Chapter 2
Averaging in Time-Invariant Networks

Abstract This chapter studies the basic averaging dynamics on a fixed network. This linear dynamics is also called “consensus” dynamics, because under suitable assumptions it brings the states associated with the nodes to converge to the same value. Section 2.1 introduces the rendezvous problem, which serves us as the main motivation to seek consensus, and states the main results of the chapter. Section 2.2 solves the consensus problem in the special case of symmetric regular graphs, while the general solution, which is based on the notion of stochastic matrix, is presented in Sect. 2.3. The subsequent sections provide further insights into the averaging dynamics, namely about its speed of convergence (Sects. 2.4 and 2.7) and its consensus value (Sect. 2.5). Meanwhile, Sect. 2.6 presents some classical examples of stochastic matrices associated with a graph, such as simple random walks and Metropolis walks. Finally, Sect. 2.8 concentrates on reversible stochastic matrices and their properties.

2.1 Rendezvous and Consensus

One of the simplest examples of coordinated control is the so-called *rendezvous problem*. Assume that units have dynamics of type $x_v(t+1) = x_v(t) + u_v(t)$ for all $t \in \mathbb{Z}_{\geq 0}$ with $x_v(t)$ and $u_v(t) \in \mathbb{R}^n$ for all $v \in V$ and that the control goal is to make all units converge their state to the same point. We can think of them as moving agents with the state representing position. This is known as the *rendezvous problem*: There are many variants of this problem, and the one we are addressing is just the basic and simplest instance. What are exactly the issues we want to analyze? Here is a brief list:

(a) Given a graph G, find out whether there exists a control scheme $u_v = g_v(x)$ adapted to G such that the state evolutions governed by the equations $x_v(t+1) = x_v(t) + g_v(x)$ all converge to the same point, namely for all initial conditions $x(0)$, there exists $x^\star \in \mathbb{R}^n$ such that

$$\lim_{t\to+\infty} x_v(t) = x^\star\,, \quad \forall v \in V. \tag{2.1}$$

F. Fagnani and P. Frasca, *Introduction to Averaging Dynamics over Networks*, Lecture Notes in Control and Information Sciences 472,
https://doi.org/10.1007/978-3-319-68022-4_2

(b) In case when (a) has a positive answer, we would like to find effective ways for producing the control scheme. Indeed, in general, there will be many possible control schemes and the choice can be dictated to optimize certain performance indices:

(b1) the velocity of convergence to the rendezvous point;
(b2) the displacement of $x^\star$ from the initial condition.

Both indices will be defined precisely later on.

Notice that without further assumptions, the problem as stated in (a) is always solvable and with no communication among units. It is sufficient to put $u_v = -x_v$, and we will have that $x_v(t) = 0$ for all v and for all $t \geq 1$: In control theory, this is known as a "deadbeat control." The reason why this is not a feasible solution is the following. This solution implicitly requires that units have already agreed to make 0 their rendezvous point, and in other terms, they have already coordinated off-line. This prior coordination is something we want to avoid; moreover, the origin may be far off from their initial condition and thus an unreasonable choice (in general not optimizing (b2)). We make the following extra assumption on the rendezvous point $x^\star$ which automatically drops out the deadbeat control scheme above: We require that, translating all initial conditions $x_v(0) \to x_v(0) + b$ with the same vector, also the rendezvous point translates the same way $x^\star \to x^\star + b$. We will refer to this as to the *translation invariance requirement*.

As it is customary in control theory, it is natural to seek, *in primis*, a linear solution to this problem, namely to consider controllers of type

$$u_v(t) = \sum_{w \in V} K_{vw} x_w(t) \tag{2.2}$$

where $K \in \mathbb{R}^{V \times V}$ is a gain matrix. Coupling with the unit dynamics, we thus obtain

$$x_v(t+1) = \sum_{w \in V} P_{vw} x_w(t) \tag{2.3}$$

where $P = I + K$.

This type of models (2.3) has applications much broader than just in the rendezvous problem for mobile agents. Instead of a position, the state $x_v(t)$ can as well be interpreted as an estimation or as an opinion on some fact possessed by unit v at time t and the common convergence to the same value is a phenomenon known as *consensus*. Later on, we will provide more details on such possible applicative contexts.

Notice that the dimension of the state does not play any particular role in the dynamics (2.3) as all components of the state vectors $x_v(t)$ evolve separately all with the same dynamics given by the matrix P. For this reason, from now on, we will assume that the state $x_v(t)$ of each unit is one-dimensional, namely a *scalar*. In this setting, (2.3) can be rewritten in more compact form simply as

$$x(t+1) = Px(t) \tag{2.4}$$

so that $x(t) = P^t x(0)$. The translation invariance, in this context, amounts to require that $P^t \mathbf{1} \to \mathbf{1}$ for $t \to +\infty$. Since $P^{t+1}\mathbf{1} = PP^t\mathbf{1}$ then converges both to $\mathbf{1}$ and to $P\mathbf{1}$, the translation invariance is also equivalent to require $P\mathbf{1} = \mathbf{1}$ (each row of P sums to 1).

Notice moreover that the feedback law (2.2) is adapted to G if K (or equivalently P) is adapted to G. Therefore, in order to exhibit a solution to the rendezvous problem with translation invariance, it is sufficient to exhibit $P \in \mathbb{R}^{V\times V}$ adapted to G such that $P\mathbf{1} = \mathbf{1}$. The following result, which will be proven in the next sections, is an elegant and simple solution.

Theorem 2.1 (Consensus) *Suppose G has a globally reachable vertex $v^\star$. Then the rendezvous problem with the translation-invariant requirement is solvable over G. A possible solution is given by any matrix $P \in \mathbb{R}^{V\times V}$ satisfying the following properties:*

(Pa) $P_{vw} \geq 0$ *for every $v, w \in V$;*
(Pb) $P\mathbf{1} = \mathbf{1}$*;*
(Pc) *For every $v \neq w$, $P_{vw} > 0 \Leftrightarrow (v, w) \in E$;*
(Pd) $P_{v^\star v^\star} > 0$.

It turns out that matrices as P sharing (Pa) and (Pb) have very special properties: They are called *stochastic* and appear in many different contexts, one of these being Markov chain theory. Property (Pc) says that $\mathcal{G}_P$ and G can only possibly differ in their self-loops.

There is an additional nice property of these systems. Being P stochastic, its Laplacian is $L(P) = I - P$. Consequently, we may write (2.4) as $x(t+1) = x(t) - L(P)x(t)$, which becomes $x_v(t+1) = x_v(t) + \sum_w P_{vw}(x_w(t) - x_v(t))$ componentwise. We observe that this expression only involves the state of v and differences between the states of v and of its neighbors w. Then, there is no need for the nodes to exchange information in an absolute reference frame, but only relative information suffices.

Before presenting the key results for stochastic matrices and proving Theorem 2.1 and some generalizations, we will work out a special case, which explains how matrices like P above come naturally into the picture.

2.2 Averaging on Symmetric Regular Graphs

Notice that if the underlying graph was the complete one, the rendezvous problem would have a very simple solution: It would be enough for all the units to compute the barycenter $\overline{x}(t) := N^{-1}\sum_{v\in V} x_v(t)$ (where $N := |V|$) and implement the control law $u_v(t) = \overline{x}(t) - x_v(t)$ which would yield $x_v(t) = \overline{x}(t)$ for all v and all positive t. This law implies that at time $t = 1$, all units have already reached consensus exactly in the barycenter of the initial state $\overline{x}(0)$. It is then immediate to see that $x_v(t) = \overline{x}(0)$ for every $t \geq 1$. The type of matrix P we obtain in this case is $P = N^{-1}\mathbf{1}\mathbf{1}^*$, a very special stochastic matrix with all elements equal to $1/N$.

This solution is not admissible for a general graph, but its main idea can be adapted. Indeed, it is sufficient to replace the barycenter $\overline{x}(t)$ with a local version of it, namely for each unit to use a local barycenter based on the units to which it is connected through the graph. Precisely, given a graph $G = (V, E)$, each unit $v \in V$ computes at time t

$$\overline{x}_v(t) := \frac{1}{d_v^{\text{out}}} \sum_{w \in V} (A_G)_{vw} x_w(t)$$

and implements the dynamics $x_v(t+1) = x_v(t) + \tau(\overline{x}_v(t) - x_v(t))$. The parameter $\tau > 0$ indicates the velocity at which unit v is following the local barycenter and will play a crucial role in the rest of this section. In compact matrix form, we obtain that $x(t+1) = Px(t)$ where

$$P = I + \tau(D_G^{-1} A_G - I) = I - \tau D_G^{-1} L(G). \tag{2.5}$$

It is easy to see that $\tau \in (0, 1]$ guarantees that P is a stochastic matrix adapted to the graph G.

