\theta)$ is the length of the shortest arc containing all angles, i.e., the smallest $\gamma(\theta)$ such that $\theta \in \overline{\Gamma}_{\mathrm{arc}}(\gamma(\theta))$. Given $\theta \in \mathbb{T}^n$, the *order parameter* is the centroid of $(\theta_1, \ldots, \theta_n)$ understood as points on the unit circle in the complex plane $\mathbb{C}$:

$$r(\theta)\, \mathrm{e}^{\psi(\theta)\mathrm{i}} := \frac{1}{n} \sum_{j=1}^{n} \mathrm{e}^{\theta_j \mathrm{i}}\,.$$

where recall $\mathrm{i} = \sqrt{-1}$. Show that

    (i) if $\gamma(\theta) \in [0, \pi]$, then $r(\theta) \in [\cos(\gamma(\theta)/2), 1]$.

The order parameter magnitude r is known to measure synchronization. Show the following statements:

    (iv) if all oscillators are phase-synchronized, then $r = 1$, and

    (v) if all oscillators are spaced equally on the unit circle (the so-called *splay state*), then $r = 0$.

E17.3 **Order parameter and mean-field dynamics.** Show that the Kuramoto model (17.2) is equivalent to the so-called mean-field model (17.4) with the order parameter r defined in (17.3).

E17.4 **Multiplicity of equilibria in the Kuramoto model.** A common misconception in the literature is that the Kuramoto model has a unique equilibrium set in the phase cohesive set $\{\theta \in \mathbb{T}^n \mid |\theta_i - \theta_j| < \pi/2 \text{ for all } \{i, j\} \in E\}$. Consider now the example of a Kuramoto oscillator network defined over a symmetric cycle graph with identical unit weights and zero natural frequencies. The equilibria are determined by

$$0 = \sin(\theta_i - \theta_{i-1}) + \sin(\theta_i - \theta_{i+1})\,,$$

where $i \in \{1, \ldots, n\}$ and all indices are evaluated modulo n. Show that for $n > 4$ there are at least two disjoint equilibrium sets in the phase cohesive set $\{\theta \in \mathbb{T}^n \mid |\theta_i - \theta_j| < \pi/2 \text{ for all } \{i, j\} \in E\}$.

E17.5 **Potential and order parameter.** Recall $U(\theta) = \sum_{\{i,j\} \in E} a_{ij}\big(1 - \cos(\theta_i - \theta_j)\big)$. Prove $U(\theta) = \frac{Kn}{2}(1 - r^2)$ for a complete homogeneous graph with coupling strength $a_{ij} = K/n$.

E17.6 **Analysis of the two-node case.** Present a complete analysis of a system of two coupled oscillators:

$$\dot{\theta}_1 = \omega_1 - a_{12} \sin(\theta_1 - \theta_2)\,,$$
$$\dot{\theta}_2 = \omega_2 - a_{21} \sin(\theta_2 - \theta_1)\,,$$

where $a_{12} = a_{21}$ and $\omega_1 + \omega_2 = 0$. When do equilibria exist? What are their stability properties and their basins of attraction?

E17.7 **Averaging analysis of coupled oscillators in a semi-circle.** Consider the coupled oscillator model (17.5) with $\theta \in \Gamma_{\mathrm{arc}}(\gamma)$ for some $\gamma < \pi$. Show that the coordinate transformations $x_i = \tan(\theta_i)$, with $x_i \in \mathbb{R}$, gives the averaging system (17.6) with $b_{ij} \geq a_{ij} \cos(\gamma/2)$.

E17.8 **Phase synchronization in spring network.** Consider the spring network from Example #1 in Section 13.3 with identical oscillators, no external torques, and a connected, undirected, and weighted graph:

$$m_i \ddot{\theta}_i + d_i \dot{\theta}_i + \sum_{j=1}^{n} a_{ij} \sin(\theta_i - \theta_j) = 0\,, \quad i \in \{1, \ldots, n\}\,.$$

Prove the phase synchronization result (in Theorem 17.6) for this spring network.

E17.9 **Synchronization on acyclic graphs.** For frequencies $\sum_{i=1}^{n} \omega_i = 0$, consider the coupled oscillator model

$$\dot{\theta}_i = -\sum_{j=1}^{n} a_{ij} \sin(\theta_i - \theta_j).$$

Assume the adjacency matrix A with elements $a_{ij} = a_{ji} \in \{0, 1\}$ is associated to an undirected, connected, and acyclic graph. Show that the following statements are equivalent:

(i) there exists a locally stable frequency-synchronized solution in the set $\{\theta \in \mathbb{T}^n \mid |\theta_i - \theta_j| < \pi/2 \text{ for all } \{i, j\} \in E\}$,

(ii) $\left\| B^{\mathsf{T}} L^{\dagger} \omega \right\|_{\infty} < 1$, where B and L are the network incidence and Laplacian matrices.

Hint: *Follow the derivation in Appendix 9.5.*

E17.10 **Distributed averaging-based PI control for coupled oscillators.** Consider a set of n controllable coupled oscillators governed by the second-order dynamics

$$\dot{\theta}_i = \omega_i, \tag{E17.1a}$$

$$m_i \dot{\omega}_i = -d_i \omega_i - \sum_{j=1}^{n} a_{ij} \sin(\theta_i - \theta_j) + u_i, \tag{E17.1b}$$

where $i \in \{1, \dots, n\}$ is the index set, each oscillator has the state $(\theta_i, \omega_i) \in \mathbb{T}^1 \times \mathbb{R}$, $u_i \in \mathbb{R}$ is a control input to oscillator i, and $m_i > 0$ and $d_i > 0$ are the inertia and damping coefficients. The oscillators are coupled through an undirected, connected, and weighted graph $G = (V, E, A)$ with node set $V = \{1, \dots, n\}$, edge set $E \subset V \times V$, and adjacency matrix $A = A^{\mathsf{T}} \in \mathbb{R}^{n \times n}$. To reject disturbances affecting the oscillators, consider the distributed averaging-based integral controller (see Exercise E6.17)

$$u_i = -q_i, \tag{E17.2a}$$

$$\dot{q}_i = w_i - \sum_{j=1}^{n} b_{ij}(q_i - q_j), \tag{E17.2b}$$

where $q_i \in \mathbb{R}$ is a controller state for each agent $i \in \{1, \dots, n\}$, and the matrix B with elements b_{ij} is the adjacency matrix of an undirected and connected graph. Your tasks are as follows:

(i) characterize the set of equilibria $(\theta^\star, \omega^\star, q^\star)$ of the closed-loop system (E17.1)-(E17.2),

(ii) show that all trajectories converge to the set of equilibria, and

(iii) show that the phase synchronization set $\{\theta \in \mathbb{T}^n \mid \theta_i = \theta_j \text{ for all } i, j \in \{1, \dots, n\}\}$ together with $\omega = q = \mathbb{0}_n$ is an equilibrium and that it is locally asymptotically stable.

Robotic Coordination and Formation Control

In this chapter we present some methods and ideas related to coordination in robotic networks with relative sensing. We discuss rendezvous, flocking, and formation control problems.

18.1 Coordination in relative sensing networks

We consider the following setup for the coordination of n autonomous mobile robots (referred to as agents) in a planar environment:

(i) **Agent dynamics:** We consider a simple and fully actuated agent model: $\dot{p}_i = u_i$, where $p_i \in \mathbb{R}^2$ and $u_i \in \mathbb{R}^2$ are the position and steering control input of agent i.

(ii) **Relative sensing model:** We consider the following sensing model.

- Each agent is equipped with onboard sensors only and has no communication devices.
- The sensing topology is encoded by an undirected and connected graph $G = (V, E)$
- Each agent i can measure the relative position of neighboring agents: $p_i - p_j$ for $\{i, j\} \in E$.

To formalize the relative sensing model, we introduce an arbitrary orientation and labeling $k \in \{1, \ldots, |E|\}$ for each undirected edge $\{i, j\} \in E$. Recall the incidence matrix $B \in \mathbb{R}^{n \times |E|}$ of the associated oriented graph and define the $2n \times 2|E|$ matrix $\hat{B} = B \otimes I_2$ via the Kronecker product (see Exercise E7.21). The *Kronecker product* $A \otimes B$ is the "element-wise" matrix product so that each scalar entry A_{ij} of A is replaced by a block-entry $A_{ij} \cdot B$ in the matrix $A \otimes B$. For example, if B is given by

$$B = \begin{bmatrix} +1 & 0 & 0 & 0 \\ -1 & +1 & -1 & 0 \\ 0 & -1 & 0 & +1 \\ 0 & 0 & +1 & -1 \end{bmatrix}, \text{ then } \hat{B} \text{ is given by } \hat{B} = B \otimes I_2 = \begin{bmatrix} +I_2 & 0 & 0 & 0 \\ -I_2 & +I_2 & -I_2 & 0 \\ 0 & -I_2 & 0 & +I_2 \\ 0 & 0 & +I_2 & -I_2 \end{bmatrix}.$$

With this notation the vector of relative positions is given by $e = \hat{B}^\mathsf{T} p$.
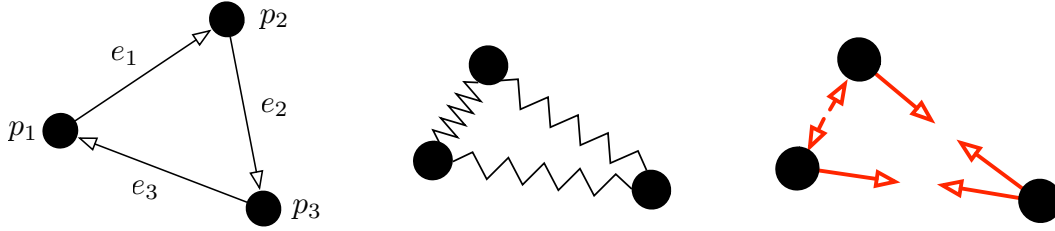
Figure 18.1: A cycle graph with three agents. The first panel shows the agents embedded in the plane $\mathbb{R}^2$ with positions p_i and relative positions e_i. The second panel shows the artificial potentials as springs connecting the robots, and the third panel shows the resulting forces.

(iii) **Geometric objective:** The objective is to achieve desired geometric configuration which can be expressed as a function of relative distances $\|p_i - p_j\|$ for each $\{i, j\} \in E$. Examples include rendezvous ($\|p_i - p_j\| = 0$), collision avoidance ($\|p_i - p_j\| > 0$), and desired relative spacings ($\|p_i - p_j\| = d_{ij} > 0$).

(iv) **Potential-based control:** We specify the geometric objective for each edge $\{i, j\} \in E$ as the minimum of an *artificial potential function* $V_{ij} : D_{ij} \subset \mathbb{R} \to \mathbb{R}_{\geq 0}$. We require the potential functions to be twice continuously differentiable on their domain D_{ij}.

It is instructive to think of $V_{ij}(\|p_i - p_j\|)$ as a spring coupling neighboring agents $\{i, j\} \in E$. The resulting spring forces acting on agents i and j are $f_{ij}(p_i - p_j) = -\frac{\partial}{\partial p_i} V_{ij}(\|p_i - p_j\|)$ and $f_{ji}(p_i - p_j) = -f_{ij}(p_i - p_j) = -\frac{\partial}{\partial p_j} V_{ij}(\|p_i - p_j\|)$; see Figure 18.1 for an illustration. The overall network potential function is then

$$V(p) = \sum_{\{i,j\}\in E} V_{ij}(\|p_i - p_j\|) \,.$$

We design the associated *gradient descent control* law as

$$\dot{p}_i = u_i = -\frac{\partial V(p)}{\partial p_i} = -\sum_{\{i,j\}\in E} \frac{\partial}{\partial p_i} V_{ij}(\|p_i - p_j\|) = \sum_{\{i,j\}\in E} f_{ij}(p_i - p_j) \,, \quad i \in \{1, \dots, n\} \,.$$

In vector form the control reads as the negative gradient flow

$$\dot{p} = u = -\frac{\partial V(p)}{\partial p}^{\mathsf{T}} = \hat{B} \cdot \operatorname{diag}(\{f_{ij}\}_{\{i,j\}\in E}) \circ \hat{B}^{\mathsf{T}} p \,. \tag{18.1}$$

The closed-loop relative sensing network (18.1) is illustrated in Figure 18.2.

Controllers based on artificial potential functions induce a lot of structure in the closed-loop system. Recall the set of 2-dimensional orthogonal matrices $\mathrm{O}(2) = \{R \in \mathbb{R}^2 \mid RR^{\mathsf{T}} = I_2\}$, introduced in Exercise E2.14, as the set of 2-dimensional rotations and reflections.

Lemma 18.1 (Symmetries of relative sensing networks). *Consider the closed-loop relative sensing network* (18.1) *with an undirected and connected graph $G = (V, E)$. For every initial condition $p_0 \in \mathbb{R}^{2n}$, we have that*
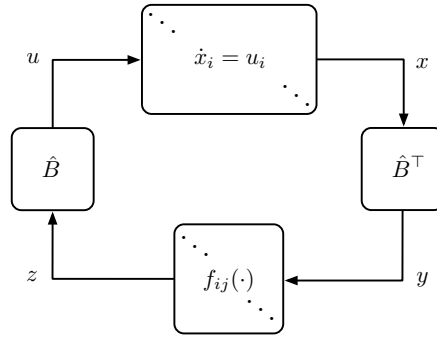
Figure 18.2: Closed-loop diagram of the relative sensing network (18.1).

(i) *the center of mass is stationary:* $\mathrm{average}(p(t)) = \mathrm{average}(p_0)$ *for all* $t \geq 0$; *and*

(ii) *the closed-loop* $\dot{p} = -\frac{\partial V(p)}{\partial p}^{\mathsf{T}}$ *is invariant under rigid body transformations: if* $\xi_i = Rp_i + q$, *where* $R \in \mathrm{O}(2)$ *and* $q \in \mathbb{R}^2$ *is a translation vector, then* $\dot{\xi} = -\frac{\partial V(\xi)}{\partial \xi}^{\mathsf{T}}$.

Proof. Regarding statement (i), since $\sum_{i=1}^{n} \dot{p}_i = 0$, it follows that $\sum_{i=1}^{n} p_i(t) = \sum_{i=1}^{n} p_{i0}$.