Let us now analyze the special case when G is *symmetric and d-regular*. In this case, $P = I - \tau d^{-1} L(G)$ is also symmetric. Assuming that $0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_N$ are the eigenvalues of the Laplacian $L(G)$, we obtain that the eigenvalues of P are simply given by $\mu_i = 1 - \tau d^{-1}\lambda_i$ (with $\mu_1 = 1$). Moreover, the two matrices $L(G)$ and P share the same orthonormal basis of eigenvectors ξ_i's (with $\xi_1 = N^{-1/2}\mathbf{1}$). We can thus write the usual orthonormal decomposition of P

$$P = \sum_{i=1}^{N} \mu_i \xi_i \xi_i^* = N^{-1}\mathbf{1}\mathbf{1}^* + \sum_{i=2}^{N} \mu_i \xi_i \xi_i^*$$

which yields, by orthonormality, $P^t = N^{-1}\mathbf{1}\mathbf{1}^* + \sum_{i=2}^{N} \mu_i^t \xi_i \xi_i^*$. The evolution of the state configuration is thus

$$P^t x(0) = \bar{x}(0)\mathbf{1} + \sum_{i=2}^{N} \mu_i^t \xi_i \xi_i^* x(0)$$

Notice now that

$$||P^t x(0) - \bar{x}(0)\mathbf{1}||^2 = \left\|\sum_{i=2}^{N} \mu_i^t \xi_i \xi_i^* x(0)\right\|^2 = \sum_{i=2}^{N} |\mu_i|^{2t} |\xi_i^* x(0)|^2$$

Since $1 = \mu_1 \geq \mu_2 \geq \cdots \geq \mu_N$, putting $\rho_2 := \max\{|\mu_2|, |\mu_N|\}$, we obtain

$$||P^t x(0) - \bar{x}(0)\mathbf{1}||^2 \leq \rho_2^{2t} \sum_{i=2}^{N} |\xi_i^* x(0)|^2 \leq \rho_2^{2t} ||x(0)||^2,$$

which can be rewritten as

$$||P^t x(0) - \bar{x}(0)\mathbf{1}|| \leq \rho_2^t ||(I - N^{-1}\mathbf{1}\mathbf{1}^\top)x(0)||. \tag{2.6}$$

This bound shows that if $\rho_2 < 1$, then $x(t) = P^t x(0) \to \bar{x}(0)\mathbf{1}$ for $t \to +\infty$, namely all states converge to a consensus point, which turns out to be again the barycenter of the initial state conditions $\bar{x}(0)$. Moreover, (2.6) actually shows that ρ_2 dictates the speed of convergence of the dynamics toward consensus. Under which conditions can we guarantee that $\rho_2 < 1$? Because of the way ρ_2 is defined, we must have $|\mu_2|, |\mu_N| < 1$. If G is not connected, we know that $\lambda_2 = 0$ and, consequently, $\mu_2 = 1$: Indeed, in this case, it is clear that consensus can not be reached in general since the network is composed of completely separated components. Instead, if G is connected, then $\lambda_2 > 0$ and, consequently, $1 > \mu_2 \geq \mu_N$. Hence, the only extra condition that needs to be satisfied is $\mu_N > -1$, namely $1 - \tau d^{-1}\lambda_N > -1$. This is equivalent to $\tau < \frac{2d}{\lambda_N}$. Considering that (see Exercise 1.17) $\lambda_N \leq 2d$, a sufficient condition which guarantees consensus is $\tau < 1$. We can summarize the above discussion in the following result.

Proposition 2.1 (Consensus on symmetric regular graphs) *Let G be a symmetric, d-regular, and connected graph. Then, the dynamics* (2.4)–(2.5)*, with $\tau \in (0, 1)$, guarantees convergence to consensus, where the consensus point is the barycenter of the initial state and convergence happens at an exponential rate given by ρ_2.*

In the next section, we will present a number of general results on stochastic matrices and we will be able to generalize this result to more general graphs, by dropping the assumptions of symmetry, regularity, and—to a certain extent—connectivity of the underlying graph.

2.3 Stochastic Matrices and Averaging

In general, a matrix $P \in \mathbb{R}^{V \times V}$ such that $P_{vw} \geq 0$ for every $v, w \in V$ is called a *nonnegative matrix*. A nonnegative matrix $P \in \mathbb{R}^{V \times V}$ satisfying the row sum condition $P\mathbf{1} = \mathbf{1}$ is said to be a *stochastic matrix*. With these new concepts, we can restate properties (Pa)-(Pb)-(Pc) above by saying that P is a stochastic matrix adapted to G.

As already noticed, P behaves as a local averaging operator: Given a vector $x \in \mathbb{R}^V$, the component v of Px is a weighted average of the values x_w for $w \in \mathcal{N}_v^{\text{out}}$. There is also an interesting *flux* interpretation of the adjoint operator. Given $\zeta \in \mathbb{R}^V$, $(\zeta^* P)_v = \sum \zeta_w P_{wv}$ can be interpreted as follows: From each node w, the quantity ζ_w will flow through the outgoing edges splitting according to the weights P_{wv} as v varies among the out neighbors of w. Hence, $\sum \zeta_w P_{wv}$ is the total new quantity present at node v.

Moreover, a stochastic matrix is the main ingredient of a Markov chain, a special stochastic process such that the future only depends on the past through the present state and states are finite in number. Given a stochastic matrix $P \in \mathbb{R}^{V \times V}$, the term P_{vw} can be interpreted as the probability of making a transition from state v to state w: If you associate each state with the node of the associated graph $\mathcal{G}_P$, you can imagine to be sitting at state v and to walk along one of the available outgoing edges from v according to the various probabilities P_{vw}. In this way, you construct what is called a random walk on the graph G. In this probabilistic setting, flows can be interpreted as probabilities: If $\zeta \in \mathbb{R}^V$ is a probability vector where ζ_v indicates the probability that at the initial instant the state is equal to v, then $(\zeta^* P)_v$ indicates the probability of finding the process in state v at the next time.

The first general observation to be done on stochastic matrices is that the set of stochastic matrices is closed under a number of important operations (whose elementary proof is left to the reader):

(1) If $P, Q \in \mathbb{R}^{V \times V}$ are stochastic, then $\lambda P + (1 - \lambda) Q$ is stochastic for any $\lambda \in (0, 1)$.
(2) If $P, Q \in \mathbb{R}^{V \times V}$ are stochastic, then PQ is stochastic. In particular, P^t is stochastic, for any $t \in \mathbb{N}$.
(3) If P_n is a sequence of stochastic matrices such that $P_n \to P$ for $n \to +\infty$, then P is stochastic.

Properties (1) and (3) say that the set of stochastic matrices form a compact convex subset of $[0, 1]^{V \times V}$.

We are now almost ready to state and prove the main result of this chapter, which investigates the behavior of the powers of a stochastic matrix, proposing minimal conditions to get convergence. The proof is based on the following lemma, which shall also be used later in these notes.

Lemma 2.1 (Contraction principle) *Let $Q \in \mathbb{R}^{V \times V}$ be a stochastic matrix such that there exist $\alpha > 0$ and $m \in V$ such that $Q_{vm} \geq \alpha$ for all $v \in V$. Then, for all $x \in \mathbb{R}^V$, it holds true that $y = Qx$ satisfies*

$$\max_{v \in V} y_v - \min_{v \in V} y_v \leq (1 - \alpha)\big(\max_{v \in V} x_v - \min_{v \in V} x_v\big)$$

Proof Note that

$$\begin{aligned}
y_v = \sum_{w\in V} Q_{vw}x_w &= \sum_{w\in V} Q_{vw}(x_w - \min_{u\in V} x_u) + \sum_{w\in V} Q_{vw} \min_{u\in V} x_u \\
&\geq \alpha(x_m - \min_{u\in V} x_u) + \min_{u\in V} x_u \\
&= \alpha x_m + (1-\alpha) \min_{u\in V} x_u.
\end{aligned}$$

Similarly,

$$\begin{aligned}
y_v = \sum_{w\in V} Q_{vw}x_w &= \sum_{w\in V} Q_{vw}(x_w - \max_{u\in V} x_u) + \sum_{w\in V} Q_{vw} \max_{u\in V} x_u \\
&\leq \alpha(x_m - \max_{u\in V} x_u) + \max_{u\in V} x_u \\
&= \alpha x_m + (1-\alpha) \max_{u\in V} x_u.
\end{aligned}$$

Putting these two inequalities together gives:

$$\begin{aligned}
\max_{u\in V} y_u - \min_{u\in V} y_u &\leq \alpha x_m + (1-\alpha)\max_{u\in V} x_u - \alpha x_m - (1-\alpha)\min_{u\in V} x_u \\
&= (1-\alpha)(\max_{u\in V} x_u - \min_{u\in V} x_u),
\end{aligned}$$

that is the thesis. □

The main result is then the following.

Theorem 2.2 (Convergence to consensus) *Let $P \in \mathbb{R}^{V\times V}$ be a stochastic matrix such that $\mathcal{G}_P$ admits a globally reachable aperiodic vertex. Then, the following two equivalent facts hold true.*

(i) The dynamics (2.4) is such that, for any initial condition $x(0) = x^0 \in \mathbb{R}^V$, there exists a scalar α such that

$$x(t) = P^t x(0) \to \alpha\mathbf{1} \qquad t \to +\infty.$$

In other terms, all components $x_v(t)$ converge to the same consensus value α.

(ii) There exists a vector $\pi \in \mathbb{R}^V$ such that $\pi_v \geq 0$ for all v, $\sum_v \pi_v = 1$, and

$$\lim_{t\to+\infty} P^t = \mathbf{1}\pi^*. \tag{2.7}$$

In other terms, P^t converges to a matrix having all rows equal to the row vector π^*.

Furthermore, $\alpha = \pi^* x(0)$.