Regarding statement (ii), first, notice that potential function is invariant under translations since $V(p) = V(p + \mathbb{1}_n \otimes q)$ for any translation $q \in \mathbb{R}^2$. Second, notice that the potential function is invariant under rotations and reflections since $V_{ij}(\|R(p_i - p_j)\|) = V_{ij}(\|p_i - p_j\|)$ and thus $V(\hat{R}p) = V(p)$ where $\hat{R} = I_n \otimes R$. From the chain rule we obtain $\frac{\partial}{\partial p}V(\hat{R}p)\hat{R} = \frac{\partial}{\partial p}V(p)$ or $\frac{\partial}{\partial p}V(\hat{R}p) = \frac{\partial}{\partial p}V(p)\hat{R}^{\mathsf{T}}$. By combining these insights when changing coordinates via $\xi_i = Rp_i + q$ (or $\xi = \hat{R}p + \mathbb{1}_n \otimes q$), we find that

$$\dot{\xi} = \hat{R}\dot{p} = -\hat{R}\frac{\partial V(p)}{\partial p}^{\mathsf{T}} = -\left(\frac{\partial V(p)}{\partial p}\hat{R}^{\mathsf{T}}\right)^{\mathsf{T}} = -\frac{\partial V(\hat{R}p)}{\partial p}^{\mathsf{T}} = -\frac{\partial V(\xi)}{\partial \xi}^{\mathsf{T}}.$$

∎

Example 18.2 (The linear-quadratic rendezvous problem). An undirected consensus system is a relative sensing network coordination problem where the objective is *rendezvous*: $p_i = p_j$ for all $\{i, j\} \in E$. For each edge $\{i, j\} \in E$ consider the artificial potential $V_{ij} : \mathbb{R}^{2n} \to \mathbb{R}_{\geq 0}$ which has a minimum at the desired objective. For example, for the quadratic potential function

$$V_{ij}(p_i - p_j) = \frac{1}{2}a_{ij}\|p_i - p_j\|_2^2,$$

the overall potential function is obtained as the Laplacian potential $V(p) = \frac{1}{2}p^{\mathsf{T}}\hat{L}p$, where $\hat{L} = L \otimes I_2$. The resulting gradient descent control law gives rise to the linear Laplacian flow

$$\dot{p}_i = u_i = -\frac{\partial}{\partial p_i}V(p) = -\sum_{\{i,j\}\in E} a_{ij}(p_i - p_j). \tag{18.2}$$

So far, we analyzed the consensus problem (18.2) using matrix theory and exploiting the linearity of the problem. In the following, we introduce numerous tools that will allow us to analyze nonlinear consensus-type interactions and more general nonlinear dynamical systems. •

18.2 A nonlinear rendezvous problem

Consider the nonlinear rendezvous system

$$\dot{p}_i = f_i(p) = - \sum_{\{i,j\} \in E} g_{ij}(p_i - p_j)\,, \tag{18.3}$$

where (for each $\{i,j\} \in E$) $g_{ij} = g_{ji}$ is a continuously differentiable, strictly increasing, and anti-symmetric function satisfying $e \cdot g_{ij}(e) \geq 0$ and $g_{ij}(e) = 0$ if and only if $e = 0$. Notice that the linearization of the system around the consensus subspace may be zero and thus not very informative, for example, when $g_{ij}(e) = \|e\|^2 e$. The nonlinear rendezvous system (18.3) can be written as a negative gradient flow:

$$\dot{p}_i = -\frac{\partial}{\partial p_i} V(p) = - \sum_{j=1}^{n} \frac{\partial}{\partial p_i} V_{ij}(\|p_i - p_j\|)\,.$$

with the associated edge potential function $V_{ij}(\|p_i - p_j\|) = \int_0^{\|p_i - p_j\|} g_{ij}(\chi)\, d\chi$.

Theorem 18.3 (Nonlinear rendezvous). *Consider the nonlinear rendezvous system (18.3) with an undirected and connected graph $G = (V, E)$. Assume that the associated edge potential functions $V_{ij}(\|p_i - p_j\|) = \int_0^{\|p_i - p_j\|} g_{ij}(\chi)\, d\chi$ are radially unbounded. For every initial condition $p_0 \in \mathbb{R}^{2n}$, we have that*

(i) the center of mass is stationary: $\operatorname{average}(p(t)) = \operatorname{average}(p_0)$ *for all* $t \geq 0$; *and*

(ii) $\lim_{t \to \infty} p(t) = \mathbb{1}_n \otimes \operatorname{average}(p_0)$.

Proof. Note that the nonlinear rendezvous system (18.3) is the negative gradient system defined by the network potential function

$$V(p) = \sum_{\{i,j\} \in E} V_{ij}(\|p_i - p_j\|)\,.$$

Recall from Lemma 18.1 that the center of mass is stationary, and observe that the function $V(p)$ is radially unbounded with exception of the direction $\operatorname{span}(\mathbb{1}_{2n})$ associated with a translation of the stationary center of mass. Thus, for every initial condition $p_0 \in \mathbb{R}^{2n}$, the set of points (with fixed center of mass)

$$\{p \in \mathbb{R}^{2n} \mid \operatorname{average}(p) = \operatorname{average}(p_0)\,, V(p) \leq V(p_0)\}$$

defines a compact set. By the Krasovskiĭ-LaSalle Invariance Principle in Theorem 14.7, each solution converges to the largest invariant set contained in

$$\left\{ p \in \mathbb{R}^{2n} \;\middle|\; \operatorname{average}(p(t)) = \operatorname{average}(p_0)\,, V(p) \leq V(p_0)\,, \frac{\partial V(p)}{\partial p} = \mathbb{0}_n^{\mathsf{T}} \right\}\,.$$

It follows that the only positive limit set is the set of equilibria: $\lim_{t \to \infty} p(t) \in \operatorname{span}(\mathbb{1}_n \otimes \operatorname{average}(p_0))$. ∎

18.3 Flocking and Formation Control

In flocking control, the objective is that the robots should mimic the behavior of fish schools and bird flocks and attain a pre-scribed formation defined by a set of distance constraints. Given an undirected graph $G(V, E)$ and a distance constraint d_{ij} for every edge $\{i, j\} \in E$, a formation is defined by the set

$$\mathcal{F} = \{p \in \mathbb{R}^{2n} \mid \|p_i - p_j\|_2 = d_{ij} \quad \forall \{i, j\} \in E\}.$$

We embed the graph G into the plane $\mathbb{R}^2$ by assigning to each node i a location $p_i \in \mathbb{R}^2$. We refer to the pair (G, p) as a *framework*, and we denote the set of frameworks $(G, \mathcal{F})$ as the *target formation*. A target formation is a realization of $\mathcal{F}$ in the configuration space $\mathbb{R}^2$. A triangular example is shown in Figure 18.3.
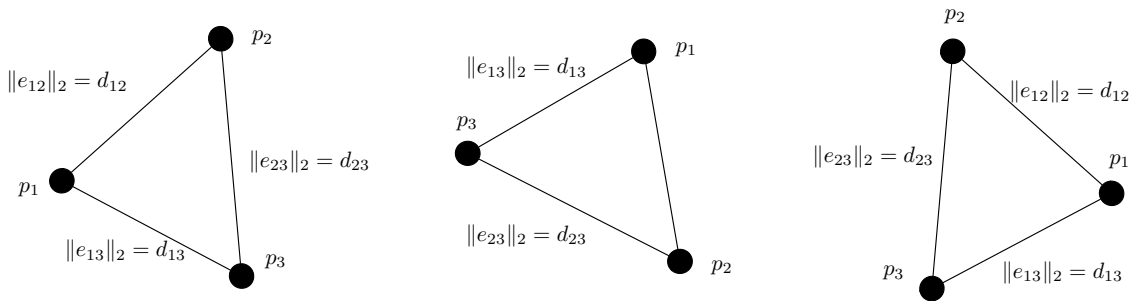
Figure 18.3: A triangular formation specified by the distance constraints d_{12}, d_{13}, and d_{23}. The left subfigure shows one possible target formation, the middle subfigure shows a rotation of this target formation, and the right subfigure shows a "flip" of the left target formation. All of these triangles satisfy the specified distance constraints and are elements of $\mathcal{F}$.

We make the following three observations on the geometry of the target formation:

- To be non-empty, the formation $\mathcal{F}$ has to be *realizable* in the plane. For example, for the triangular formation in Figure 18.3 the distance constraints d_{ij} need to satisfy the triangle inequalities:

$$d_{12} \leq d_{13} + d_{23}, \quad d_{23} \leq d_{12} + d_{13}, \quad d_{13} \leq d_{12} + d_{23}.$$

- A framework (G, p) with $p \in \mathcal{F}$ is *invariant under rigid body transformations*, that is, rotation or translation, as seen in Figure 18.3. Hence, the formation $\mathcal{F}$ is a set of at least of "dimension 3".

- The formation $\mathcal{F}$ may consist of *multiple disconnected components*. For instance, for the triangular example in Figure 18.3 there is no continuous deformation from the left framework to the right "flipped" framework, even though both are target formations. In the state space $\mathbb{R}^6$, this absence of a continuous deformation corresponds to two disconnected components of the set $\mathcal{F}$.

To steer the agents towards the target formation, consider an artificial potential function for each edge $\{i, j\} \in E$ which mimics the Hookean potential of a spring with rest length d_{ij}:

$$V_{ij}(\|p_i - p_j\|) = \frac{1}{2} (\|p_i - p_j\|_2 - d_{ij})^2.$$

Since this potential function is not differentiable, we choose the modified potential function

$$V_{ij}(\|p_i - p_j\|) = \frac{1}{2}\left(\|p_i - p_j\|_2^2 - d_{ij}^2\right)^2 . \tag{18.4}$$

The resulting closed loop under the gradient descent control law $u = -\frac{\partial}{\partial p}V(p)$ is given by

$$\dot{p}_i = u_i = -\frac{\partial}{\partial p_i}V(p) = -2\sum_{\{i,j\}\in E}\left(\|p_i - p_j\|_2^2 - d_{ij}^2\right)\cdot(p_i - p_j). \tag{18.5}$$

Observe that the set of equilibria of the closed loop (18.5) is the set of critical points of $V(p)$ which is a strict super-set of the target formation $\mathcal{F}$. For example, it includes the set of points when two neighbors are collocated: $p_i = p_j$ for $\{i,j\} \in E$. In the following, we show convergence to the equilibrium set.

Theorem 18.4 (Flocking). *Consider the nonlinear flocking system* (18.5) *with an undirected and connected graph $G = (V, E)$ and a realizable formation $\mathcal{F}$. For every initial condition $p_0 \in \mathbb{R}^n$, we have that*

- *the center of mass is stationary:* $\mathrm{average}(p(t)) = \mathrm{average}(p_0)$ *for all $t \geq 0$; and*

- *the agents asymptotically converge to the set of critical points of the potential function.*

Proof. As in the proof of Theorem 18.3, the center of mass is stationary and the potential is non-increasing:

$$\dot{V}(p) = -\left\|\frac{\partial V(p)}{\partial p}^{\mathsf{T}}\right\|^2 \leq 0 .$$

Observe further that for a fixed initial center of mass, the sublevel sets of $V(p)$ form a compact set. By the Krasovskiĭ-LaSalle Invariance Principle in Theorem 14.7, $p(t)$ converges to the largest invariant set contained in

$$\left\{p \in \mathbb{R}^{2n} \;\middle|\; \mathrm{average}(p(t)) = \mathrm{average}(p_0), V(p) \leq V(p_0), \frac{\partial V(p)}{\partial p} = \mathbb{0}_n^{\mathsf{T}}\right\} .$$

It follows that the positive limit set is the set of critical points of the potential function. ∎

Observe that Theorem 18.4 guarantees at most convergence to the critical points of the potential function. Depending on the problem scenario of interest, we still have to investigate which of these critical points are locally asymptotically stable or unstable on a case-by-case basis; see Exercise E18.3 for an application to a linear formation and Section 18.4 for a more general analysis.

The above Theorem 18.4 also holds true for non-smooth potential functions $V_{ij} :]0, \infty[\to \mathbb{R}$ that satisfy

(P1) regularity: $V_{ij}(\xi)$ is defined and twice continuously-differentiable on $]0, \infty[$;

(P2) distance specification: $f_{ij}(\xi) = \frac{\partial}{\partial \xi}V_{ij}(\xi) = 0$ if and only if $\xi = d_{ij}$;

(P3) mutual attractivity: $f_{ij}(\xi) = \frac{\partial}{\partial \xi}V_{ij}(\xi)$ is strictly monotone increasing; and

(P4) collision avoidance: $\lim_{\xi \to 0} V_{ij}(\xi) = \infty$.

(a) Artificial potential functions

(b) Induced artificial spring forces
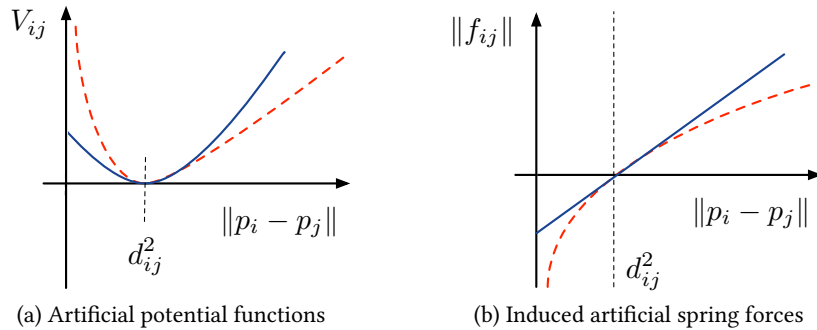
Figure 18.4: Illustration of the quadratic potential function (18.4) (blue solid plot) and a logarithmic barrier potential function (red dashed plot) that approaches ∞ as two neighboring agents become collocated

An illustration of possible potential functions can be found in Figure 18.4. These potential functions can also be modified to include input constraints; see Exercise E18.1.

Theorem 18.5 (Flocking with collision avoidance). *Consider the negative gradient flow* (18.1) *with an undirected and connected graph* $G = (V, E)$, *a realizable formation* $\mathcal{F}$, *and artificial potential functions satisfying* (P1)–(P1). *Then, for every initial condition* $p_0 \in \mathbb{R}^{2n}$ *satisfying* $p_i(0) \neq p_j(0)$ *for all* $\{i, j\} \in E$,

(i) *the solution to the non-smooth dynamical system exists for all times* $t \geq 0$;

(ii) *the center of mass* $\mathrm{average}(p(t)) = \mathrm{average}(p(0))$ *is stationary for all* $t \geq 0$;

(iii) *neighboring robots will not collide, that is,* $p_i(t) \neq p_j(t)$ *for all* $\{i, j\} \in E$ *and for all* $t \geq 0$; *and*

(iv) *the agents asymptotically converge to the set of critical points of the potential function.*

Proof. The proof of Theorem 18.5 is identical to that of Theorem 18.4 after realizing that, for initial conditions satisfying $p_i(0) \neq p_j(0)$ for all $\{i, j\} \in E$, the dynamics are confined to the compact and forward invariant set

$$\left\{ p \in \mathbb{R}^{2n} \mid \mathrm{average}(p(t)) = \mathrm{average}(p_0), V(p) \leq V(p_0) \right\}.$$

Within this set, the dynamics (18.5) are twice continuously differentiable and collisions are avoided. ∎

At this point we should ask ourselves the following three questions:

(i) Do the agents actually stop, that is, does there exist an $p_\infty \in \mathbb{R}^n$ so that $\lim_{t\to\infty} p(t) = p_\infty$?

(ii) The formation $\mathcal{F}$ is a subset of the set of critical points of the potential function. How can we render this particular subset stable (amongst possible other critical points)? What are the other critical points?

(iii) Does our specification of the target formation make sense? For example, in Figure 18.5 the target formation can be infinitesimally deformed, such that the resulting geometric configurations are not congruent.

The answers to all this question is tied to a graph-theoretic concept called *rigidity*.
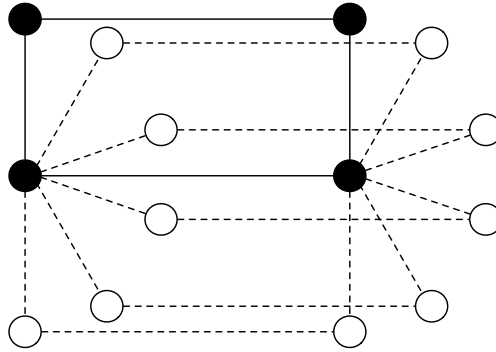
Figure 18.5: A rectangular target formation among four robots, which is specified by four distance constraints. The initial geometric configuration (solid circles) can be continuously deformed such that the resulting geometric configuration is not congruent anymore. All of the displayed configurations are part of the target formation set and satisfy the distance constraints, even the case when the agents are collinear.

18.4   Rigidity and stability of the target formation

To introduce the notion of graph rigidity, we view the undirected graph $G = (V, E)$ as a framework (G, p) embedded in the plane $\mathbb{R}^2$. Given a framework (G, p), we define the *rigidity function* $r_G(p)$ as

$$r_G : \mathbb{R}^{2n} \to \mathbb{R}^{|E|} \quad , \quad r_G(p) \triangleq \left[\dots , \|p_i - p_j\|_2^2 , \dots \right]^{\mathsf{T}},$$

where each component in $r_G(p)$ corresponds the length of the relative position $p_i - p_j$ for $\{i, j\} \in E$.

Definition 18.6 (Rigidity). *Given an undirected graph $G(V, E)$ and $p \in \mathbb{R}^{2n}$, the framework (G, p) is said to be* rigid *if there is an open neighbourhood $\mathcal{U}$ of p such that if $q \in \mathcal{U}$ and $r_G(p) = r_G(q)$, then (G, p) is congruent to (G, q).*

An example of a rigid and non-rigid framework is shown in Figure 18.6.

(a) A flexible framework
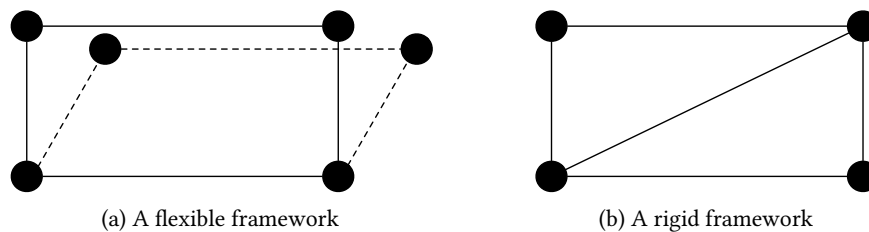
(b) A rigid framework

Figure 18.6: The framework in Figure 18.6a is not rigid since a slight perturbation of the upper two points of the framework results in a framework that is not congruent to the original one although their rigidity functions coincide. If an additional cross link is added to the framework as in Figure 18.6b, small perturbations that do not change the rigidity function result in a congruent framework. Thus, the framework in Figure 18.6b is rigid.

Although rigidity is a very intuitive concept, its definition does not provide an easily verifiable condition, especially if one is interested in finding the exact neighbourhood $\mathcal{U}$ where the framework is rigid. The

following "linearized rigidity concept" offers an easily checkable algebraic condition. The idea is to allow an infinitesimally small perturbation ∂p of the framework (G, p) while keeping the rigidity function constant up to first order. Then the first order Taylor approximation of the rigidity function r_G about p is

$$r_G(p + \partial p) = r_G(p) + \frac{\partial r_G(p)}{\partial p} \, \partial p + \mathcal{O}(\|\partial p\|^2) \, .$$

The rigidity function then remains constant up to first order if $\partial p \in \text{kernel} \left(\frac{\partial r_G(p)}{\partial p} \right)$. The matrix $\frac{\partial r_G(p)}{\partial p} \in \mathbb{R}^{|E| \times 2n}$ is called the *rigidity matrix* of the graph G. If the perturbation ∂p is a rigid body motion, that is a translation and rotation of the framework, then, by Definition 18.6, the framework is still rigid. Thus, the dimension of the kernel of the rigidity matrix is at least 3. The idea that rigidity is preserved under infinitesimal perturbations motivates the following definition of infinitesimal rigidity.

Definition 18.7 (Infinitesimal rigidity). *Given an undirected graph $G(V, E)$ and $p \in \mathbb{R}^{2n}$, the framework (G, p) is said to be infinitesimally rigid if* $\dim \left(\text{kernel} \left(\frac{\partial r_G(p)}{\partial p} \right) \right) = 3$ *or equivalently if* $\text{rank} \left(\frac{\partial r_G(p)}{\partial p} \right) = 2n - 3$.

If a framework is infinitesimally rigid, then it is also rigid but the converse is not necessarily true (Asimow and Roth, 1979). Also note that an infinitesimally rigid framework must have at least $2n - 3$ edges E. If it has exactly $2n - 3$ edges, then we call it a *minimally rigid* framework. Finally, if (G, p) is infinitesimally rigid at p, so is (G, p') for p' in an open neighborhood of p. Thus, infinitesimal rigidity is a *generic property* that depends almost only on the graph G and not on the specific point $p \in \mathbb{R}^{2n}$. Throughout the literature (infinitesimally, minimally) rigid frameworks are often denoted as (infinitesimally, minimally) rigid graphs.

Example 18.8 (Rigidity and infinitesimal rigidity of triangular formation). Consider the triangular framework in Figure 18.7a and the collapsed triangular framework in Figure 18.7b which are both embeddings of the same triangular graph. The rigidity function for both frameworks is given by

$$r_G(p) = \begin{bmatrix} \|p_2 - p_1\|^2 \\ \|p_3 - p_2\|^2 \\ \|p_1 - p_3\|^2 \end{bmatrix} \, .$$

(a) A rigid and infinitesimally rigid framework (triangle inequalities are strict)

(b) A rigid but not infinitesimally rigid framework (triangle inequalities are equalities)
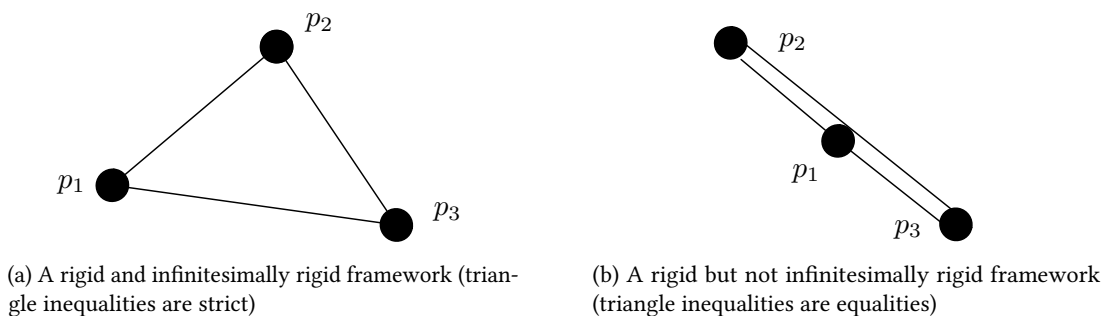
Figure 18.7: Infinitesimal rigidity properties of a framework with three points

Both frameworks are rigid but only the left framework is infinitesimally rigid. To see this, consider the rigidity matrix

$$\frac{\partial r_G(p)}{\partial p} = 2 \begin{bmatrix} p_1^\mathsf{T} - p_2^\mathsf{T} & p_2^\mathsf{T} - p_1^\mathsf{T} & \mathbb{0}_2^\mathsf{T} \\ \mathbb{0}_2^\mathsf{T} & p_2^\mathsf{T} - p_3^\mathsf{T} & p_3^\mathsf{T} - p_2^\mathsf{T} \\ p_1^\mathsf{T} - p_3^\mathsf{T} & \mathbb{0}_2^\mathsf{T} & p_3^\mathsf{T} - p_1^\mathsf{T} \end{bmatrix} .$$

The rank of the rigidity matrix at a collinear point is $2 < 2n - 3$. Hence, the collapsed triangle in Figure 18.7b is not infinitesimally rigid. All non-collinear realizations are infinitesimally and minimally rigid. Hence, the triangular framework in Figure 18.7a is generically minimally rigid (for almost every $p \in \mathbb{R}^6$).●

Minimally rigid graphs can be constructed by adding a new node with two undirected edges to an existing minimally rigid graph; see Figure 18.8. This construction is known under the name *Henneberg* sequence.
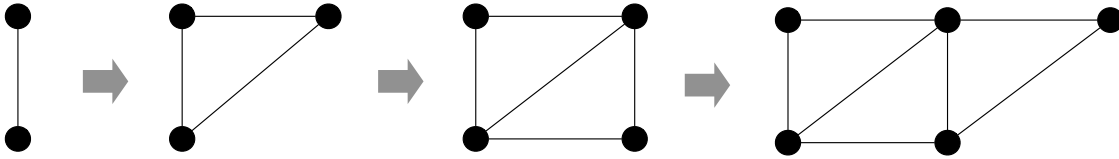
Figure 18.8: Construction of a minimally rigid graph by means of Henneberg sequence

The flocking result in Theorem 18.4 identifies the critical points of the potential function as the positive limit set. For minimally rigid graphs, we can perform a more insightful stability analysis. To do so, we first reformulate the formation control problem in the coordinates of the relative positions $e = \hat{B}^\mathsf{T} p$. The rigidity function can be conveniently rewritten in terms of the relative positions $e_{ij} = p_i - p_j$ for every edge $\{i, j\} \in E$:

$$r_G : B^\mathsf{T} \mathbb{R}^{2n} \to \mathbb{R}^{|E|} \quad , \quad r_G(e) = \left[\ldots, \|e_{ij}\|_2^2, \ldots \right]^\mathsf{T}.$$

The rigidity matrix is then obtained in terms of the relative positions as

$$R(e) \triangleq \frac{\partial r_G(e)}{\partial p} = \frac{\partial r_G(e)}{\partial e} \cdot \frac{\partial e}{\partial p} = 2 \operatorname{diag}(e^\mathsf{T}) \cdot \hat{B}^\mathsf{T} .$$

Consider the shorthand $r_G(e) - \mathbf{d} = \left[\ldots, \|p_i - p_j\|_2^2 - d_{ij}^2, \ldots \right]^\mathsf{T}$. Then the closed-loop formation control equations (18.5) can be reformulated in terms of relative positions as

$$\dot{e} = \hat{B}^\mathsf{T} \dot{p} = \hat{B}^\mathsf{T} u = -2 \hat{B}^\mathsf{T} \hat{B} \operatorname{diag}(e)(r_G(e) - \mathbf{d}) = -\hat{B}^\mathsf{T} R(e)^\mathsf{T}(r_G(e) - \mathbf{d}) . \tag{18.6}$$

The associated initial condition $e_0 = \hat{B}^\mathsf{T} p_0$ is a vector in $\operatorname{image}(\hat{B}^\mathsf{T})$.

Theorem 18.9 (Stability of minimally rigid formations (Dörfler and Francis, 2010)). *Consider the nonlinear flocking system (18.5) with an undirected and connected graph $G = (V, E)$ and a realizable and minimally rigid formation $\mathcal{F}$. For every initial condition $p_0 \in \mathbb{R}^n$, we have that*

- *the center of mass is stationary:* $\operatorname{average}(p(t)) = \operatorname{average}(p_0)$ *for all $t \geq 0$;*

- *the agents asymptotically converge to the set*

$$W_{p_0} = \{ p \in \mathbb{R}^{2n} \mid \text{average}(p) = \text{average}(p_0) \,, V(p) \le V(p_0) \,, \| R(e)^{\mathsf{T}} \, [r_G(e) - \mathbf{d}] \|_2 = \mathbb{0}_{|E|} \} \,.$$

In particular, the limit set W_{p_0} is a union of realizations of the target formation (G, p) with $p \in W_{p_0} \cap \mathcal{F}$ and the set of points $p \in W_{p_0}$ where the framework (G, p) is not infinitesimally rigid; and

- *For every $p_0 \in \mathbb{R}^{2n}$ such that the framework (G, p) is minimally rigid for all p in the set*

$$\{ p \in \mathbb{R}^{2n} \mid \text{average}(p) = \text{average}(p_0) \,, V(p) \le V(p_0) \} \,,$$

the agents converge exponentially fast to a stationary target formation (G, p_∞) with $p_\infty \in W_{p_0} \cap \mathcal{F}$.