Let $s \in V$ be the aperiodic vertex which is reachable from all others. This means that there exists $t^\star \in \mathbb{N}$ such that $Q := P^{t^\star}$ is such that $Q_{vs} > 0$ for all $v \in V$. Let $\alpha = \min\{Q_{vs} \,:\, v \in V\} > 0$. Then, letting $y^0 \in \mathbb{R}^V$ and $y^1 = Qy^0$, Lemma 2.1 implies that

$$\max_{v\in V} y^1_v - \min_{v\in V} y^1_v \leq (1-\alpha)(\max_{v\in V} y^0_v - \min_{v\in V} y^0_v).$$

Fix now $x(0) = x^0 \in \mathbb{R}^V$ arbitrarily and consider (2.4). Define $M_t = \max_{v\in V} x_v(t)$ and $m_t = \min_{v\in V} x_v(t)$, and notice that, since the components of $x(t)$ are convex combinations of those of $x(t-1)$, the sequences M_t and m_t are bounded and, respectively, nonincreasing and nondecreasing (hence convergent). Hence, also $\Delta_t = M_t - m_t$ converges. For the previous argument, moreover, it holds that $\Delta_{nt^\star} \leq (1-\alpha)^n \Delta_0$. This implies that $\Delta_{nt^\star} \to 0$ for $n \to +\infty$. Hence, all components of $x(t)$ will converge to the same limit, thus proving the first claim. If we apply this result choosing $x(0) = e_w$, the wth element of the canonical basis of $\mathbb{R}^V$, we obtain that all elements of the wth column of P^t will converge to the same limit. This clearly yields the second claim.

□

Theorem 2.2 immediately yields Theorem 2.1. An important special case is discussed in the following remark.

Remark 2.1 (*Irreducibility*) A stochastic matrix P for which $\mathcal{G}_P$ is strongly connected is called *irreducible*. A stochastic matrix is said to be aperiodic if $\mathcal{G}_P$ is aperiodic. Hence, Theorem 2.2 applies to the important case when P is irreducible and aperiodic. Notice that for symmetric P these two properties are equivalent to the assumptions in Theorem 2.2.

We now briefly discuss the *necessity* of the assumptions in Theorem 2.2.

Remark 2.2 (*Aperiodicity*) Notice that it is not necessary that all units have access to their own state. It is instead sufficient that the globally reachable node is aperiodic; hence, for instance, it is sufficient that there is a self-loop in this node. The fact that some assumption of aperiodicity is necessary for convergence follows by considering the simple example of a strongly connected graph with two nodes and no self-loops. The only possible stochastic matrix adapted to such a graph is

$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

Notice that $P^{2t} = I$ for all t, and therefore, P does not yield consensus. A few properties of periodic matrices are discussed in Exercise 2.3.

Remark 2.3 (*Connectivity*) If the graph $\mathcal{G}_P$ has two (or more) sinks, the matrix P can be written with the block structure

$$P = \begin{bmatrix} R_1 & R_2 & R_3 \\ 0 & P_1 & 0 \\ 0 & 0 & P_2 \end{bmatrix}$$

Since the powers of P inherit its block structure, the entries of no column of P^t can converge to the same value in general. This reasoning shows that *global reachability of a node is necessary for convergence to consensus*.

The convergence theorem is illustrated by the following example and in Fig. 2.1.

Example 2.1 (*An irreducible, aperiodic stochastic matrix*) Consider the stochastic matrix

$$P = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 0 & 2/3 \end{bmatrix}$$

It is evident that P is irreducible and aperiodic. Let us compute the invariant probability π. From $\pi^* P = \pi^*$, we get

$$\begin{cases} -\frac{1}{2}\pi_1 + \frac{1}{3}\pi_2 + \frac{1}{3}\pi_3 = 0 \\ \frac{1}{2}\pi_1 - \frac{2}{3}\pi_2 = 0 \\ \frac{1}{3}\pi_2 - \frac{1}{3}\pi_3 = 0 \end{cases}$$

which immediately yields $\pi_2 = \pi_3$ and $\pi_1 = \frac{4}{3}\pi_2$. Using the normalization condition $\pi_1 + \pi_2 + \pi_3 = 1$, we finally get $\pi = \left(\frac{2}{5}, \frac{3}{10}, \frac{3}{10}\right)^*$.

Theorem 2.2 also contains further information useful to address issue (b) presented at the beginning of the chapter. We shall make this information explicit in the next two results, as well as in the following sections. The first result is about the spectrum of the update matrix.

Corollary 2.1 *Let $P \in \mathbb{R}^{V \times V}$ be a stochastic matrix such that $\mathcal{G}_P$ admits a globally reachable aperiodic node. Then,*

(i) 1 *is an algebraically simple eigenvalue whose eigenspace is generated by* $\mathbf{1}$*;*

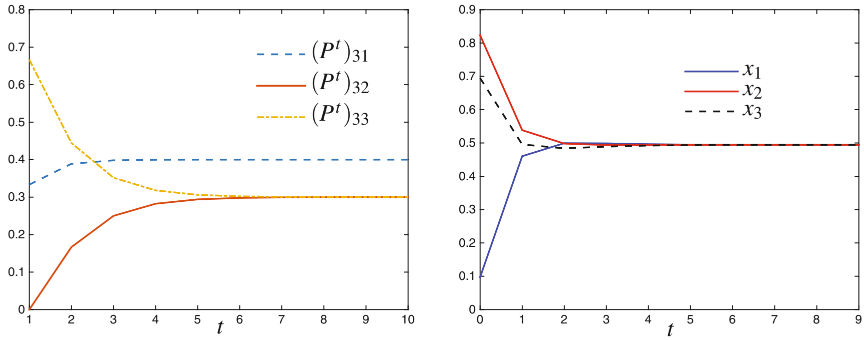


Fig. 2.1 Illustration of convergence for Example 2.1. The *left* diagram plots the entries of the third row of P^t that converges to π^*. The *right* diagram plots the associated averaging dynamics (2.4) from random initial conditions within (0, 1)

(ii) Any other eigenvalue μ of P is such that $|\mu| < 1$.

Proof Suppose indeed that P satisfies the assumptions of Theorem 2.2 and let $\xi \in \mathbb{R}^V$ be an eigenvector of P with eigenvalue μ. Then, for $t \to +\infty$,

$$\mu^t \xi = P^t \xi \to \mathbf{1}\pi^* \xi$$

This immediately yields that either $\mu = 1$ and ξ is a multiple of $\mathbf{1}$, or $|\mu| < 1$. This remark yields (ii) and says that 1 is a geometrically simple eigenvalue (the corresponding eigenspace has dimension 1). It remains to show that 1 is also algebraically simple. This follows using similar arguments showing that the presence of a nontrivial Jordan block relative to the eigenvalue 1 will imply that P^t would grow unbounded contrarily to what is asserted in Theorem 2.2. □

The second result further investigates the structure of the limit matrix.

Corollary 2.2 *Let $P \in \mathbb{R}^{V \times V}$ be a stochastic matrix such that $\mathcal{G}_P$ admits a globally reachable aperiodic node. Consider the vector π as in Theorem 2.2. Then, $\pi^* P = \pi^*$, and π is the only vector sharing this property and the normalization condition $\sum_v \pi_v = 1$.*

Proof A very well-known fact of linear algebra says that P and P^* have the same eigenvalues. This implies that there must exists $\zeta \in \mathbb{R}^V$ such that $\zeta^* P = \zeta^*$. This yields, for $t \to +\infty$,

$$\zeta^* = \zeta^* P^t \to \zeta^* \mathbf{1}\pi^*$$

Hence, ζ is necessarily a multiple of π. In other words, this shows that π is a left eigenvalue of P relative to the eigenvalue 1. Since 1 is also algebraically simple as a left eigenvalue, the uniqueness result immediately follows. □

In the flux interpretation presented at the beginning of this section, the equation $\pi^* P = \pi^*$ can be interpreted as a "stationary regime": The flux is not modifying the quantity π_v present in every node. For this reason, and because of the normalization to 1, π is called *stationary* or *invariant probability measure*. Note that the invariant probability measure needs not to be unique in general—find an example as an exercise—, but is unique when there is a globally reachable node; see Exercise 2.3.

2.4 Convergence Rate and Eigenvalues

This section deals more precisely with question (b1) defined at the beginning of this chapter, that is with convergence speed. The speed of convergence of (2.7) is dictated by the magnitude of the eigenvalues of P. We start by recalling the following result, which is a standard fact in the stability of linear dynamical systems.

Lemma 2.2 *Let $M \in \mathbb{R}^{V\times V}$ be any matrix and let λ_i be its eigenvalues. Let $\rho = \max|\lambda_i|$ be the spectral radius of M. Then, for every $\epsilon > 0$, there exists a constant C_ϵ such that*

$$||M^t x_0||_2 \le C_\epsilon(\rho+\epsilon)^t ||x_0||_2 \quad \textit{for all } t.$$

A simple application of this lemma allows us to obtain the following result:

Proposition 2.2 (Convergence rate) *Let $P \in \mathbb{R}^{V\times V}$ be a stochastic matrix such that $\mathcal{G}_P$ admits a globally reachable aperiodic node. Consider all its eigenvalues μ_i but 1 and put $\rho_2 = \max\{|\mu_i| < 1\}$. Then, for every $\epsilon > 0$, there exists a constant C_ϵ such that*

$$||(P^t - \mathbf{1}\pi^*)x_0||_2 \le C_\epsilon(\rho_2+\epsilon)^t ||x_0||_2 \quad \textit{for all } t.$$

Proof Put $Q := P - \mathbf{1}\pi^*$, and notice that $Q^t = P^t - \mathbf{1}\pi^*$ (check this for exercise). Notice moreover that $Q\mathbf{1} = 0$. Consider now the subspace W of $\mathbb{R}^V$ orthogonal to the vector π, and notice that if $w \in W$, then $Pw = Qw$ and $\pi^* Pw = \pi^* w = 0$. In other terms, the subspace W is invariant for P and Q, and on W, the two matrices P and Q coincide. The eigenvalues of P and Q on W are exactly given by the eigenvalues of P different from 1. Wrapping up, we have that Q has eigenvalues μ_i plus the eigenvalue 0; hence, it is asymptotically stable, and the result follows from Lemma 2.2. □

The parameter ρ_2, introduced in the statement of the corollary above, is also called the *second eigenvalue* of P, and the difference $1-\rho_2$ the *spectral gap* of P. The above result essentially says that convergence to rendezvous happens exponentially fast as ρ_2^t. Actually, this is only approximately true because of the arbitrarily small ϵ we have to fix. The ϵ is needed because of the possible presence in P of nontrivial Jordan blocks (which is when the algebraic and geometric dimension of an eigenspace does not coincide). When P is symmetric, things are much simpler: We can indeed follow the proof of Equation (2.6) above and prove the following result which extends Proposition 2.1 to general symmetric matrices P, possibly adapted to nonregular graphs.