Proof. Consider the potential function (18.4), which reads in e-coordinates as

$$V(e) = \frac{1}{2} \| r_G(e) - \mathbf{d} \|^2 \,, \tag{18.7}$$

In the space of relative positions the target formation set $\hat{B}^{\mathsf{T}} \mathcal{F}$ is compact since the translational invariance is removed. Also the sublevel sets of $V(e)$ are compact, and the derivative along the trajectories of (18.6) is

$$\frac{\partial V(e)}{\partial e} \dot{e} = -2 \, [r_G(e) - \mathbf{d}]^{\mathsf{T}} \, \text{diag}(e^{\mathsf{T}}) \hat{B}^{\mathsf{T}} R(e) [r_G(e) - \mathbf{d}] = -[r_G(e) - \mathbf{d}]^{\mathsf{T}} R(e) R(e)^{\mathsf{T}} [r_G(e) - \mathbf{d}] \le 0 \,.$$

Notice that $V(e(t))$ is non-increasing, and for every $c \ge 0$ the sublevel set

$$\Omega(c) := \{ e \in \text{Im}(\hat{B}^{\mathsf{T}}) \mid V(e) \le c \}$$

is forward invariant. By the Krasovskiĭ-LaSalle Invariance Principle, for every initial condition $e_0 \in \text{image}(\hat{B}^{\mathsf{T}})$ the associated solution of (18.6) converges to the largest invariant set in

$$W_{e_0} = \{ e \in \text{image}(\hat{B}^{\mathsf{T}}) \mid V(e) \le V(e_0) \,, \| R(e)^{\mathsf{T}} \, [r_G(e) - \mathbf{d}] \|_2 = \mathbb{0}_{|E|} \} \,.$$

In particular, the limit set W_{e_0} includes (i) realizations of the target formation (G, p) with $p \in W_{p_0} \cap \mathcal{F}$, $e = \hat{B}^{\mathsf{T}} p$, and $[r_G(e) - \mathbf{d}] = \mathbb{0}_{|E|}$, and (ii) the set of points $e \in W_{e_0}$ where the rigidity matrix $R(e)^{\mathsf{T}} \in \mathbb{R}^{n \times |E|}$ loses rank corresponding to points $p \in W_{p_0}$ where the framework (G, p) is not infinitesimally rigid.

Due to minimal rigidity of the target formation the matrix $R(e)^{\mathsf{T}} \in \mathbb{R}^{2n \times m}$ has full rank $|E| = 2n - 3$ for all $e \in \hat{B}^{\mathsf{T}} \mathcal{F}$, or said differently $R(e) R(e)^{\mathsf{T}}$ has no zero eigenvalues for all $e \in \hat{B}^{\mathsf{T}} \mathcal{F}$. The minimal eigenvalue of $R(e) R(e)^{\mathsf{T}}$ is positive for all $e \in \hat{B}^{\mathsf{T}} \mathcal{F}$ and thus (due to continuity of eigenvalues with respect to the matrix elements) also in an open neighborhood of $\hat{B}^{\mathsf{T}} \mathcal{F}$. In particular, for any strictly positive $\lambda > 0$, we can find $\rho = \rho(\lambda)$ so that everywhere in the sublevel set $\Omega(\rho)$ the matrix $R(e) R(e)^{\mathsf{T}}$ is positive definite with eigenvalues lower-bounded by λ. Formally, ρ is obtained by

$$\rho = \text{argmax}_{e, \tilde{\rho}} \; \tilde{\rho}$$
$$\text{subject to} \;\; e \in \Omega(\tilde{\rho})$$
$$\min_{e \in \Omega(\tilde{\rho})} \text{eig} \left(R(e) R(e)^{\mathsf{T}} \right) \ge \lambda \,.$$

Then, for all $e \in \Omega(\rho)$, we can upper-bound the derivative of $V(e)$ along trajectories as

$$\dot{V}(e) \leq -\lambda \|r_G(e) - \mathbf{d}\|^2 = -2\lambda \, V(e) \,. \tag{18.8}$$

By the Grönwall-Bellman Comparison Lemma in Exercise E14.2, we have that for every $e_0 \in \Omega(\rho)$, $V(e(t)) \leq V(e_0)e^{-2\lambda t}$. It follows that the the target formation set (parameterized in terms of relative positions) $\hat{B}^{\mathsf{T}}\mathcal{F}$ is exponentially stable with $\Omega(\rho)$ as guaranteed region of attraction.

Although the e-dynamics (18.6) and the p-dynamics (18.5) both have the formation $\mathcal{F}$ as a limit set, convergence of the e-dynamics does not automatically imply convergence to a stationary target formation (but only convergence of the point-to-set distance to $\mathcal{F}$). To establish stationarity, we rewrite the p-dynamics (18.5) as

$$p(t) = p_0 + \int_0^t f(\tau) \, d\tau \,, \tag{18.9}$$

where $f(t) = -\hat{B}\,\mathrm{diag}(e(t))\big(v(e(t) - \mathbf{d}\big)$. Due to the exponential convergence rate of the e-dynamics in W_{e_0} the function $f(t)$ is exponentially decaying in time and thus an integrable ($\mathcal{L}_1$) function. It follows that the integral on the right-hand side of (18.9) exists even in the limit as $t \to \infty$ and thus a solution of the p-dynamics converges to a finite point in $\mathcal{F}$, that is, the agents converge to a stationary target formation. In conclusion, for every $p_0 \in \mathbb{R}^{2n}$ satisfying $e_0 = \hat{B}^{\mathsf{T}}p_0 \in \Omega(\rho)$, the agents converge exponentially fast to a stationary target formation.   ∎

18.5   Historical notes and further reading

Recent surveys and tutorials on rigidity and formation control include (Anderson et al., 2008; Oh et al., 2015). The presentation in this chapter borrows ideas from (Dörfler and Francis, 2010; Krick et al., 2009).

Rigidity theory is a classic topic in graph theory; relevant references include (Asimow and Roth, 1979; White and Whiteley, 1983). A theory of network localization based on graph rigidity is provided in (Aspnes et al., 2006).

Theorem 18.9 formulated for minimally rigid formations can also be extended to more redundant infinitesimally rigid formations; see (Oh et al., 2015).

18.6 Exercises

E18.1   **Consensus with input constraints.** Consider a set of n agents each with first-order dynamics $\dot{x}_i = u_i$.

    (i) Design a consensus protocol that respects input constraints $u_i(t) \in [-1, 1]$ for all $t \geq 0$, and prove that your protocol achieves consensus.

        Hint: Adopt the hyperbolic tangent function (or the arctangent function) and Theorem 18.3.

    (ii) Extend the protocol and the proof to the case of second-order dynamics $\ddot{x}_i = u_i$ to achieve consensus of position states and convergence of velocity states to zero.

        Hint: Recall Examples 14.5 and 14.8.

E18.2   **Pentagon formation.** Consider $n = 5$ agents that should form a pentagon with unit side lengths according to the formation control protocol (18.5). Design a graph so that the pentagon formation is locally asymptotically stable.

E18.3   **Global analysis of a linear formation.** Consider two agents with positions $p_i = (x_i, y_i) \in \mathbb{R}^2$, $i \in \{1, \ldots, 2\}$, with controllable integrator dynamics $\dot{p}_i = u_i$, where $u_i \in \mathbb{R}^2$ is the steering command that serves as control input. The two agents have access to only relative position measurements $p_1 - p_2$. Your tasks are as follows:

    (i) propose a control law for u_1 and u_2 as function of the relative position $p_1 - p_2$ and a design parameter $d_{12} > 0$ so that the agents achieve a desired distance $\|p_1 - p_2\| = d_{12} > 0$ in steady state (possibly next to other undesired equilibria).

    (ii) study the convergence properties of the closed loop under your proposed control law.

    (iii) show that your proposed control law (or a modification thereof) achieves that almost all trajectories converge to the desired formation. Possibly you need to modify your controller accordingly.

Bibliography

R. P. Abelson. Mathematical models of the distribution of attitudes under controversy. In N. Frederiksen and H. Gulliksen, editors, *Contributions to Mathematical Psychology*, volume 14, pages 142–160. Holt, Rinehart, & Winston, 1964. ISBN 0030430100.

J. A. Acebrón, L. L. Bonilla, C. J. P. Vicente, F. Ritort, and R. Spigler. The Kuramoto model: A simple paradigm for synchronization phenomena. *Reviews of Modern Physics*, 77(1):137–185, 2005. doi:10.1103/RevModPhys.77.137.

D. Acemoglu and A. Ozdaglar. Opinion dynamics and learning in social networks. *Dynamic Games and Applications*, 1(1):3–49, 2011. doi:10.1007/s13235-010-0004-1.

D. Acemoglu, G. Como, F. Fagnani, and A. Ozdaglar. Opinion fluctuations and disagreement in social networks. *Mathematics of Operation Research*, 38(1):1–27, 2013. doi:10.1287/moor.1120.0570.

R. P. Agaev and P. Y. Chebotarev. The matrix of maximum out forests of a digraph and its applications. *Automation and Remote Control*, 61(9):1424–1450, 2000. URL https://arxiv.org/pdf/math/0602059.

H. J. Ahn and B. Hassibi. Global dynamics of epidemic spread over complex networks. In *IEEE Conf. on Decision and Control*, pages 4579–4585, Florence, Italy, Dec. 2013. doi:10.1109/CDC.2013.6760600.

L. J. S. Allen. Some discrete-time SI, SIR, and SIS epidemic models. *Mathematical Biosciences*, 124(1):83–105, 1994. doi:10.1016/0025-5564(94)90025-6.

B. D. O. Anderson, C. Yu, B. Fidan, and J. M. Hendrickx. Rigid graph control architectures for autonomous formations. *IEEE Control Systems*, 28(6):48–63, 2008. doi:10.1109/MCS.2008.929280.

M. Arcak. Passivity as a design tool for group coordination. *IEEE Transactions on Automatic Control*, 52(8):1380–1390, 2007. doi:10.1109/TAC.2007.902733.

M. Arcak, C. Meissen, and A. Packard. *Networks of Dissipative Systems: Compositional Certification of Stability, Performance, and Safety*. Springer, 2016. ISBN 978-3-319-29928-0. doi:10.1007/978-3-319-29928-0.

A. Arenas, A. Díaz-Guilera, J. Kurths, Y. Moreno, and C. Zhou. Synchronization in complex networks. *Physics Reports*, 469(3):93–153, 2008. doi:10.1016/j.physrep.2008.09.002.

V. I. Arnol'd. *Ordinary Differential Equations*. Springer, 1992. ISBN 3-540-54813-0. Translation of the third Russian edition by R. Cooke.

L. Asimow and B. Roth. The rigidity of graphs, II. *Journal of Mathematical Analysis and Applications*, 68(1): 171–190, 1979. doi:doi:10.1016/0022-247X(79)90108-2.

J. Aspnes, T. Eren, D. K. Goldenberg, A. S. Morse, W. Whiteley, Y. R. Yang, B. D. O. Anderson, and P. Belhumeur. A theory of network localization. *IEEE Transactions on Mobile Computing*, 5(12):1663–1678, 2006. doi:10.1109/TMC.2006.174.

B. Avramovic, P. V. Kokotović, J. R. Winkelman, and J. H. Chow. Area decomposition for electromechanical models of power systems. *Automatica*, 16(6):637–648, 1980. doi:10.1016/0005-1098(80)90006-0.

N. Azizan Ruhi and B. Hassibi. SIRS epidemics on complex networks: Concurrence of exact Markov chain and approximated models. In *IEEE Conf. on Decision and Control*, pages 2919–2926, Dec. 2015. doi:10.1109/CDC.2015.7402660.

H. Bai, M. Arcak, and J. Wen. *Cooperative Control Design*. Springer, 2011. ISBN 1461429072.

S. Baigent. Lotka-Volterra Dynamics — An Introduction. Unpublished Lecture Notes, University of College, London, Mar. 2010. URL http://www.ltcc.ac.uk/media/london-taught-course-centre/documents/Bio-Mathematics-(APPLIED).pdf.

D. Bajović, J. Xavier, J. M. F. Moura, and B. Sinopoli. Consensus and products of random stochastic matrices: Exact rate for convergence in probability. *IEEE Transactions on Signal Processing*, 61(10):2557–2571, 2013. doi:10.1109/TSP.2013.2248003.

B. Bamieh, M. R. Jovanovic, P. Mitra, and S. Patterson. Coherence in large-scale networks: Dimension-dependent limitations of local feedback. *IEEE Transactions on Automatic Control*, 57(9):2235–2249, 2012. doi:10.1109/TAC.2012.2202052.

E. A. Barbashin and N. N. Krasovskiĭ. On global stability of motion. *Doklady Akademii Nauk SSSR*, 86(3): 453–456, 1952. (In Russian).

P. Barooah and J. P. Hespanha. Estimation on graphs from relative measurements. *IEEE Control Systems*, 27 (4):57–74, 2007. doi:10.1109/MCS.2007.384125.

P. Barooah and J. P. Hespanha. Estimation from relative measurements: Electrical analogy and large graphs. *IEEE Transactions on Signal Processing*, 56(6):2181–2193, 2008. doi:10.1109/TSP.2007.912270.

A. Barrat, M. Barthlemy, and A. Vespignani. *Dynamical Processes on Complex Networks*. Cambridge University Press, 2008. ISBN 0521879507.

D. Bauso and G. Notarstefano. Distributed n-player approachability and consensus in coalitional games. *IEEE Transactions on Automatic Control*, 60(11):3107–3112, 2015. doi:10.1109/TAC.2015.2411873.

A. Bavelas. Communication patterns in task-oriented groups. *Journal of the Acoustical Society of America*, 22:725–730, 1950. doi:10.1121/1.1906679.

F. Benezit, V. Blondel, P. Thiran, J. Tsitsiklis, and M. Vetterli. Weighted gossip: Distributed averaging using non-doubly stochastic matrices. In *IEEE International Symposium on Information Theory*, pages 1753–1757, June 2010. doi:10.1109/ISIT.2010.5513273.

M. Benzi, G. H. Golub, and J. Liesen. Numerical solution of saddle point problems. *Acta Numerica*, 14:1–137, 2005. doi:10.1017/S0962492904000212.

A. R. Bergen and D. J. Hill. A structure preserving model for power system stability analysis. *IEEE Transactions on Power Apparatus and Systems*, 100(1):25–35, 1981. doi:10.1109/TPAS.1981.316883.

A. Berman and R. J. Plemmons. *Nonnegative Matrices in the Mathematical Sciences*. SIAM, 1994. ISBN 978-0-89871-321-3.

N. Biggs. *Algebraic Graph Theory*. Cambridge University Press, 2 edition, 1994. ISBN 0521458978.

N. Biggs. Algebraic potential theory on graphs. *Bulletin of the London Mathematical Society*, 29(6):641–682, 1997. doi:10.1112/S0024609397003305.

D. Bindel, J. Kleinberg, and S. Oren. How bad is forming your own opinion? *Games and Economic Behavior*, 92:248–265, 2015. doi:10.1016/j.geb.2014.06.004.

F. Blanchini and S. Miani. *Set-Theoretic Methods in Control*. Springer, 2015. ISBN 9783319179322.

V. D. Blondel and A. Olshevsky. How to decide consensus? A combinatorial necessary and sufficient condition and a proof that consensus is decidable but NP-hard. *SIAM Journal on Control and Optimization*, 52(5):2707–2726, 2014. doi:10.1137/12086594X.

S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D. U. Hwang. Complex networks: Structure and dynamics. *Physics Reports*, 424(4-5):175–308, 2006. doi:10.1016/j.physrep.2005.10.009.

N. Bof, R. Carli, and L. Schenato. On the performance of consensus based versus Lagrangian based algorithms for quadratic cost functions. In *European Control Conference*, pages 160–165, Aalborg, Denmark, June 2016. doi:10.1109/ECC.2016.7810280.

B. Bollobás. *Modern Graph Theory*. Springer, 1998. ISBN 0387984887.

S. Bolognani, S. Del Favero, L. Schenato, and D. Varagnolo. Consensus-based distributed sensor calibration and least-square parameter identification in WSNs. *International Journal of Robust and Nonlinear Control*, 20(2):176–193, 2010. doi:10.1002/rnc.1452.

P. Bonacich. Technique for analyzing overlapping memberships. *Sociological Methodology*, 4:176–185, 1972a. doi:10.2307/270732.

P. Bonacich. Factoring and weighting approaches to status scores and clique identification. *Journal of Mathematical Sociology*, 2(1):113–120, 1972b. doi:10.1080/0022250X.1972.9989806.

S. P. Borgatti and M. G. Everett. A graph-theoretic perspective on centrality. *Social Networks*, 28(4):466–484, 2006. doi:10.1016/j.socnet.2005.11.005.

S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah. Randomized gossip algorithms. *IEEE Transactions on Information Theory*, 52(6):2508–2530, 2006. doi:10.1109/TIT.2006.874516.

U. Brandes. Centrality: concepts and methods. Slides, May 2006. URL http://vw.indiana.edu/netsci06. The International Workshop/School and Conference on Network Science.

U. Brandes and T. Erlebach. *Network Analysis: Methodological Foundations*. Springer, 2005. ISBN 3540249796.

L. Breiman. *Probability*, volume 7 of *Classics in Applied Mathematics*. SIAM, 1992. ISBN 0-89871-296-3. Corrected reprint of the 1968 original.

S. Brin and L. Page. The anatomy of a large-scale hypertextual Web search engine. *Computer Networks*, 30: 107–117, 1998. doi:10.1016/S0169-7552(98)00110-X.

E. Brown, P. Holmes, and J. Moehlis. Globally coupled oscillator networks. In E. Kaplan, J. E. Marsden, and K. R. Sreenivasan, editors, *Perspectives and Problems in Nonlinear Science: A Celebratory Volume in Honor of Larry Sirovich*, pages 183–215. Springer, 2003. doi:10.1007/978-0-387-21789-5_5.

A. M. Bruckstein, N. Cohen, and A. Efrat. Ants, crickets, and frogs in cyclic pursuit. Technical Report CIS 9105, Center for Intelligent Systems, Technion, Haifa, Israel, July 1991. URL http://www.cs.technion.ac.il/tech-reports.

J. Buck. Synchronous rhythmic flashing of fireflies. II. *Quarterly Review of Biology*, 63(3):265–289, 1988. doi:10.1086/415929.

F. Bullo, J. Cortés, and S. Martínez. *Distributed Control of Robotic Networks*. Princeton University Press, 2009. ISBN 978-0-691-14195-4. URL http://www.coordinationbook.info.

Z. Burda, J. Duda, J. M. Luck, and B. Waclaw. Localization of the maximal entropy random walk. *Physical Review Letters*, 102:160602, 2009. doi:10.1103/PhysRevLett.102.160602.

M. Cao, A. S. Morse, and B. D. O. Anderson. Agreeing asynchronously. *IEEE Transactions on Automatic Control*, 53(8):1826–1838, 2008. doi:10.1109/TAC.2008.929387.

Y. Cao, W. Yu, W. Ren, and G. Chen. An overview of recent progress in the study of distributed multi-agent coordination. *IEEE Transactions on Industrial informatics*, 9(1):427–438, 2013. doi:10.1109/TII.2012.2219061.

R. Carli and S. Zampieri. Network clock synchronization based on the second-order linear consensus algorithm. *IEEE Transactions on Automatic Control*, 59(2):409–422, 2014. doi:10.1109/TAC.2013.2283742.

R. Carli, A. Chiuso, L. Schenato, and S. Zampieri. A PI consensus controller for networked clocks synchronization. In *IFAC World Congress*, volume 41, pages 10289–10294, 2008a. doi:10.3182/20080706-5-KR-1001.01741.

R. Carli, F. Fagnani, A. Speranzon, and S. Zampieri. Communication constraints in the average consensus problem. *Automatica*, 44(3):671–684, 2008b. doi:10.1016/j.automatica.2007.07.009.

R. Carli, F. Garin, and S. Zampieri. Quadratic indices for the analysis of consensus algorithms. In *IEEE Information Theory and Applications Workshop*, pages 96–104, San Diego, CA, USA, Feb. 2009. doi:10.1109/ITA.2009.5044929.

C. Castellano, S. Fortunato, and V. Loreto. Statistical physics of social dynamics. *Reviews of Modern Physics*, 81(2):591–646, 2009. doi:10.1103/RevModPhys.81.591.

H. Caswell. *Matrix Population Models*. Sinauer Associates, 2 edition, 2006. ISBN 087893121X.

A. Cayley. On the Theory of Analytic Forms Called Trees. *Philosophical Magazine*, 13:19–30, 1857.

A. G. Chandrasekhar, H. Larreguy, and J. P. Xandri. Testing models of social learning on networks: Evidence from a lab experiment in the field. Working Paper 21468, National Bureau of Economic Research, August 2015.

N. D. Charkes, P. T. M. Jr, and C. Philips. Studies of skeletal tracer kinetics. I. Digital-computer solution of a five-compartment model of [18f] fluoride kinetics in humans. *Journal of Nuclear Medicine*, 19(12): 1301–1309, 1978.

S. Chatterjee and E. Seneta. Towards consensus: Some convergence theorems on repeated averaging. *Journal of Applied Probability*, 14(1):89–97, 1977. doi:10.2307/3213262.

A. Cherukuri, B. Gharesifard, and J. Cortes. Saddle-point dynamics: Conditions for asymptotic stability of saddle points. *SIAM Journal on Control and Optimization*, 55(1):486–511, 2017. doi:10.1137/15M1026924.

N. G. Chetaev. *The Stability of Motion*. Pergamon, 1961. Translation from Russian by M. Nadler.

N. Chopra and M. W. Spong. On exponential synchronization of Kuramoto oscillators. *IEEE Transactions on Automatic Control*, 54(2):353–357, 2009. doi:10.1109/TAC.2008.2007884.

J. H. Chow, editor. *Time-Scale Modeling of Dynamic Networks with Applications to Power Systems*. Lecture Notes in Control and Information Sciences. Springer, 1982. ISBN 978-3-540-12106-0.

J. H. Chow and P. Kokotović. Time scale modeling of sparse dynamic networks. *IEEE Transactions on Automatic Control*, 30(8):714–722, 1985. doi:10.1109/TAC.1985.1104055.

J. H. Chow, J. Cullum, and R. A. Willoughby. A sparsity-based technique for identifying slow-coherent areas in large power systems. *IEEE Transactions on Power Apparatus and Systems*, 103(3):463–473, 1984. doi:10.1109/TPAS.1984.318724.

F. H. Clarke, Y. Ledyaev, R. J. Stern, and P. R. Wolenski. *Nonsmooth Analysis and Control Theory*. Springer, 1998. ISBN 0387983368.

R. Cogburn. The ergodic theory of Markov chains in random environments. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 66(1):109–128, 1984. doi:10.1007/BF00532799.

G. Como, K. Savla, D. Acemoglu, M. A. Dahleh, and E. Frazzoli. Robust distributed routing in dynamical networks – Part I: Locally responsive policies and weak resilience. *IEEE Transactions on Automatic Control*, 58(2):317–332, 2013. doi:10.1109/TAC.2012.2209951.

S. Coogan and M. Arcak. A compartmental model for traffic networks and its dynamical behavior. *IEEE Transactions on Automatic Control*, 60(10):2698–2703, 2015. doi:10.1109/TAC.2015.2411916.

J. Cortés. Discontinuous dynamical systems. *IEEE Control Systems*, 28(3):36–73, 2008. doi:10.1109/MCS.2008.919306.

E. Cristiani, B. Piccoli, and A. Tosin. *Multiscale Modeling of Pedestrian Dynamics*. Springer, 2014. ISBN 978-3-319-06619-6.

S. M. Crook, G. B. Ermentrout, M. C. Vanier, and J. M. Bower. The role of axonal delay in the synchronization of networks of coupled cortical oscillators. *Journal of Computational Neuroscience*, 4(2):161–172, 1997. doi:10.1023/A:1008843412952.

H. Daido. Quasientrainment and slow relaxation in a population of oscillators with random and frustrated interactions. *Physical Review Letters*, 68(7):1073–1076, 1992. doi:10.1103/PhysRevLett.68.1073.

J. M. Danskin. The theory of max-min, with applications. *SIAM Journal on Applied Mathematics*, 14(4): 641–664, 1966. doi:10.1137/0114053.

K. C. Das and P. Kumar. Some new bounds on the spectral radius of graphs. *Discrete Mathematics*, 281(1): 149–161, 2004. doi:10.1016/j.disc.2003.08.005.

P. J. Davis. *Circulant Matrices*. John Wiley & Sons, 1979. ISBN 0-471-05771-1. A Wiley-Interscience Publication, Pure and Applied Mathematics.

T. A. Davis and Y. Hu. The University of Florida sparse matrix collection. *ACM Transactions on Mathematical Software*, 38(1):1–25, 2011. doi:10.1145/2049662.2049663.

M. H. DeGroot. Reaching a consensus. *Journal of the American Statistical Association*, 69(345):118–121, 1974. doi:10.1080/01621459.1974.10480137.

P. M. DeMarzo, D. Vayanos, and J. Zwiebel. Persuasion bias, social influence, and unidimensional opinions. *The Quarterly Journal of Economics*, 118(3):909–968, 2003. doi:10.1162 /00335530360698469.

R. Diestel. *Graph Theory*, volume 173 of *Graduate Texts in Mathematics*. Springer, 2 edition, 2000. ISBN 3642142788.

J.-G. Dong and L. Qiu. Complex Laplacians and applications in multi-agent systems, 2014. URL https://arxiv.org/pdf/1406.1862.

F. Dörfler and F. Bullo. On the critical coupling for Kuramoto oscillators. *SIAM Journal on Applied Dynamical Systems*, 10(3):1070–1099, 2011. doi:10.1137/10081530X.

F. Dörfler and F. Bullo. Exploring synchronization in complex oscillator networks, 2012. URL http://arxiv.org/pdf/1209.1335. Extended version including proofs.

F. Dörfler and F. Bullo. Kron reduction of graphs with applications to electrical networks. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 60(1):150–163, 2013. doi:10.1109/TCSI.2012.2215780.

F. Dörfler and F. Bullo. Synchronization in complex networks of phase oscillators: A survey. *Automatica*, 50 (6):1539–1564, 2014. doi:10.1016/j.automatica.2014.04.012.

F. Dörfler and B. Francis. Geometric analysis of the formation problem for autonomous robots. *IEEE Transactions on Automatic Control*, 55(10):2379–2384, 2010. doi:10.1109/TAC.2010.2053735.

F. Dörfler, J. W. Simpson-Porco, and F. Bullo. Electrical networks and algebraic graph theory: Models, properties, and applications. *Proceedings of the IEEE*, Sept. 2017. Submitted.

G. Droge, H. Kawashima, and M. B. Egerstedt. Continuous-time proportional-integral distributed optimisation for networked systems. *Journal of Control and Decision*, 1(3):191–213, 2014. doi:10.1080/23307706.2014.926622.

C. L. DuBois. UCI Network Data Repository, 2008. URL http://networkdata.ics.uci.edu.

D. Easley and J. Kleinberg. *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge University Press, 2010. ISBN 0521195330.

L. Euler. Solutio Problematis ad Geometriam Situs Pertinentis. *Commentarii Academiae Scientiarum Imperialis Petropolitanae*, 8:128–140, 1741. Also in Opera Omnia (1), Vol. 7, 1-10.

F. Fagnani. Consensus dynamics over networks. Lecture notes for Winter School on Complex Networks, INRIA. Downloaded on 12/23/2016, Jan. 2014. URL http://www-sop.inria.fr/members/Giovanni.Neglia/complexnetworks14.

F. Fagnani and S. Zampieri. Randomized consensus algorithms over large scale networks. *IEEE Journal on Selected Areas in Communications*, 26(4):634–649, 2008. doi:10.1109/JSAC.2008.080506.

A. Fall, A. Iggidr, G. Sallet, and J.-J. Tewa. Epidemiological models and Lyapunov functions. *Mathematical Modelling of Natural Phenomena*, 2(1):62–68, 2007. doi:10.1051/mmnp:2008011.

L. Farina and S. Rinaldi. *Positive Linear Systems: Theory and Applications*. John Wiley & Sons, 2000. ISBN 0471384569.

M. Fiedler. Algebraic connectivity of graphs. *Czechoslovak Mathematical Journal*, 23(2):298–305, 1973. URL http://dml.cz/dmlcz/101168.

D. Fife. Which linear compartmental systems contain traps? *Mathematical Biosciences*, 14(3):311–315, 1972. doi:10.1016/0025-5564(72)90082-X.

S. Fortunato. Community detection in graphs. *Physics Reports*, 486(3-5):75–174, 2010. doi:10.1016/j.physrep.2009.11.002.

D. M. Foster and J. A. Jacquez. Multiple zeros for eigenvalues and the multiplicity of traps of a linear compartmental system. *Mathematical Biosciences*, 26(1):89–97, 1975. doi:10.1016/0025-5564(75)90096-6.

L. R. Foulds. *Graph Theory Applications*. Springer, 1995. ISBN 0387975993.

B. A. Francis and M. Maggiore. *Flocking and Rendezvous in Distributed Robotics*. Springer, 2016. ISBN 978-3-319-24727-4.

P. Frasca, R. Carli, F. Fagnani, and S. Zampieri. Average consensus on networks with quantized communication. *International Journal of Robust and Nonlinear Control*, 19(16):1787–1816, 2009. doi:10.1002/rnc.1396.

L. C. Freeman. A set of measures of centrality based on betweenness. *Sociometry*, 40(1):35–41, 1977. doi:10.2307/3033543.

J. R. P. French. A formal theory of social power. *Psychological Review*, 63(3):181–194, 1956. doi:10.1037/h0046123.

N. E. Friedkin. Theoretical foundations for centrality measures. *American Journal of Sociology*, 96(6): 1478–1504, 1991. doi:10.1086/229694.