Corollary 2.3 (Convergence rate for symmetric P) *Let $P \in \mathbb{R}^{V\times V}$ be a symmetric stochastic matrix such that $\mathcal{G}_P$ is strongly connected and aperiodic. Then,*

$$||(P^t - N^{-1}\mathbf{1}\mathbf{1}^*)x_0||_2 \le \rho_2^t ||x_0||_2$$

Example 2.2 (*An irreducible, aperiodic stochastic matrix*) Consider the stochastic matrix P defined in Example 2.1. An easy computation shows that the characteristic polynomial of P is given by

$$p(\lambda) := \det(\lambda I - P) = (\lambda - 1)(\lambda - 1/6)^2$$

Therefore, $\rho_2 = 1/6$.

Once we have established that ρ_2 is the right parameter to analyze the speed of convergence, it remains to understand how it depends on the graph: This analysis will be done, on certain families, in Sect. 2.7.

2.5 Consensus Point

Another important question—mentioned in (b2) at the beginning of this chapter—regards the location of the consensus point with respect to the initial condition. As we know from Theorem 2.2, this is completely determined by the left eigenvector π of P. First, notice that the optimization problem:

$$\min_{y\in\mathbb{R}} \sum_{v\in V} |y - y_v|^2$$

has solution given by the barycenter $y = N^{-1} \sum_v y_v$. The rendezvous problem has the barycenter as meeting point if and only if $\pi = N^{-1}\mathbf{1}$. When will this happen? The answer is very simple: if and only if $\mathbf{1}^* P = \mathbf{1}^*$, namely if all columns of P also sum up to 1. When this happens, P is called a *doubly stochastic* matrix. A particular case is when P is symmetric.

What about the possibility to construct a doubly stochastic matrix over a preassigned graph? Is that always possible? The answer is on the negative. Before showing this fact, we introduce another concept which will also be useful later on.

Definition 2.1 (*Sub-stochastic matrix*) A nonnegative matrix $P \in \mathbb{R}^{V\times V}$ is said to be *sub-stochastic* if $\sum_w P_{vw} \leq 1$ for all $v \in V$, and there exists at least one $v \in V$ for which the inequality is strict. Such node will be called a *deficiency node* of P.

There are a few useful facts about sub-stochastic matrices, which the reader is encouraged to verify on his/her own and which are gathered in the following proposition.

Proposition 2.3 (Sub-stochastic matrices) *Let $P \in \mathbb{R}^{V\times V}$ be a sub-stochastic matrix.*

- *Then, P^t is sub-stochastic for all t. More precisely, if we let $V_t^\star$ to be the set of deficiency nodes of P^t, then*

$$V_t^\star \subseteq V_{t+1}^\star \qquad \textit{for all } t \in \mathbb{N}.$$

- *If, moreover, P is such that from every node v there is a path in $\mathcal{G}_P$ to a deficiency node, then there exists $t^\star$ such that $V_{t^\star}^\star = V$ (all nodes for $P^{t^\star}$ are deficiency nodes).*

Actually, this fact implies a simple condition for the stability of sub-stochastic matrices.

Proposition 2.4 (Stability of sub-stochastic matrices) *Let $P \in \mathbb{R}^{V\times V}$ be a sub-stochastic matrix such that from every node v there is a path in $\mathcal{G}_P$ to a deficiency node. Then, P is asymptotically stable.*

Proof Let $t^\star$ be defined as in Proposition 2.3 and let $\nu = \max_v \sum_w P^{t^\star}_{vw} < 1$. Given any $t \in \mathbb{N}$, write $t = nt^\star + r$ with $r \in \{0, \dots, t^\star - 1\}$ and $n \in \mathbb{N}$, and notice that $P^t\mathbf{1} \leq P^{nt^\star}\mathbf{1} \leq \nu^n\mathbf{1}$ (where inequalities have to be understood componentwise). This inequality implies that P^t converges to 0. □

The following result characterizes the "zero pattern" of the invariant probability measure of stochastic matrices and shows that it is not always possible to construct a doubly stochastic matrix on a given graph.

Proposition 2.5 (Positivity of invariant probability measure) *Let $P \in \mathbb{R}^{V\times V}$ be a stochastic matrix such that $\mathcal{G}_P$ admits a globally reachable node $v^\star$. Let π be its invariant probability measure. Then, $\pi_v \neq 0$ if and only if v and $v^\star$ are in the same strongly connected component of $\mathcal{G}_P$.*

Proof Let $V^\star$ be the set of nodes corresponding to the connected component containing $v^\star$, and let $V^{\star\star} = V \setminus V^\star$. Ordering nodes in such a way that the first ones are those in $V^{\star\star}$, we get that P has the following block structure

$$P = \begin{bmatrix} Q & R \\ 0 & S \end{bmatrix}$$

where $Q \in \mathbb{R}^{V^{\star\star}\times V^{\star\star}}$, $R \in \mathbb{R}^{V^{\star\star}\times V^\star}$, and $S \in \mathbb{R}^{V^\star\times V^\star}$. By the assumption made, it follows that Q is sub-stochastic satisfying the assumptions of Proposition 2.4. On the other hand, P^t will have the following block structure:

$$P^t = \begin{bmatrix} Q^t & R_t \\ 0 & S^t \end{bmatrix}.$$

If we partition accordingly $\pi = (\pi^{\star\star}, \pi^\star)$, we then obtain $(\pi^{\star\star})^* Q^t = \pi^{\star\star}$ for all t. This yields $\pi^{\star\star} = 0$. We now prove that instead $\pi^\star_v > 0$ for every $v \in V^\star$. Assume, by contradiction, that there exists $w \in V^\star$ such that $\pi_w = 0$. The relation $\sum_{v\in V^\star} \pi_v P_{vw} = 0$ yields $\pi_v = 0$ for every $v \in N^{in}_w$. A straightforward inductive argument now shows that $\pi_v = 0$ for all $v \in V^*$ for which there exists a path from v to w. By the definition of V^*, this implies that $\pi_v = 0$ for all $v \in V^*$. But this says that π is the 0 vector and can not be an invariant probability. The proof is thus complete. □

This result implies that for a matrix to be doubly stochastic, its associated graph must be strongly connected.

2.6 Stochastic Matrices Adapted to a Graph

In this section, we focus on the problem of finding, given a graph, a stochastic/doubly stochastic matrix adapted to it. One solution is based on assigning equal weight to all outgoing edges of a node, similarly to what we did in (2.5):

$$P = D_G^{-1} A_G. \tag{2.8}$$

This matrix is known as the *simple random walk* (SRW) matrix associated with G. This name is explained by a probabilistic interpretation. Let us think of token that is performing a random walk on the nodes of the graph, according to the following rule: From each node, transitions happen with equal probability along all the available edges. Then, the rows of the SRW matrix are the transition probabilities from each node. More generally, we can consider, for $\tau \in (0, 1)$,

$$P = (1-\tau)I + \tau D_G^{-1} A_G,$$

which is also stochastic and is called the *lazy* SRW. If G contains a globally reachable aperiodic vertex, then $D_G^{-1} A_G$ yields consensus. On the other hand, even if the globally reachable vertex of G is not aperiodic, the lazy SRW yields consensus for any $\tau \in (0, 1)$ since aperiodicity is automatically gained from the presence of the identity part. Notice that P is not symmetric even when G is symmetric (unless G is also regular which was the case studied in Sect. 2.2). However, the case when G is symmetric is very special since in this case it is very simple to compute the invariant probability measure as $\pi_v = d_v/d$ where $d = \sum_u d_u$. Indeed,

$$(\pi^* P)_w = \sum_v \pi_v P_{vw} = \sum_{v \in \mathcal{N}_w} \frac{d_v}{d} \frac{1}{d_v} = \frac{|\mathcal{N}_w|}{d} = \pi_w.$$

Notice that the invariant measure for the SRW (or for the lazy version) is the uniform vector $N^{-1}\mathbf{1}$ if and only if G is regular. For symmetric nonregular graphs, there is however an alternative construction yielding a symmetric stochastic matrix. It is sufficient to define, for any $v \neq w$,

$$P_{vw} := (A_G)_{vw} \min\left\{\frac{1}{d_v}, \frac{1}{d_w}\right\}.$$

It is easy to see that, with this choice, the off-diagonal terms of any row of P sum up to a value which is not greater than 1. To complete, we define P on the diagonal terms in such a way to make it a stochastic matrix. Notice that P is symmetric by construction and is called the *Metropolis random walk*.Next, we present an example of Metropolis random walk.

Example 2.3 (*Random walks on a symmetric graph*) Let G be the graph represented in Fig. 2.2. This graph is not regular, and its degree matrix is $D = \text{diag}(2, 2, 2, 3, 1)$.