N. E. Friedkin and E. C. Johnsen. Social influence networks and opinion change. In S. R. Thye, E. J. Lawler, M. W. Macy, and H. A. Walker, editors, *Advances in Group Processes*, volume 16, pages 1–29. Emerald Group Publishing Limited, 1999. ISBN 0762304529.

N. E. Friedkin and E. C. Johnsen. *Social Influence Network Theory: A Sociological Examination of Small Group Dynamics*. Cambridge University Press, 2011. ISBN 9781107002463.

N. E. Friedkin and E. C. Johnsen. Two steps to obfuscation. *Social Networks*, 39:12–13, 2014. doi:10.1016/j.socnet.2014.03.008.

N. E. Friedkin, P. Jia, and F. Bullo. A theory of the evolution of social power: Natural trajectories of interpersonal influence systems along issue sequences. *Sociological Science*, 3:444–472, 2016. doi:10.15195/v3.a20.

F. G. Frobenius. Über matrizen aus nicht negativen Elementen. 1912. doi:10.3931/e-rara-18865. Königliche Gesellschaft der Wissenschaften.

T. M. J. Fruchterman, , and E. M. Reingold. Graph drawing by force-directed placement. *Software: Practice and Experience*, 21(11):1129–1164, 1991. doi:10.1002/spe.4380211102.

P. A. Fuhrmann and U. Helmke. *The Mathematics of Networks of Linear Systems*. Springer, 2015. ISBN 3319166468.

F. R. Gantmacher. *The Theory of Matrices*, volume 1 and 2. Chelsea, New York, 1959. ISBN 0-8218-1376-5 and 0-8218-2664-6. Translation of German edition by K. A. Hirsch.

C. Gao, J. Cortés, and F. Bullo. Notes on averaging over acyclic digraphs and discrete coverage control. *Automatica*, 44(8):2120–2127, 2008. doi:10.1016/j.automatica.2007.12.017.

F. Garin and L. Schenato. A survey on distributed estimation and control applications using linear consensus algorithms. In A. Bemporad, M. Heemels, and M. Johansson, editors, *Networked Control Systems*, LNCIS, pages 75–107. Springer, 2010. doi:10.1007/978-0-85729-033-5_3.

B. Gharesifard and J. Cortes. Distributed continuous-time convex optimization on weight-balanced digraphs. *IEEE Transactions on Automatic Control*, 59(3):781–786, 2014. doi:10.1109/TAC.2013.2278132.

A. K. Ghosh, B. Chance, and E. K. Pye. Metabolic coupling and synchronization of NADH oscillations in yeast cell populations. *Archives of Biochemistry and Biophysics*, 145(1):319–331, 1971. doi:10.1016/0003-9861(71)90042-7.

G. Giorgi and S. Komlósi. Dini derivatives in optimization — Part I. *Rivista di Matematica Per Le Scienze Economiche e Sociali*, 15(1):3–30, 1992. doi:10.1007/BF02086523.

D. Gleich. Spectral Graph Partitioning and the Laplacian with Matlab, Jan. 2006. URL https://www.cs.purdue.edu/homes/dgleich/demos/matlab/spectral/spectral.html. (Last retrieved on May 30, 2016.).

D. F. Gleich. PageRank beyond the Web. *SIAM Review*, 57(3):321–363, 2015. doi:10.1137/140976649.

C. D. Godsil and G. F. Royle. *Algebraic Graph Theory*. Springer, 2001. ISBN 0387952411.

R. Goebel, R. G. Sanfelice, and A. R. Teel. *Hybrid Dynamical Systems: Modeling, Stability, and Robustness*. Princeton University Press, 2012. ISBN 9780691153896.

B. S. Goh. Global stability in two species interactions. *Journal of Mathematical Biology*, 3(3-4):313–318, 1976. doi:10.1007/BF00275063.

B. S. Goh. Stability in models of mutualism. *American Naturalist*, pages 261–275, 1979. doi:10.1086/283384.

B.-S. Goh. *Management and Analysis of Biological Populations*. Elsevier, 1980. ISBN 978-0-444-41793-0.

M. Grant and S. Boyd. CVX: Matlab software for disciplined convex programming, version 2.1, Oct. 2016. URL http://cvxr.com/cvx.

J. Guckenheimer and P. Holmes. *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*, volume 42 of *Applied Mathematical Sciences*. Springer, 1990. ISBN 0387908196.

H. Guo, M. Li, and Z. Shuai. A graph-theoretic approach to the method of global Lyapunov functions. *Proceedings of the American Mathematical Society*, 136(8):2793–2802, 2008. doi:10.1090/S0002-9939-08-09341-6.

A. Gushchin, E. Mallada, and A. Tang. Phase-coupled oscillators with plastic coupling: Synchronization and stability. *IEEE Transactions on Network Science and Engineering*, 3(4):240–256, 2016. doi:10.1109/TNSE.2016.2605096.

I. Gutman and W. Xiao. Generalized inverse of the Laplacian matrix and some applications. *Bulletin (AcadÃľmie serbe des sciences et des arts. Classe des sciences mathÃľmatiques et naturelles. Sciences mathÃľmatiques)*, 129(29):15–23, 2004. URL http://emis.ams.org/journals/BSANU/29/2.html.

W. H. Haddad, V. Chellaboina, and Q. Hui. *Nonnegative and Compartmental Dynamical Systems*. Princeton University Press, 2010. ISBN 0691144117.

W. M. Haddad and V. Chellaboina. *Nonlinear Dynamical Systems and Control: A Lyapunov-Based Approach*. Princeton University Press, 2008. ISBN 9780691133294.

W. Hahn. *Stability of Motion.* Springer, 1967.

F. Harary. A criterion for unanimity in French's theory of social power. In D. Cartwright, editor, *Studies in Social Power*, pages 168–182. University of Michigan, 1959. ISBN 0879442301. URL http://psycnet.apa.org/psycinfo/1960-06701-006.

F. Harary. *Graph Theory.* Addison-Wesley, 1969. URL http://www.dtic.mil/dtic/tr/fulltext/u2/705364.pdf.

T. Hatanaka, Y. Igarashi, M. Fujita, and M. W. Spong. Passivity-based pose synchronization in three dimensions. *IEEE Transactions on Automatic Control*, 57(2):360–375, 2012. doi:10.1109/TAC.2011.2166668.

Y. Hatano and M. Mesbahi. Agreement over random networks. *IEEE Transactions on Automatic Control*, 50 (11):1867–1872, 2005. doi:10.1109/TAC.2005.858670.

J. M. Hendrickx. *Graphs and Networks for the Analysis of Autonomous Agent Systems.* PhD thesis, Departement d'Ingenierie Mathematique, Université Catholique de Louvain, Belgium, Feb. 2008.

J. M. Hendrickx and J. N. Tsitsiklis. Convergence of type-symmetric and cut-balanced consensus seeking systems. *IEEE Transactions on Automatic Control*, 58(1):214–218, 2013. doi:10.1109/TAC.2012.2203214.

J. P. Hespanha. *Linear Systems Theory.* Princeton University Press, 2009. ISBN 0691140219.

H. W. Hethcote. An immunization model for a heterogeneous population. *Theoretical Population Biology*, 14(3):338–349, 1978. doi:10.1016/0040-5809(78)90011-4.

H. W. Hethcote. The mathematics of infectious diseases. *SIAM Review*, 42(4):599–653, 2000. doi:10.1137/S0036144500371907.

M. W. Hirsch and S. Smale. *Differential Equations, Dynamical Systems and Linear Algebra.* Academic Press, 1974. ISBN 0123495504.

J. Hofbauer and K. Sigmund. *Evolutionary Games and Population Dynamics.* Cambridge University Press, 1998. ISBN 052162570X.

L. Hogben, editor. *Handbook of Linear Algebra.* Chapman and Hall/CRC, 2 edition, 2013. ISBN 1466507284.

Y. Hong, J. Hu, and L. Gao. Tracking control for multi-agent consensus with an active leader and variable topology. *Automatica*, 42(7):1177–1182, 2006. doi:10.1016/j.automatica.2006.02.013.

Y. Hong, L. Gao, D. Cheng, and J. Hu. Lyapunov-based approach to multiagent systems with switching jointly connected interconnection. *IEEE Transactions on Automatic Control*, 52(5):943–948, 2007. doi:10.1109/TAC.2007.895860.

F. C. Hoppensteadt and E. M. Izhikevich. Synchronization of laser oscillators, associative memory, and optical neurocomputing. *Physical Review E*, 62(3):4010–4013, 2000. doi:10.1103/PhysRevE.62.4010.

R. A. Horn and C. R. Johnson. *Matrix Analysis.* Cambridge University Press, 1985. ISBN 0521386322.

R. A. Horn and C. R. Johnson. *Topics in Matrix Analysis*. Cambridge University Press, 1994. ISBN 0521467136.

J. Hu and Y. Hong. Leader-following coordination of multi-agent systems with coupling time delays. *Physica A: Statistical Mechanics and its Applications*, 374(2):853–863, 2007. doi:10.1016/j.physa.2006.08.015.

Y. Hu. Efficient, high-quality force-directed graph drawing. *Mathematica Journal*, 10(1):37–71, 2005. URL http://www.mathematica-journal.com/issue/v10i1/.

C. Huygens. *Horologium Oscillatorium*. Paris, France, 1673.

H. Ishii and R. Tempo. Distributed randomized algorithms for the PageRank computation. *IEEE Transactions on Automatic Control*, 55(9):1987–2002, 2010. doi:10.1109/TAC.2010.2042984.

H. Ishii and R. Tempo. The PageRank problem, multiagent consensus, and web aggregation: A systems and control viewpoint. *IEEE Control Systems*, 34(3):34–53, 2014. doi:10.1109/MCS.2014.2308672.

M. O. Jackson. *Social and Economic Networks*. Princeton University Press, 2010. ISBN 0691148201.

J. A. Jacquez and C. P. Simon. Qualitative theory of compartmental systems. *SIAM Review*, 35(1):43–79, 1993. doi:10.1137/1035003.

A. Jadbabaie and A. Olshevsky. On performance of consensus protocols subject to noise: Role of hitting times and network structure, 2015. URL https://arxiv.org/pdf/1508.00036.

A. Jadbabaie, J. Lin, and A. S. Morse. Coordination of groups of mobile autonomous agents using nearest neighbor rules. *IEEE Transactions on Automatic Control*, 48(6):988–1001, 2003. doi:10.1109/TAC.2003.812781.

S. Jafarpour and F. Bullo. Synchronization of Kuramoto oscillators via cutset projections. *IEEE Transactions on Automatic Control*, Nov. 2017. Submitted.

S. Jafarpour, V. Purba, S. V. Dhople, B. Johnson, and F. Bullo. Robust stability of grid-tied inverter networks, Dec. 2017. Working Paper.

G. Jongen, J. Anemüller, D. Bollé, A. C. C. Coolen, and C. Perez-Vicente. Coupled dynamics of fast spins and slow exchange interactions in the XY spin glass. *Journal of Physics A: Mathematical and General*, 34 (19):3957–3984, 2001. doi:10.1088/0305-4470/34/19/302.

A. Kashyap, T. Başar, and R. Srikant. Quantized consensus. *Automatica*, 43(7):1192–1203, 2007. doi:10.1016/j.automatica.2007.01.002.

L. Katz. A new status index derived from sociometric analysis. *Psychometrika*, 18(1):39–43, 1953. doi:10.1007/BF02289026.

T. Kavitha, C. Liebchen, K. Mehlhorn, D. Michail, R. Rizzi, T. Ueckerdt, and K. A. Zweig. Cycle bases in graphs characterization, algorithms, complexity, and applications. *Computer Science Review*, 3(4):199–243, 2009. doi:10.1016/j.cosrev.2009.08.001.

D. Kempe, A. Dobra, and J. Gehrke. Gossip-based computation of aggregate information. In *IEEE Symposium on Foundations of Computer Science*, pages 482–491, Washington, DC, Oct. 2003. doi:10.1109/SFCS.2003.1238221.

H. K. Khalil. *Nonlinear Systems*. Prentice Hall, 3 edition, 2002. ISBN 0130673897.

A. Khanafer, T. Başar, and B. Gharesifard. Stability of epidemic models over directed graphs: A positive systems approach. *Automatica*, 74:126–134, 2016. doi:10.1016/j.automatica.2016.07.037.

G. Kirchhoff. Über die Auflösung der Gleichungen, auf welche man bei der Untersuchung der linearen Verteilung galvanischer Ströme geführt wird. *Annalen der Physik und Chemie*, 148(12):497–508, 1847. doi:10.1002/andp.18471481202.

I. Z. Kiss, Y. Zhai, and J. L. Hudson. Emerging coherence in a population of chemical oscillators. *Science*, 296(5573):1676–1678, 2002. doi:10.1126/science.1070757.

M. S. Klamkin and D. J. Newman. Cyclic pursuit or "the three bugs problem". *American Mathematical Monthly*, 78(6):631–639, 1971. doi:10.4169/amer.math.monthly.122.04.377.

D. J. Klein, P. Lee, K. A. Morgansen, and T. Javidi. Integration of communication and control using discrete time Kuramoto models for multivehicle coordination over broadcast networks. *IEEE Journal on Selected Areas in Communications*, 26(4):695–705, 2008. doi:10.1109/JSAC.2008.080511.

J. M. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46(5):604–632, 1999. doi:10.1145/324133.324140.

G. Kozyreff, A. G. Vladimirov, and P. Mandel. Global coupling with time delay in an array of semiconductor lasers. *Physical Review Letters*, 85(18):3809–3812, 2000. doi:10.1103/PhysRevLett.85.3809.

D. Krackhardt. Cognitive social structures. *Social Networks*, 9(2):109–134, 1987. doi:10.1016/0378-8733(87)90009-8.

N. N. Krasovskiĭ. *Stability of Motion. Applications of Lyapunov's Second Method to Differential Systems and Equations with Delay*. Stanford University Press, 1963. Translation of the 1959 edition in Russian by J. L. Brenner.

L. Krick, M. E. Broucke, and B. Francis. Stabilization of infinitesimally rigid formations of multi-robot networks. *International Journal of Control*, 82(3):423–439, 2009. doi:10.1080/00207170802108441.

J. Kunegis. KONECT: the Koblenz network collection. In *International Conference on World Wide Web Companion*, pages 1343–1350, Rio de Janeiro, Brazil, May 2013. doi:10.1145/2487788.2488173.

Y. Kuramoto. Self-entrainment of a population of coupled non-linear oscillators. In H. Araki, editor, *Int. Symposium on Mathematical Problems in Theoretical Physics*, volume 39 of *Lecture Notes in Physics*, pages 420–422. Springer, 1975. ISBN 978-3-540-07174-7. doi:10.1007/BFb0013365.

Y. Kuramoto. *Chemical Oscillations, Waves, and Turbulence*. Springer, 1984. ISBN 0387133224.

J. L. Lagrange. *Mécanique analytique*. Chez la Veuve Desaint, Paris, 1788. Translation of revised edition: Lagrange (1997).

J. L. Lagrange. *Analytical Mechanics*. Number 191 in Boston Studies in the Philosophy of Science. Kluwer Academic Publishers, Dordrecht, 1997. ISBN 0-7923-4349-2. Translation of the 1811 edition in French by A. Boissonnade and V. N. Vagliente.

A. Lajmanovich and J. A. Yorke. A deterministic model for gonorrhea in a nonhomogeneous population. *Mathematical Biosciences*, 28(3):221–236, 1976. doi:10.1016/0025-5564(76)90125-5.

J. P. LaSalle. Some extensions of Liapunov's second method. *IRE Trans. Circuit Theory*, CT-7:520–527, 1960. doi:10.1109/TCT.1960.1086720.

J. P. LaSalle. Stability theory for ordinary differential equations. *Journal of Differential Equations*, 4:57–65, 1968. doi:10.1016/0022-0396(68)90048-X.

J. P. LaSalle. *The Stability of Dynamical Systems*. SIAM, 1976. ISBN 9780898710229. doi:10.1137/1.9781611970432.

N. E. Leonard, T. Shen, B. Nabet, L. Scardovi, I. D. Couzin, and S. A. Levin. Decision versus compromise for animal groups in motion. *Proceedings of the National Academy of Sciences*, 109(1):227–232, 2012. doi:10.1073/pnas.1118318108.

P. H. Leslie. On the use of matrices in certain population mathematics. *Biometrika*, 3(3):183–212, 1945. doi:10.1007/978-3-642-81046-6_26.

Z. Li, Z. Duan, G. Chen, and L. Huang. Consensus of multiagent systems and synchronization of complex networks: A unified viewpoint. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 57(1):213–224, 2010. doi:10.1109/TCSI.2009.2023937.

P. Lin and Y. Jia. Average consensus in networks of multi-agents with both switching topology and coupling time-delay. *Physica A: Statistical Mechanics and its Applications*, 387(1):303–313, 2008. doi:10.1016/j.physa.2007.08.040.

Z. Lin, B. Francis, and M. Maggiore. Necessary and sufficient graphical conditions for formation control of unicycles. *IEEE Transactions on Automatic Control*, 50(1):121–127, 2005. doi:10.1109/TAC.2004.841121.

Z. Lin, B. Francis, and M. Maggiore. State agreement for continuous-time coupled nonlinear systems. *SIAM Journal on Control and Optimization*, 46(1):288–307, 2007. doi:10.1137/050626405.

Z. Lin, W. Ding, G. Yan, C. Yu, and A. Giua. Leader–follower formation via complex Laplacian. *Automatica*, 49(6):1900–1906, 2013. doi:10.1016/j.automatica.2013.02.055.

C. Liu, D. R. Weaver, S. H. Strogatz, and S. M. Reppert. Cellular construction of a circadian clock: Period determination in the suprachiasmatic nuclei. *Cell*, 91(6):855–860, 1997. doi:10.1016/S0092-8674(00)80473-0.

A. J. Lohwater. Introduction to Inequalities. Unpublished Lecture Notes, reproduced with permission of Marjorie Lohwater, 1982. URL http://www.mediafire.com/?1mw1tkgozzu.

S. Łojasiewicz. Sur les trajectoires du gradient d'une fonction analytique. *Seminari di Geometria 1982-1983*, pages 115–117, 1984. Istituto di Geometria, Dipartimento di Matematica, Università di Bologna, Italy.

A. J. Lotka. Analytical note on certain rhythmic relations in organic systems. *Proceedings of the National Academy of Sciences*, 6(7):410–415, 1920. doi:10.1073/pnas.6.7.410.

E. Lovisari, F. Garin, and S. Zampieri. Resistance-based performance analysis of the consensus algorithm over geometric graphs. *SIAM Journal on Control and Optimization*, 51(5):3918–3945, 2013. doi:10.1137/110857428.

D. G. Luenberger. *Introduction to Dynamic Systems: Theory, Models, and Applications.* John Wiley & Sons, 1979. ISBN 0471025941.

D. G. Luenberger and Y. Ye. *Linear and Nonlinear Programming.* Springer, 4 edition, 2008. ISBN 9780387745022.

A. M. Lyapunov. *Obščaya zadača ob ustoĭčivosti dviženiya.* Fakul′teta i Khar′kovskogo Matematicheskogo Obshchestva, Kharkov, 1892. Translation: Lyapunov (1992).

A. M. Lyapunov. *The General Problem of the Stability of Motion.* Taylor & Francis, 1992. Translation from Russian by A. T. Fuller.

E. Mallada, R. A. Freeman, and A. K. Tang. Distributed synchronization of heterogeneous oscillators on networks with arbitrary topology. *IEEE Transactions on Control of Network Systems*, 3(1):1–12, 2016. doi:10.1109/TCNS.2015.2428371.

J. R. Marden, G. Arslan, and J. S. Shamma. Joint strategy fictitious play with inertia for potential games. *IEEE Transactions on Automatic Control*, 54(2):208–220, 2009. doi:10.1109/TAC.2008.2010885.

J. A. Marshall, M. E. Broucke, and B. A. Francis. Formations of vehicles in cyclic pursuit. *IEEE Transactions on Automatic Control*, 49(11):1963–1974, 2004. doi:10.1109/TAC.2004.837589.

S. Martínez, J. Cortés, and F. Bullo. Motion coordination with distributed information. *IEEE Control Systems*, 27(4):75–88, 2007. doi:10.1109/MCS.2007.384124.

I. Matei, J. S. Baras, and C. Somarakis. Convergence results for the linear consensus problem under Markovian random graphs. *SIAM Journal on Control and Optimization*, 51(2):1574–1591, 2013. doi:10.1137/100816870.

A. Mauroy, P. Sacré, and R. J. Sepulchre. Kick synchronization versus diffusive synchronization. In *IEEE Conf. on Decision and Control*, pages 7171–7183, Maui, HI, USA, Dec. 2012. doi:10.1109/CDC.2012.6425821.

J. C. Maxwell. On governors. *Proceedings of the Royal Society. London. Series A. Mathematical and Physical Sciences*, 16:270–283, 1868.

W. Mei, S. Mohagheghi, S. Zampieri, and F. Bullo. On the dynamics of deterministic epidemic propagation over networks. *Annual Reviews in Control*, 44:116–128, 2017. doi:10.1016/j.arcontrol.2017.09.002.

R. Merris. Laplacian matrices of a graph: A survey. *Linear Algebra and its Applications*, 197:143–176, 1994. doi:10.1016/j.laa.2011.11.032 3374.

M. Mesbahi and M. Egerstedt. *Graph Theoretic Methods in Multiagent Networks*. Princeton University Press, 2010. ISBN 9781400835355.

C. D. Meyer. *Matrix Analysis and Applied Linear Algebra*. SIAM, 2001. ISBN 0898714540.

D. C. Michaels, E. P. Matyas, and J. Jalife. Mechanisms of sinoatrial pacemaker synchronization: A new hypothesis. *Circulation Research*, 61(5):704–714, 1987. doi:10.1161/01.RES.61.5.704.

B. Mohar. The Laplacian spectrum of graphs. In Y. Alavi, G. Chartrand, O. R. Oellermann, and A. J. Schwenk, editors, *Graph Theory, Combinatorics, and Applications*, volume 2, pages 871–898. John Wiley & Sons, 1991. ISBN 0471532452. URL http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.96.2577.

L. Moreau. Stability of continuous-time distributed consensus algorithms. In *IEEE Conf. on Decision and Control*, pages 3998–4003, Nassau, Bahamas, 2004. doi:10.1109/CDC.2004.1429377.

L. Moreau. Stability of multiagent systems with time-dependent communication links. *IEEE Transactions on Automatic Control*, 50(2):169–182, 2005. doi:10.1109/TAC.2004.841888.

S. Muthukrishnan, B. Ghosh, and M. H. Schultz. First-and second-order diffusive methods for rapid, coarse, distributed load balancing. *Theory of Computing Systems*, 31(4):331–354, 1998. doi:10.1007/s002240000092.

Z. Néda, E. Ravasz, T. Vicsek, Y. Brechet, and A.-L. Barabási. Physics of the rhythmic applause. *Physical Review E*, 61(6):6987–6992, 2000. doi:10.1103/PhysRevE.61.6987.

A. Nedić, A. Olshevsky, A. Ozdaglar, and J. N. Tsitsiklis. On distributed averaging algorithms and quantization effects. *IEEE Transactions on Automatic Control*, 54(11):2506–2517, 2009. doi:10.1109/TAC.2009.2031203.

Y. Nesterov. Efficiency of coordinate descent methods on huge-scale optimization problems. *SIAM Journal on Optimization*, 22(2):341–362, 2012. doi:10.1137/100802001.

M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45(2):167–256, 2003. doi:10.1137/S003614450342480.

M. E. J. Newman. *Networks: An Introduction*. Oxford University Press, 2010. ISBN 0199206651.

V. Nikiforov. Some inequalities for the largest eigenvalue of a graph. *Combinatorics, Probability and Computing*, 11(2):179–189, 2002. doi:10.1017/S0963548301004928.

C. Nowzari, V. M. Preciado, and G. J. Pappas. Analysis and control of epidemics: A survey of spreading processes on complex networks. *IEEE Control Systems*, 36(1):26–46, 2016. doi:10.1109/MCS.2015.2495000.

I. Noy-Meir. Desert ecosystems. I. Environment and producers. *Annual Review of Ecology and Systematics*, pages 25–51, 1973. URL http://www.jstor.org/stable/2096803.

E. P. Odum. *Fundamentals of Ecology*. Saunders Company, 1959.

K.-K. Oh, M.-C. Park, and H.-S. Ahn. A survey of multi-agent formation control. *Automatica*, 53:424–440, 2015. doi:10.1016/j.automatica.2014.10.022.

R. Olfati-Saber and R. M. Murray. Consensus problems in networks of agents with switching topology and time-delays. *IEEE Transactions on Automatic Control*, 49(9):1520–1533, 2004. doi:10.1109/TAC.2004.834113.

R. Olfati-Saber, E. Franco, E. Frazzoli, and J. S. Shamma. Belief consensus and distributed hypothesis testing in sensor networks. In P. J. Antsaklis and P. Tabuada, editors, *Network Embedded Sensing and Control. (Proceedings of NESC'05 Worskhop)*, Lecture Notes in Control and Information Sciences, pages 169–182. Springer, 2006. ISBN 3540327940. doi:10.1007/11533382_11.

A. Olshevsky. Linear time average consensus on fixed graphs and implications for decentralized optimization and multi-agent control, 2014. URL https://arxiv.org/pdf/1411.4186.

A. Olshevsky and J. N. Tsitsiklis. Convergence speed in distributed consensus and averaging. *SIAM Journal on Control and Optimization*, 48(1):33–55, 2009. doi:10.1137/060678324.

R. W. Owens. An algorithm to solve the Frobenius problem. *Mathematics Magazine*, 76(4):264–275, 2003. doi:10.2307/3219081.

L. Page. Method for node ranking in a linked database, Sept. 2001. URL https://www.google.com/patents/US6285999. US Patent 6,285,999.

D. A. Paley, N. E. Leonard, R. Sepulchre, D. Grunbaum, and J. K. Parrish. Oscillator models and collective motion. *IEEE Control Systems*, 27(4):89–105, 2007. doi:10.1109/MCS.2007.384123.

J. Pantaleone. Stability of incoherence in an isotropic gas of oscillating neutrinos. *Physical Review D*, 58(7):073002, 1998. doi:10.1103/PhysRevD.58.073002.

O. Perron. Zur Theorie der Matrices. *Mathematische Annalen*, 64(2):248–263, 1907. doi:10.1007/BF01449896.

G. Piovan, I. Shames, B. Fidan, F. Bullo, and B. D. O. Anderson. On frame and orientation localization for relative sensing networks. *Automatica*, 49(1):206–213, 2013. doi:10.1016/j.automatica.2012.09.014.

M. Pirani and S. Sundaram. On the smallest eigenvalue of grounded Laplacian matrices. *IEEE Transactions on Automatic Control*, 61(2):509–514, 2016. doi:10.1109/TAC.2015.2444191.

B. T. Polyak. Some methods of speeding up the convergence of iteration methods. *USSR Computational Mathematics and Mathematical Physics*, 4(5):1–17, 1964. doi:10.1016/0041-5553(64)90137-5.

V. H. Poor. *An Introduction to Signal Detection and Estimation*. Springer, 2 edition, 1998. ISBN 0387941738.

M. A. Porter and J. P. Gleeson. *Dynamical Systems on Networks: A Tutorial*. Springer, 2016. ISBN 978-3-319-26641-1. doi:10.1007/978-3-319-26641-1.

M. A. Porter, J.-P. Onnela, and P. J. Mucha. Communities in networks. *Notices of the AMS*, 56(9):1082–1097, 2009. URL http://www.ams.org/notices/200909/rtx090901082p.pdf.

R. Potrie and P. Monzón. Local implications of almost global stability. *Dynamical Systems*, 24(1):109–115, 2009. doi:10.1080/14689360802474657.

A. V. Proskurnikov and R. Tempo. A tutorial on modeling and analysis of dynamic social networks. Part I. *Annual Reviews in Control*, 43:65–79, 2017. doi:10.1016/j.arcontrol.2017.03.002.

A. Rantzer. Scalable control of positive systems. *European Journal of Control*, 24:72–80, 2015. doi:10.1016/j.ejcon.2015.04.004.

B. S. Y. Rao and H. F. Durrant-Whyte. A decentralized Bayesian algorithm for identification of tracked targets. *IEEE Transactions on Systems, Man & Cybernetics*, 23(6):1683–1698, 1993. doi:10.1109/21.257763.

C. Ravazzi, P. Frasca, R. Tempo, and H. Ishii. Ergodic randomized algorithms and dynamics over networks. *IEEE Transactions on Control of Network Systems*, 2(1):78–87, 2015. doi:10.1109/TCNS.2014.2367571.