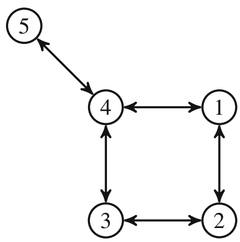


Fig. 2.2 The graph of Example 2.3

Then, the matrix corresponding to a SRW on G is

$$P = \begin{bmatrix} 1/2 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0&1&0&1&0 \\ 1&0&1&0&0 \\ 0&1&0&1&0 \\ 1&0&1&0&1 \\ 0&0&0&1&0 \end{bmatrix} = \begin{bmatrix} 0 & 1/2 & 0 & 1/2 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 1/3 & 0 & 1/3 & 0 & 1/3 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

Note that this matrix is not doubly stochastic and has -1 as an eigenvalue, because G is bipartite. Instead, the transition matrix of the Metropolis RW is

$$Q = \begin{bmatrix} 1/6 & 1/2 & 0 & 1/3 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 1/2 & 1/6 & 1/3 & 0 \\ 1/3 & 0 & 1/3 & 0 & 1/3 \\ 0 & 0 & 0 & 1/3 & 2/3 \end{bmatrix},$$

which is doubly stochastic and has second eigenvalue $\rho_2(Q) \simeq 0.7845$.

Metropolis construction produces a doubly stochastic matrix, but works for symmetric graphs only. On which graphs is it possible to construct a doubly stochastic matrix adapted to a generic graph? The following result gives us the answer.

Proposition 2.6 (Existence of doubly stochastic P) If $G = (V, E)$ is strongly connected, then there exists a doubly stochastic $P \in \mathbb{R}^{V \times V}$ such that $\mathcal{G}_P = G$.

Proof Given any circuit in G with edges

$$E' = \{(k_1, k_2), (k_2, k_3), \dots, (k_n, k_1)\} \qquad (k_i \neq k_j \text{ for } i \neq j),$$

consider the matrix $P^{(E')} \in \mathbb{R}^{V \times V}$ defined by

$$P^{(E')}_{vw} = \begin{cases} 1 & \text{if } (v, w) \in E' \\ 1 & \text{if } v = w \neq k_s \ \forall s = 1, \dots, n \\ 0 & \text{otherwise} \end{cases}$$

It is immediate to check that P has the following property: On each row and on each column, there is exactly one entry equal to 1, while all the others are equal to 0. This is what is called a *permutation matrix*, a very special case of doubly stochastic matrix. Now consider the family $\mathscr{D}$ of all possible circuits and the convex combination $P = \frac{1}{|\mathscr{D}|}\sum_{E'\in\mathscr{D}} P^{(E')}$: Clearly P is doubly stochastic and $\mathscr{G}_P \subseteq G$. Since G is strongly connected, any edge in G belongs to at least one of the subgraphs in $\mathscr{D}$ (check this as an exercise): This fact implies that $\mathscr{G}_P = G$ and the proof is complete. □

Example 2.4 (*Doubly stochastic matrix*) Consider the graph $G = (V, E)$ with $V = \{1, 2, 3\}$ and $E = \{(1, 2), (2, 3), (3, 1), (2, 1)\}$. Graph G is strongly connected and we know from the proof of Proposition 2.6 that a doubly stochastic matrix can be constructed as

$$P = \frac{1}{2}\left(\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}\right) = \begin{bmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 \end{bmatrix}.$$

The proof of Proposition 2.6 provides a method to construct a doubly stochastic matrix: The method can be easily applied on small graphs, but is not suitable to large graphs because one needs to find all possible circuits in G. In view of this difficulty, we become interested in matrices whose dominant left eigenvector is not $\mathbf{1}$, but some other vector that still guarantees a "balanced" consensus point. The following definitions go in this direction, identifying sequences of matrices that do not give "too much" weight to any node. Consider a sequence of irreducible matrices $\{Q_n\}_{n\in\mathbb{N}}$ of increasing size, together with their unique invariant measures $\pi^{(n)}$, such that $Q_n^*\pi^{(n)} = \pi^{(n)}$ and $\mathbf{1}^*\pi^{(n)} = 1$. We say that Q_n is *democratic* if $\|\pi^{(n)}\|_\infty \to 0$ as $n \to +\infty$. Clearly, a sequence of doubly stochastic matrices is democratic. A less trivial example is a SRW on a bidimensional grid $L_n \times L_m$. More in general, a family of SRWs on undirected graphs G_N is democratic if and only if $\max_v d_v / \sum_u d_u$ goes to zero as $N = |V|$ goes to infinity. This sufficient condition is not satisfied on star graphs, and indeed, the lazy SRW on S_n is

$$P_{\mathrm{SRW}}(\tau) = \begin{bmatrix} 1-\tau & \tau/n\mathbf{1}_n^* \\ \tau\mathbf{1}_n & (1-\tau)I_n \end{bmatrix},$$

which is not democratic. However, a democratic sequence on star graphs can be constructed as

$$P_{\mathrm{dem}} = \begin{bmatrix} 0 & \frac{1}{n}\mathbf{1}_n^* \\ \frac{1}{n}\mathbf{1}_n & (1-\frac{1}{n})I_n \end{bmatrix}.$$

Other examples of democratic matrices are given in the Exercises.

2.7 Convergence Rate: Examples

We have already seen in Proposition 2.2 that the spectral radius of P determines the rate of convergence to consensus. In this section, the spectral radius is studied in some examples and its connection with convergence time is made more explicit.

We first present an example of a family of SRWs on a simple graph.

Example 2.5 (*SRW on cycles*) Let $G = C_n$ be the symmetric cycle graph with n vertices. Since C_n is 2-regular, we have that the matrix of the SRW is

$$P = \text{circ}([0, 1/2, 0, \dots, 0, 1/2]).$$

The eigenvalues of P can be computed as we did for the Laplacian eigenvalues of C_n in Example 1.5. Namely, the eigenvalues of P are

$$\mu_k(P) = \cos\left(\frac{2\pi}{n}k\right).$$

Note that -1 is an eigenvalue if and only if n is even: This corresponds to the graph being bipartite. If instead n is odd, the second eigenvalue of P is

$$\rho_2 = \max\left\{\left|\cos\left(\frac{2\pi}{n}\right)\right|, \left|\cos\left(\frac{2\pi}{n}\frac{n-1}{2}\right)\right|\right\} = \cos\frac{\pi}{n}.$$

In order to control the convergence properties of a matrix adapted to C_n, we may define the family of matrices $P_\tau = (1-\tau)I + \tau P$, with the parameter $\tau \in (0, 1]$. A matrix in this family corresponds to a lazy SRW. In this case, $\mu_k(\tau; P) = 1 - \tau + \tau\cos\left(\frac{2\pi}{n}k\right)$ and

$$\rho_2(\tau) = \begin{cases} \max\{|1-\tau+\tau\cos\frac{2\pi}{n}|, |1-2\tau|\} & \text{if } n \text{ is even} \\ \max\{|1-\tau+\tau\cos\frac{2\pi}{n}|, |1-\tau(1+\cos\frac{\pi}{n})|\} & \text{if } n \text{ is odd.} \end{cases}$$

It is clear that a choice of $\tau \in (0, 1)$ allows to have ρ_2 smaller than 1 for every n. Furthermore, if n is even,

$$\rho_2(\tau) = \begin{cases} 1-\tau(1-\cos\frac{2\pi}{n}) & \text{if } \tau \le \frac{2}{3-\cos\frac{2\pi}{n}} \\ 2\tau - 1 & \text{otherwise} \end{cases}$$

and the minimum is achieved for $\tau = \frac{2}{3-\cos\frac{2\pi}{n}}$. If instead n is odd,

$$\rho_2(\tau) = \begin{cases} 1-\tau(1-\cos\frac{2\pi}{n}) & \text{if } \tau \le \frac{2}{2+\cos\frac{\pi}{n}-\cos\frac{2\pi}{n}} \\ \tau(1+\cos\frac{\pi}{n}) - 1 & \text{otherwise.} \end{cases}$$

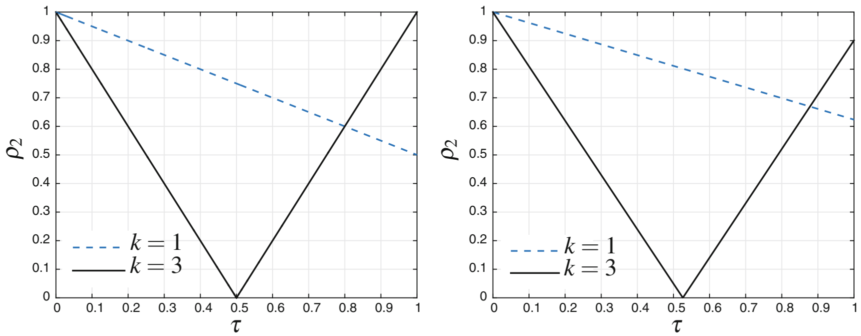


Fig. 2.3 Optimization in Example 2.5. Plots show the function $|\mu_k(\tau; P)|$ for $k = 1$ and $k = \lfloor \frac{n}{2} \rfloor$ assuming $n = 6$ (*left plot*) and $n = 7$ (*right plot*), respectively

and the minimum is achieved for $\tau = \frac{2}{2+\cos\frac{\pi}{n}-\cos\frac{2\pi}{n}}$. This optimization is illustrated in Fig. 2.3

The simple random walk can be also easily studied on the complete graph.

Example 2.6 (*SWR on complete graphs*) The SWR matrix on the complete graph is $P = \frac{1}{N-1}(\mathbf{1}\mathbf{1}^* - I)$, resulting in $\rho_2 = \frac{1}{N-1}$. If we instead consider the lazy version $P(\tau) = (1-\tau)I + \tau P$, we obtain $\rho_2(\tau) = 1 - \frac{N}{N-1}\tau$, which vanishes when $\tau = 1 - \frac{1}{N}$. Indeed, $P(1-\frac{1}{N}) = \frac{1}{N}\mathbf{1}\mathbf{1}^*$ gives consensus in one step.