N. Reff. Spectral properties of complex unit gain graphs. *Linear Algebra and its Applications*, 436(9): 3165–3176, 2012. doi:10.1016/j.laa.2011.10.021.

W. Ren. On consensus algorithms for double-integrator dynamics. *IEEE Transactions on Automatic Control*, 53(6):1503–1509, 2008a. doi:10.1109/TAC.2008.924961.

W. Ren. Synchronization of coupled harmonic oscillators with local interaction. *Automatica*, 44:3196–3200, 2008b. doi:10.1016/j.automatica.2008.05.027.

W. Ren and W. Atkins. Second-order consensus protocols in multiple vehicle systems with local interactions. In *AIAA Guidance, Navigation, and Control Conference and Exhibit*, pages 15–18, San Francisco, CA, USA, Aug. 2005. doi:10.2514/6.2005-6238.

W. Ren and R. W. Beard. Consensus seeking in multiagent systems under dynamically changing interaction topologies. *IEEE Transactions on Automatic Control*, 50(5):655–661, 2005. doi:10.1109/TAC.2005.846556.

W. Ren and R. W. Beard. *Distributed Consensus in Multi-vehicle Cooperative Control*. Communications and Control Engineering. Springer, 2008. ISBN 978-1-84800-014-8.

W. Ren, R. W. Beard, and E. M. Atkins. Information consensus in multivehicle cooperative control. *IEEE Control Systems*, 27(2):71–82, 2007. doi:10.1109/MCS.2007.338264.

C. W. Reynolds. Flocks, herds, and schools: A distributed behavioral model. *Computer Graphics*, 21(4): 25–34, 1987. doi:10.1145/37402.37406.

F. D. Sahneh, C. Scoglio, and P. Van Mieghem. Generalized epidemic mean-field model for spreading processes over multilayer complex networks. *IEEE/ACM Transactions on Networking*, 21(5):1609–1620, 2013. doi:10.1109/TNET.2013.2239658.

V. R. Saksena, J. O'Reilly, and P. V. Kokotović. Singular perturbations and time-scale methods in control theory: Survey 1976-1983. *Automatica*, 20(3):273–293, 1984. doi:10.1016/0005-1098(84)90044-X.

S. F. Sampson. *Crisis in a Cloister*. PhD thesis, Department of Sociology, Cornell University, 1969.

W. H. Sandholm. *Population Games and Evolutionary Dynamics*. MIT Press, 2010. ISBN 0262195879.

P. Santesso and M. E. Valcher. On the zero pattern properties and asymptotic behavior of continuous-time positive system trajectories. *Linear Algebra and its Applications*, 425(2):283–302, 2007. doi:10.1016/j.laa.2007.01.014.

L. Scardovi and R. Sepulchre. Synchronization in networks of identical linear systems. *Automatica*, 45(11): 2557–2562, 2009. doi:10.1016/j.automatica.2009.07.006.

L. Schenato and F. Fiorentin. Average TimeSynch: A consensus-based protocol for clock synchronization in wireless sensor networks. *Automatica*, 47(9):1878–1886, 2011. doi:10.1016/j.automatica.2011.06.012.

R. Sepulchre, D. A. Paley, and N. E. Leonard. Stabilization of planar collective motion: All-to-all communication. *IEEE Transactions on Automatic Control*, 52(5):811–824, 2007. doi:10.1109/TAC.2007.898077.

J. R. Silvester. Determinants of block matrices. *The Mathematical Gazette*, 84(501):460–467, 2000. doi:10.2307/3620776.

O. Simeone, U. Spagnolini, Y. Bar-Ness, and S. H. Strogatz. Distributed synchronization in wireless networks. *IEEE Signal Processing Magazine*, 25(5):81–97, 2008. doi:10.1109/MSP.2008.926661.

J. W. Simpson-Porco, F. Dörfler, and F. Bullo. Synchronization and power sharing for droop-controlled inverters in islanded microgrids. *Automatica*, 49(9):2603–2611, 2013. doi:10.1016/j.automatica.2013.05.018.

F. Sivrikaya and B. Yener. Time synchronization in sensor networks: A survey. *IEEE Network*, 18(4):45–50, 2004. doi:10.1109/MNET.2004.1316761.

S. L. Smith, M. E. Broucke, and B. A. Francis. A hierarchical cyclic pursuit scheme for vehicle networks. *Automatica*, 41(6):1045–1053, 2005. doi:10.1016/j.automatica.2005.01.001.

E. D. Sontag. *Mathematical Control Theory: Deterministic Finite Dimensional Systems*. Springer, 2 edition, 1998. ISBN 0387984895.

D. Spielman. Graphs, vectors, and matrices. *Bulletin of the American Mathematical Society*, 54(1):45–61, 2017. doi:10.1090/bull/1557.

M. W. Spong and N. Chopra. Synchronization of networked Lagrangian systems. In *Lagrangian and Hamiltonian Methods for Nonlinear Control 2006*, volume 366 of *Lecture Notes in Control and Information Sciences*, pages 47–59. Springer, 2007. ISBN 978-3-540-73889-3.

G. Strang. The fundamental theorem of linear algebra. *American Mathematical Monthly*, 100(9):848–855, 1993. doi:10.2307/2324660.

S. H. Strogatz. From Kuramoto to Crawford: Exploring the onset of synchronization in populations of coupled oscillators. *Physica D: Nonlinear Phenomena*, 143(1):1–20, 2000. doi:10.1016/S0167-2789(00)00094-4.

B. Sundararaman, U. Buy, and A. D. Kshemkalyani. Clock synchronization for wireless sensor networks: a survey. *Ad Hoc Networks*, 3(3):281–323, 2005. doi:10.1016/j.adhoc.2005.01.002.

A. Tahbaz-Salehi and A. Jadbabaie. A necessary and sufficient condition for consensus over random networks. *IEEE Transactions on Automatic Control*, 53(3):791–795, 2008. doi:10.1109/TAC.2008.917743.

A. Tahbaz-Salehi and A. Jadbabaie. Consensus over ergodic stationary graph processes. *IEEE Transactions on automatic Control*, 55(1):225–230, 2010. doi:10.1109/TAC.2009.2034054.

Y. Takeuchi. *Global Dynamical Properties of Lotka-Volterra Systems*. World Scientific Publishing, 1996. ISBN 9810224710.

Y. Takeuchi, N. Adachi, and H. Tokumaru. The stability of generalized Volterra equations. *Journal of Mathematical Analysis and Applications*, 62(3):453–473, 1978. doi:10.1016/0022-247X(78)90139-7.

P. A. Tass. A model of desynchronizing deep brain stimulation with a demand-controlled coordinated reset of neural subpopulations. *Biological Cybernetics*, 89(2):81–88, 2003. doi:10.1007/s00422-003-0425-7.

W. Thomson and P. G. Tait. *Treatise on Natural Philosophy*. Oxford University Press, 1867.

B. Touri and A. Nedić. Product of random stochastic matrices. *IEEE Transactions on Automatic Control*, 59 (2):437–448, 2014. doi:10.1109/TAC.2013.2283750.

D. J. Trudnowski, J. R. Smith, T. A. Short, and D. A. Pierre. An application of Prony methods in PSS design for multimachine systems. *IEEE Transactions on Power Systems*, 6(1):118–126, 1991. doi:10.1109/59.131054.

J. N. Tsitsiklis. *Problems in Decentralized Decision Making and Computation*. PhD thesis, Massachusetts Institute of Technology, Nov. 1984.

J. N. Tsitsiklis, D. P. Bertsekas, and M. Athans. Distributed asynchronous deterministic and stochastic gradient optimization algorithms. *IEEE Transactions on Automatic Control*, 31(9):803–812, 1986. doi:10.1109/TAC.1986.1104412.

A. J. van der Schaft. Modeling of physical network systems. *Systems & Control Letters*, 2015. ISSN 0167-6911. doi:10.1016/j.sysconle.2015.08.013.

A. J. van der Schaft and J. Wei. A Hamiltonian perspective on the control of dynamical distribution networks. In *Lagrangian and Hamiltonian Methods for Nonlinear Control 2012*, pages 24–29. 2012. doi:10.3182/20120829-3-IT-4022.00033.

P. Van Mieghem, J. Omic, and R. Kooij. Virus spread in networks. *IEEE/ACM Transactions on Networking*, 17 (1):1–14, 2009. doi:10.1109/TNET.2008.925623.

F. Varela, J. P. Lachaux, E. Rodriguez, and J. Martinerie. The brainweb: Phase synchronization and large-scale integration. *Nature Reviews Neuroscience*, 2(4):229–239, 2001. doi:10.1038/35067550.

R. S. Varga. *Matrix Iterative Analysis*. Springer, 2009. ISBN 3642051545.

M. Vidyasagar. *Nonlinear Systems Analysis*. SIAM, 2002. ISBN 9780898715262. doi:10.1137/1.9780898719185.

V. Volterra. Variations and fluctuations of the number of individuals in animal species living together. *ICES Journal of Marine Science*, 3(1):3–51, 1928. doi:10.1093/icesjms/3.1.3.

T. J. Walker. Acoustic synchrony: Two mechanisms in the snowy tree cricket. *Science*, 166(3907):891–894, 1969. doi:10.1126/science.166.3907.891.

G. G. Walter and M. Contreras. *Compartmental Modeling with Networks*. Birkhäuser, 1999. ISBN 0817640193. doi:10.1007/978-1-4612-1590-5.

J. Wang and N. Elia. Control approach to distributed optimization. In *Allerton Conf. on Communications, Control and Computing*, pages 557–561, Monticello, IL, USA, 2010. doi:10.1109/ALLERTON.2010.5706956.

Y. Wang, D. Chakrabarti, C. Wang, and C. Faloutsos. Epidemic spreading in real networks: An eigenvalue viewpoint. In *IEEE Int. Symposium on Reliable Distributed Systems*, pages 25–34, Florence, Italy, Oct. 2003. doi:10.1109/RELDIS.2003.1238052.

A. Watton and D. W. Kydon. Analytical aspects of the N-bug problem. *American Journal of Physics*, 37(2): 220–221, 1969. doi:10.1119/1.1975458.

J. T. Wen and M. Arcak. A unifying passivity framework for network flow control. *IEEE Transactions on Automatic Control*, 49(2):162–174, 2004. doi:10.1109/TAC.2003.822858.

N. L. White and W. Whiteley. The algebraic geometry of stresses in frameworks. *SIAM Journal on Algebraic and Discrete Methods*, 4(4):481–511, 1983. doi:10.1137/0604049.

A. T. Winfree. Biological rhythms and the behavior of populations of coupled oscillators. *Journal of Theoretical Biology*, 16(1):15–42, 1967. doi:10.1016/0022-5193(67)90051-3.

J. Wolfowitz. Product of indecomposable, aperiodic, stochastic matrices. *Proceedings of American Mathematical Society*, 14(5):733–737, 1963. doi:10.1090/S0002-9939-1963-0154756-3.

C. W. Wu and L. O. Chua. Application of Kronecker products to the analysis of systems with uniform linear coupling. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, 42(10): 775–778, 1995. doi:10.1109/81.473586.

T. Xia and L. Scardovi. Output-feedback synchronizability of linear time-invariant systems. *Systems & Control Letters*, 94:152–158, 2016. doi:10.1016/j.sysconle.2016.06.007.

W. Xia and M. Cao. Sarymsakov matrices and asynchronous implementation of distributed coordination algorithms. *IEEE Transactions on Automatic Control*, 59(8):2228–2233, 2014. doi:10.1109/TAC.2014.2301571.

W. Xia and M. Cao. Analysis and applications of spectral properties of grounded Laplacian matrices for directed networks. *Automatica*, 80:10–16, 2017. doi:10.1016/j.automatica.2017.01.009.

L. Xiao and S. Boyd. Fast linear iterations for distributed averaging. *Systems & Control Letters*, 53:65–78, 2004. doi:10.1016/j.sysconle.2004.02.022.

L. Xiao, S. Boyd, and S. Lall. A scheme for robust distributed sensor fusion based on average consensus. In *Symposium on Information Processing of Sensor Networks*, pages 63–70, Los Angeles, CA, USA, Apr. 2005. doi:10.1109/IPSN.2005.1440896.

L. Xiao, S. Boyd, and S.-J. Kim. Distributed average consensus with least-mean-square deviation. *Journal of Parallel and Distributed Computing*, 67(1):33–46, 2007. doi:10.1016/j.jpdc.2006.08.010.

T. Yang, S. Roy, Y. Yan, and A. Saberi. Constructing consensus controllers for networks with identical general linear agents. *International Journal of Robust and Nonlinear Control*, 21(11):1237–1256, 2011. doi:10.1002/rnc.1641.

R. A. York and R. C. Compton. Quasi-optical power combining using mutually synchronized oscillator arrays. *IEEE Transactions on Microwave Theory and Techniques*, 39(6):1000–1009, 1991. doi:10.1109/22.81670.

M. Youssef and C. Scoglio. An individual-based approach to SIR epidemics in contact networks. *Journal of Theoretical Biology*, 283(1):136–144, 2011. doi:10.1016/j.jtbi.2011.05.029.

W. Yu, G. Chen, and M. Cao. Some necessary and sufficient conditions for second-order consensus in multi-agent dynamical systems. *Automatica*, 46(6):1089–1095, 2010. doi:10.1016/j.automatica.2010.03.006.

D. Zelazo. *Graph-Theoretic Methods for the Analysis and Synthesis of Networked Dynamic Systems*. PhD thesis, University of Washington, Aug. 2009.

D. Zelazo and M. Mesbahi. Edge agreement: Graph-theoretic performance bounds and passivity analysis. *IEEE Transactions on Automatic Control*, 56(3):544–555, 2011. doi:10.1109/TAC.2010.2056730.

D. Zelazo, S. Schuler, and F. Allgöwer. Performance and design of cycles in consensus networks. *Systems & Control Letters*, 62(1):85–96, 2013. doi:10.1016/j.sysconle.2012.10.014.

Y. Zhang and Y. P. Tian. Consentability and protocol design of multi-agent systems with stochastic switching topology. *Automatica*, 45:1195–1201, 2009. doi:10.1016/j.automatica.2008.11.005.

W. X. Zhao, H. F. Chen, and H. T. Fang. Convergence of distributed randomized PageRank algorithms. *IEEE Transactions on Automatic Control*, 58(12):3255–3259, 2013. doi:10.1109/TAC.2013.2264553.

J. Zhu, Y. Tian, and J. Kuang. On the general consensus protocol of multi-agent systems with double-integrator dynamics. *Linear Algebra and its Applications*, 431(5-7):701–715, 2009. doi:10.1016/j.laa.2009.03.019.