As we have observed in Example 2.5, -1 is an eigenvalue of the SRW matrix when the graph is bipartite. To rule out such undesired case, in the next example we concentrate on a specific class of lazy SRW.

Example 2.7 (*Spectral radius of k-dimensional grids*) Let G be a d-regular graph, and consider the matrix $P = \frac{1}{d+1}(I + A)$. When the spectrum of A is known, ρ_2 can be readily computed. For instance, for k-dimensional symmetric grids C^k, it holds

$$\rho_2 = \frac{2k-1}{2k+1} + \frac{2}{2k+1}\cos\frac{2\pi}{n}.$$

If the number of nodes $N = n^k$ goes to infinity while keeping the dimension k fixed, then $\rho_2 \to 1$, and by the Taylor expansion of the cosine, we observe that[1]

$$\rho_2 = 1 - \frac{4\pi^2}{2k+1}\frac{1}{n^2} + o\left(\frac{1}{n^3}\right) = 1 - \frac{4\pi^2}{2k+1}\frac{1}{N^{2/k}} + o\left(\frac{1}{N^{3/k}}\right) \qquad \text{as } n \to \infty.$$

[1]Here and throughout the book, we will make use of the standard asymptotic notation. If f_n and g_n are two positive sequences, we say that $f_n = O(g_n)$ if f_n/g_n is upper bounded in n; that $f_n = \Theta(g_n)$ if f_n/g_n is both lower and upper bounded in n; that $f_n = o(g_n)$ if $f_n/g_n \to 0$ as $n \to \infty$; and that $f_n \sim g_n$ if $f_n/g_n \to 1$ as $n \to \infty$.

Hence, $1-\rho_2$ goes to zero at a polynomial rate. More details and more graph examples on this SRW are given in Exercise 2.12.

To better highlight the role of N, we define the *convergence time* as

$$T_\varepsilon = \inf\{t > 0 \,:\, ||P^t - \mathbf{1}\pi^*|| < \varepsilon\}.$$

On symmetric matrices, $||P^t - \mathbf{1}\pi^*|| = \rho_2^t$, so that $T_\varepsilon = \frac{\log \varepsilon^{-1}}{\log \rho_2^{-1}}$, which is in turn upper bounded by $\frac{\log \varepsilon^{-1}}{1-\rho_2}$. Moreover, $\frac{1}{1-\rho_2} \sim \frac{1}{\log \rho_2^{-1}}$ as $\rho_2 \to 1$. Hence, the inverse of the spectral gap of P can be immediately interpreted as an upper bound on the convergence time. For instance, in the example above, $T_\varepsilon \sim \frac{2k+1}{4\pi^2} N^{2/k} \log \varepsilon^{-1}$.

Remark 2.4 (*Trade-off between speed and democracy*) Consider again irreducible matrices adapted to star graphs as at the end of Sect. 2.6. You can easily verify—exercise—that P_{dem} has second eigenvalue $\rho_2(P_{\text{dem}}) = 1 - \frac{1}{n}$ (thus growing to 1 as $n \to= \infty$), whereas $\rho_2(P_{\text{SRW}}(\tau)) = 1 - \tau$. On the other hand, P_{dem} is democratic while $P_{\text{SWR}}(\tau)$ is not. This observation highlights that by choosing either matrix we are trading off speed for democracy. This trade-off exhibited by star graphs is further discussed in Exercise 2.18; see also Exercises 2.23 and 2.24 for other graphs having this feature.

2.8 Reversible Matrices

An important family of stochastic matrices, encompassing SRW and in general all symmetric stochastic matrices, is the family of *reversible* matrices. A reversible matrix can be defined starting from any nonnegative symmetric matrix $M \in \mathbb{R}^{V \times V}$ putting

$$P_{vw} = \frac{M_{vw}}{(M\mathbf{1})_v}. \tag{2.9}$$

It is immediate to check that P is stochastic and that $\pi_v = (M\mathbf{1})_v[\sum_u (M\mathbf{1})_u]^{-1}$ is an invariant probability measure of P. Notice that SRW on symmetric graphs is a special case of this construction, when M is the adjacency matrix of the graph. We have the following alternative characterization:

Proposition 2.7 (Reversibility) *Let P be a stochastic matrix. The following conditions are equivalent:*

(i) P is reversible;
(ii) There exists a nonzero and nonnegative $x \in \mathbb{R}^V$ such that

$$x_v P_{vw} = x_w P_{wv} \quad \textit{for all}\, v, w \in V. \tag{2.10}$$

Proof On the one hand, if P satisfies (2.9), it follows that

$$(M\mathbf{1})_v P_{vw} = M_{vw} = M_{wv} = (M\mathbf{1})_w P_{wv}$$

On the other hand, if P satisfies (2.10), then putting $M_{vw} := x_v P_{vw}$ we have that M is nonnegative symmetric, and it holds $(M\mathbf{1})_v = x_v$. Hence, (2.9) holds with such M. □

Condition (2.10) is often referred to as the *detailed balance* condition: We note that it implies that x is actually an invariant measure (possibly not normalized to 1) of P, because

$$\sum_v x_v P_{vw} = \sum_v x_w P_{wv} = x_v.$$

Condition (2.10) is actually stronger than the requirement that x is an invariant measure as it says that each pair of nodes v, w for which $P_{vw} > 0$ must balance the exchange flow between each other. The reason for the name "reversible" becomes clear when we interpret it in the probabilistic framework considering P as the transition matrix of a Markov chain X_t having initial probability vector π satisfying the condition (2.10). Then, the left and right terms of (2.10) can be interpreted, respectively, as $\mathbb{P}(X_t = v, X_{t+1} = w)$ and $\mathbb{P}(X_t = w, X_{t+1} = v)$.

It is possible to generalize to reversible matrices most of the results obtained for symmetric matrices: The key fact is that reversible matrices are *diagonalizable* as we show below. Let $P \in \mathbb{R}^{V\times V}$ be a reversible, irreducible, aperiodic stochastic matrix, and let $\pi \in \mathbb{R}^V$ be its invariant probability measure. Consider D_π the diagonal matrix such that $(D_\pi)_{vv} = \pi_v$, and define $A = D_\pi^{1/2} P D_\pi^{-1/2}$. Reversibility implies (check this) that A is symmetric. Let ϕ_j's, for $j \in \{1, \dots, n\}$, be an orthonormal basis of eigenvectors for A with correspondent real eigenvalues μ_j. It is immediate to check that $\pi^{1/2}$ is indeed an eigenvector with eigenvalue 1. Therefore, we will assume that $\phi_1 = \pi^{1/2}$ and $\mu_1 = 1$. A straightforward verification shows that the $\psi_j = D_\pi^{-1/2}\phi_j$ are eigenvectors of P with eigenvalue μ_j. The ψ_j's together with $\mathbf{1}$ do form a basis of eigenvectors of P which is thus diagonalizable. Using the usual orthonormal splitting expression for A, we can write

$$A^t = \pi^{1/2}(\pi^{1/2})^* + \sum_{j\geq 2} \mu_j^t \phi_j \phi_j^*,$$

from which we can derive the following useful representation for P^t

$$P^t = \mathbf{1}\pi^* + D_\pi^{-1/2} \sum_{j\geq 2} \mu_j^t \phi_j \phi_j^* D_\pi^{1/2}.$$

From this expression, we can estimate the speed of convergence as in the symmetric case (see Problem 2.19 for details). Moreover, we can extend the theory developed for the Laplacian $L(P) = I - P$, when P is a symmetric matrix, to the case when

P is reversible. The idea for the extension simply involves replacing the Euclidean scalar product with the product induced by π, which is $\langle x, y\rangle_\pi :=< x, D_\pi y >= \sum_v \pi_v x_v y_v$. In particular, the following results, extending Propositions 1.9 and 1.9, hold true (their proof is left to the reader).

Proposition 2.8 (Dirichlet form for reversible matrices) *Assume that P is a reversible stochastic matrix with invariant probability measure π. For every $x \in \mathbb{R}^V$, it holds*

$$\langle x, L(P)x\rangle_\pi = \frac{1}{2}\sum_{v,w} P_{vw}\pi_v(x_v - x_w)^2. \tag{2.11}$$

Proposition 2.9 (Variational characterization for reversible matrices) *Assume that P is a reversible stochastic matrix with invariant probability measure π and second largest eigenvalue μ_2. Let λ_2 be the second smallest eigenvalue of $L(P)$. It holds*

$$\lambda_2 = (1 - \mu_2) = \min_{x\neq 0, \langle x,\mathbf{1}\rangle_\pi = 0} \frac{\langle x, L(P)x\rangle_\pi}{\langle x, x\rangle_\pi}. \tag{2.12}$$

A useful technique to upper bound the spectral gap of a reversible stochastic matrix P is through the so-called *bottleneck ratio*, a sort of index measuring how well the “flow” represented by the matrix is spreading along the underlying graph. Suppose π is the usual invariant probability measure of P, and for every $S \subset V$, define $\pi(S) = \sum_{v\in S}\pi_v$ and

$$Q(S, S^c) = \sum_{v\in S, w\notin S} \pi_v P_{vw}$$

(check as an exercise that $Q(S, S^c) = Q(S^c, S)$ for all $S \subset V$). Then, we define

$$\Phi(S) := \frac{Q(S, S^c)}{\pi(S)}$$

and the bottleneck ratio of P as

$$\Phi_* := \min_{S:\pi(S)\leq \frac{1}{2}} \Phi(S).$$

In the flow interpretation $Q(S, S^c)$ represents the total flow exiting S (assuming we are at the stationary regime), while $\Phi(S)$ the fraction of flow exiting S with respect to the total flow exiting from the nodes in S. We have the following result:

Proposition 2.10 (Cheeger bound) *Let μ_2 be the second largest eigenvalue of a reversible matrix P, and let Φ_* be the bottleneck ratio of P. Then,*

$$1 - \mu_2 \leq 2\Phi_*.$$

Proof Given $S \subseteq V$, consider the vector $\phi \in \mathbb{R}^V$ defined by $\phi_v = \pi(S^c)$ if $v \in S$, and $\phi_v = -\pi(S)$ if $v \in S^c$. Then, from Proposition 2.8 and the detailed balance condition (2.10), it follows that

$$\begin{aligned}
\langle \phi, L(P)\phi \rangle_\pi &= \frac{1}{2} \sum_{v,w} \pi_v P_{vw} (\phi_v - \phi_w)^2 \\
&= \sum_{v \in S, w \notin S} \pi_v P_{vw} (\phi_v - \phi_w)^2 \\
&= \sum_{v \in S, w \notin S} \pi_v P_{vw} (\pi(S) + \pi(S^c))^2 \\
&= \sum_{v \in S, w \notin S} \pi_v P_{vw} = Q(S, S^c)\,.
\end{aligned}$$

On the other hand,

$$\langle \phi, \phi \rangle_\pi = \sum_v \pi_v \phi_v^2 = \sum_{v \in S} \pi_v \pi(S^c)^2 + \sum_{w \notin S} \pi_w \pi(S)^2 = \pi(S)\pi(S^c).$$

From the variational characterization of Proposition 2.9, and assuming $\pi(S) \le 1/2$, we thus conclude

$$\lambda_2 \le \frac{\langle \phi, L(P)\phi \rangle_\pi}{\langle \phi, \phi \rangle_\pi} = \frac{Q(S, S^c)}{\pi(S)\pi(S^c)} \le 2\Phi(S).$$

Since this inequality holds for all S such that $\pi(S) \le \frac{1}{2}$, the upper bound is proved. □

Notice that, since $\rho_2 \ge \mu_2$, we can also bound the spectral gap by

$$1 - \rho_2 \le 2\Phi_*.$$

In the case when P is the SRW on a symmetric graph $G = (V, E)$, the bottleneck ratio takes a peculiar form which is convenient to work out:

$$\Phi(S) = \frac{\displaystyle\sum_{v \in S, w \in S^c} \frac{d_v}{|E|} (A_G)_{vw} \frac{1}{d_v}}{\displaystyle\sum_{v \in S} \frac{d_v}{|E|}} = \frac{\sum_{v \in S, w \in S^c} (A_G)_{vw}}{\sum_{v \in S} d_v} \tag{2.13}$$

This equation says that $\Phi(S)$ equals the *fraction of edges which start inside S and end outside S*.

Example 2.8 (*Graphs with a bottleneck*) Given two graphs $G_1 = (V_1, E_1)$, $G_2 = (V_2, E_2)$ and a symmetric set of edges $E_3 \subseteq (V_1 \times V_2) \cup (V_2 \times V_1)$, we can consider the interconnected graph $G = (V_1 \cup V_2, E_1 \cup E_2 \cup E_3)$. Following (2.13), for the SRW on the graph G, we have that the bottleneck can be estimated as

$$\Phi_* \leq \Phi(V_1) = \frac{|E_3|}{2|E_1| + |E_3|}$$

For instance, consider the case when $|V_1| = |V_2| = n$, G_1 and G_2 are both complete, and $|E_3| = 2$ (namely, there is just one edge and its inverse) connecting the two complete graphs (this is called *barbell graph*). Then, $\Phi_* \leq (n(n-1)+1)^{-1}$.

Similar reasonings can be applied to other families of matrices; see for instance Exercise 2.20 on Metropolis random walks. For completeness, we report that also a lower bound on the spectral gap involving the bottleneck ratio can be obtained [30, Theorem 13.14].

Proposition 2.11 (Reverse Cheeger bound) *Let μ_2 be the second largest eigenvalue of a reversible matrix P, and let Φ_* be the bottleneck ratio of P. Then,*

$$1 - \mu_2 \geq \frac{{\Phi_*}^2}{2}.$$

Exercises

Exercises are divided into five groups, respectively, devoted to some basic facts, to the rate of convergence, to the consensus value, to reversible matrices, and to miscellaneous arguments.

First Examples and Concepts

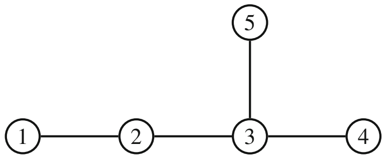
Exercise 2.1 (*SRW example*) Consider the graph G in Fig. 2.4, and let P be the transition matrix relative to the simple random walk on G.

(i) Write P.
(ii) Compute P^9_{13}.
(iii) What is the multiplicity of the eigenvalue 1 of P? Why?
(iv) Is -1 an eigenvalue of P? Is P aperiodic? Why?

Exercise 2.2 (*Consensus example*) Consider the simple random walk (2.8) on the graph $G = (V, E)$ defined in Exercise 1.9, and let $x(t)$ be the evolution of the consensus algorithm associated with the corresponding stochastic matrix.

(i) Prove that $x(t)$ converges to consensus at the value α.
(ii) Find the value of α as a function of the initial condition $x(0)$.
(iii) Find $\bar{t}$ such that $\|x(\bar{t}) - |V|^{-1}\mathbf{1}\| \leq 10^{-2}\|x(0)\|$.

Fig. 2.4 The graph G of Exercise 2.1



Exercise 2.3 (*Periodic matrices*) Let P be a stochastic matrix such that $\mathcal{G}_P$ has a globally reachable node. Prove the following facts.

(i) $Q_\varepsilon = \varepsilon I + (1-\varepsilon)P$ is stochastic, and $\mathcal{G}_{Q_\varepsilon}$ has a globally reachable aperiodic node.
(ii) The scalar 1 is a simple eigenvalue of P.
(iii) The spectrum of P is contained in the closed unit disk of the complex plane.
(iv) P has a unique invariant probability measure.

Rate of Convergence

Exercise 2.4 (*SRW on complete*) Let G be the complete graph.

(i) Write down explicitly the corresponding symmetric random walk P, as in (2.8).
(ii) Compute all eigenvalues of P and, in particular, the second eigenvalue ρ_2.
(iii) Consider the lazy SRW $P_\tau = (1-\tau)I + \tau P$, and compute the corresponding ρ_2.

Exercise 2.5 (*SRW on complete bipartite*) Consider the complete bipartite graph $K_{m,n}$ as defined in Example 1.2

(i) Write down explicitly the corresponding symmetric random walk P, as in (2.8).
(ii) Compute all eigenvalues of P and, in particular, check that -1 is always an eigenvalue.
(iii) Consider the lazy SRW $P_\tau = (1-\tau)I + \tau P$, and compute the corresponding ρ_2.

Exercise 2.6 (*SRW on grids*) Let $G = C_n \times C_m$ be the symmetric two-dimensional toroidal graph with $n \times m$ vertices.

(i) Observe that the corresponding symmetric random walk P is a Cayley matrix.
(ii) Compute all eigenvalues of P and find when -1 is an eigenvalue.
(iii) When n and m are both odd, compute the second eigenvalue ρ_2 of P.
(iv) Consider the lazy SRW $P_\tau = (1-\tau)I + \tau P$, and compute the corresponding ρ_2 for every value of n and m.
(v) As τ varies in $[0, 1]$, compute the maximal value of the spectral gap $1-\rho_2$ (you may assume that n and m are sufficiently large).

Exercise 2.7 (*Symmetric cycle*) Let $n \in \mathbb{N}$, and consider the symmetric cycle graph C_n with adjacency matrix A_n. Consider the matrix $P_n = \frac{1}{3}(I + A_n)$ corresponding to a lazy random walk on C_n.

(i) Verify that P_n is a lazy simple random walk on C_n.

(ii) Let $\rho_2^{(n)}$ be the second largest eigenvalue of P_n. Using the formula for the eigenvalues of circulant matrices in Proposition 1.13, find an expression for $\rho_2^{(n)}$.
(iii) Find a function $f(n)$ such that $f(n) \sim 1 - \rho_2^{(n)}$.

Exercise 2.8 (*Augmented cycle I*) Let $n \in \mathbb{N}$, and consider the following augmentation G_n of the symmetric cycle graph C_n, defined as follows: A node $i \in \{0, \dots, n-1\}$ is connected with nodes $i-2, i-1, i+1, i+2 \pmod n$. Consider the matrix P_n corresponding to the simple random walk on G_n. Let $\rho_2^{(n)}$ be the second largest eigenvalue of P_n.

(i) Using the formula for the eigenvalues of circulant matrices in Proposition 1.13, find an expression for $\rho_2^{(n)}$.
(ii) Find a function $f(n)$ such that $f(n) \sim 1 - \rho_2^{(n)}$.

Exercise 2.9 (*Augmented cycle II*) Let n be an even number, and consider the following augmentation G_n of the symmetric cycle graph C_n, defined as follows: A node $i \in \{0, \dots, n-1\}$ is connected with nodes $i-1, i+1, i+n/2 \pmod n$. Consider the matrix P_n corresponding to the simple random walk on G_n. Let $\rho_2^{(n)}$ be the second largest eigenvalue of P_n.

(i) Using the formula for the eigenvalues of circulant matrices, find an expression for $\rho_2^{(n)}$.
(ii) Find a function $f(n)$ such that $f(n) \sim 1 - \rho_2^{(n)}$.

Exercise 2.10 (*Line graph*) Let $n \in \mathbb{N}$, and consider the following matrix P_n corresponding to a random walk on the symmetric line graph L_n:

$$P_n = \frac{1}{2}\begin{bmatrix} 1 & 1 & 0 & \dots & & 0 & 0 \\ 1 & 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & 1 & \dots & 0 & 0 \\ & & & \ddots & & & \\ 0 & 0 & \dots & & 0 & 1 & 0 \\ 0 & 0 & \dots & & 1 & 0 & 1 \\ 0 & 0 & \dots & & 0 & 1 & 1 \end{bmatrix}.$$

Let $\rho_2^{(n)}$ be the second largest eigenvalue of P_n.

(i) Using the formulas for the eigenvalues of tridiagonal matrices in Exercise 1.27, verify that $\rho_2^{(n)} = \cos\frac{\pi}{n}$.
(ii) Consider the simple random walk on a symmetric cycle C_n (see Exercise 2.7), and denote by $\bar{\rho}_2^{(n)}$ the second largest eigenvalue of the associated matrix. Show that $\rho_2^{(n)} \geq \bar{\rho}_2^{(n)}$, and compute $\lim_n \dfrac{1-\rho_2^{(n)}}{1-\bar{\rho}_2^{(n)}}$.
(iii) Interpret the above results in terms of speed of convergence of the corresponding consensus algorithms on L_n and C_n.

Exercise 2.11 (*Line graph II*) Let $n \in \mathbb{N}$, and consider the symmetric line graph L_n with adjacency matrix B_n. Consider the matrix

$$Q_n = \frac{1}{3}(I + B_n) + \frac{1}{3}\begin{bmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ & & \ddots & & \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 1 \end{bmatrix}$$

corresponding to a lazy random walk on L_n. Let $\rho_2^{(n)}$ be the second largest eigenvalue of Q_n.

(i) Using Exercise 1.27, find a closed form expression for $\rho_2^{(n)}$.
(ii) Find a function $f(n)$ such that $f(n) \sim 1 - \rho_2^{(n)}$.
(iii) Compare these results with the analogous results for the simple random walk on the cycle graph C_n in Exercise 2.7.

Exercise 2.12 (*Rate comparison*) Consider cycle graphs, k-dimensional torus graphs, k-dimensional hypercubes, and De Bruijn graphs on k symbols.

(i) Observe that the graphs at hand are regular. For each of these graphs, consider its Laplacian matrix L and the stochastic matrix $P = I - \frac{1}{d+1}L = \frac{1}{d+1}(I + A)$. Let ρ_2 be the magnitude of the second largest eigenvalue of P.
(ii) Observe that the graphs at hand have real Laplacian eigenvalues. Denote them as $0 = \lambda_1 < \lambda_2 \leq \dots \leq \lambda_n$, and verify that $\rho_2 = 1 - \frac{\lambda_2}{d+1}$, where d is the degree of the graph. Observe that under the current assumptions, the rate of convergence of P is completely determined by the topology of the graph.
(iii) Compute the values of ρ_2 as functions of the graph parameters and of the number of nodes N. Derive the values in Table 2.1.
(iv) By using Taylor expansions, compute the principal part of the rate ρ_2 as $N \to +\infty$, in the cases of Table 2.1.
(v) Rank the graphs in Table 2.1 from fastest to slowest. Observe that if we consider a sequence B_h^k with k fixed and $h \in \mathbb{N}$, then ρ_2 does not depend on N.

Exercise 2.13 (*Rate on directed grids*) Let $G = \mathbf{C}_n^d$ be a directed d-dimensional torus, $P = \frac{1}{d+1}(I + A_G)$, and $N = n^d$.

(i) Verify that

$$\rho_2 = \sqrt{\frac{d^2 + 1 + 2d\cos\left(\frac{2\pi}{n}\right)}{(d+1)^2}} = 1 - \frac{2d\pi^2}{(d+1)^2}\frac{1}{n^2} + o\left(\frac{1}{n^3}\right) \text{ as } n \to \infty$$

Exercise 2.14 (*Rate comparison in SRW*) Consider the same graphs as in Exercise 2.12 and for each of them the stochastic matrix $P = \frac{1}{d}A$, corresponding to a symmetric random walk.

Table 2.1 Rates of convergence for consensus algorithms on several families of graphs; see Example 2.7 and Exercise 2.12

Graph	λ_2	d	N	ρ_2	$\rho_2(N)$
C_n	$2(1-\cos\frac{2\pi}{n})$	2	n	$\frac{1}{3}\left(1+2\cos\frac{2\pi}{n}\right)$	$\frac{1}{3}\left(1+2\cos\frac{2\pi}{N}\right)$
$C_n \times C_n$	$2(1-\cos\frac{2\pi}{n})$	4	n^2	$\frac{1}{5}(3+2\cos\frac{2\pi}{n})$	$\frac{1}{5}(3+2\cos\frac{2\pi}{\sqrt{N}})$
C_n^k	$2(1-\cos\frac{2\pi}{n})$	$2k$	n^k	$\frac{2k-1}{2k+1}+\frac{2}{2k+1}\cos\frac{2\pi}{n}$	$\frac{2k-1}{2k+1}+\frac{2}{2k+1}\cos\frac{2\pi}{N^{1/k}}$
H_k	2	k	2^k	$\frac{d-1}{d+1}$	$\frac{\log_2 N-1}{\log_2 N+1}$
B_h^k	k	k	k^h	$\frac{1}{k+1}$	$\frac{1}{N^{1/h}+1}$

(i) Compute the second eigenvalues ρ_2 as functions of the graph parameters and of the number of nodes N.
(ii) Compare your results with those in Exercise 2.12.

Exercise 2.15 (*Majority computation*) Let $G = (V, E)$ be a symmetric connected graph and $\bar{x} \in \{-1, +1\}^V$. Let $N_1 = |\{v \in V : \bar{x}_v = 1\}|$ and $N_{-1} = N - N_1$ where $N = |V|$. The agents want to estimate which state has the majority. Consider the following algorithm. Let P be an aperiodic irreducible symmetric matrix adapted to G, and define

$$\begin{cases} x(t) = P^t \bar{x} \\ \lambda(t) = \text{sign}(x(t)) \in \{-1, +1\}^V. \end{cases}$$

Clearly, if $N_1 \neq N_{-1}$, then $\lim_{t\to\infty} \lambda(t) = \bar{\lambda}\mathbf{1}$ and $\bar{\lambda} = 1$ when $N_1 > N_{-1}$. Agent v can then use $\lambda_v(t)$ as an estimation of the majority value. Let $T_{\min} = \min\{t \in \mathbb{N} : \lambda_v(t) = \bar{\lambda} \ \ \forall v \in V\}$.

(i) Estimate $T_{\min}$ in terms of the second eigenvalue ρ_2 of P and of the vector $\bar{x}$.
(ii) Estimate $T_{\min}$ when P is the SRW in the toroidal d-grid of size N.

Evaluation of the Convergence Value and Democracy

Exercise 2.16 (*Democracy and wisdom of crowds*) Consider an irreducible aperiodic matrix Q used to solve a consensus problem with $x(0) = \theta\mathbf{1} + \eta$, where θ is a scalar and η is a vector of "disturbances." We know that $x_v(t) \to x_v(\infty) = \theta + \pi^*\eta$ for all v. Assume that η_vs are random variables, independent and identically distributed with zero mean and variance σ^2 (this setup will be considered again in Sect. 4.5).

Let $\{Q_N\}_{N\in\mathbb{N}}$ be a sequence of such matrices, each with size N. According to [21], the sequence Q_N is said to be *wise* if the variance of $x_v(\infty)$ goes to 0 as N goes to infinity. Prove the following statements.

(i) Q_N is democratic if and only if it is wise.
(ii) There exists $c > 0$ such that $\frac{\pi_u}{\pi_v} \leq c$ for all indices u, v and all size N if and only if there exist two positive constants c_1 and c_2 such that $\frac{c_1}{N} \leq \pi_w \leq \frac{c_2}{N}$ for all w and all N.
(iii) The conditions at point (ii) imply that Q_N is wise.

Exercise 2.17 (*Line graph: democracy*) Let $n \in \mathbb{N}$, and consider the symmetric line graph L_n. Consider the matrix S_n associated with the simple random walk on L_n, and define

$$Q_n = \frac{1}{3}I + \frac{2}{3}S_n.$$

(i) Observe that Q_n is stochastic but not doubly stochastic, and compute $\pi^{(n)}$, the invariant probability measure of Q_n.
(ii) Compute for each component $v \in \{1, \dots, n\}$,

$$\lim_{n\to+\infty} \pi_v^{(n)} \quad \text{and} \quad \lim_{n\to+\infty} \frac{\pi_v^{(n)}}{1/n}.$$

Comment on your results, recalling that the invariant probability measure of a doubly stochastic matrix is $\frac{1}{n}\mathbf{1}$. Is Q_n democratic?
(iii) Compute $\rho_2(n)$, the second largest eigenvalue of Q_n.
(iv) Compare these figures with the corresponding results for the simple random walk on the cycle graph C_n (see Exercise 2.7).

Exercise 2.18 (*Speed and democracy on star graphs*) Consider the graph S_n, the symmetric star graph with n edges and $n+1$ nodes, whose center node is denoted as 0 and the n leaves as the elements of the set $\{1, \dots, n\}$. Then, consider the following family of adapted stochastic matrices

$$P_n = \begin{bmatrix} 1-n\alpha & \alpha & \alpha & \alpha & \dots & \alpha \\ \beta & 1-\beta & 0 & 0 & \dots & 0 \\ \beta & 0 & 1-\beta & 0 & \dots & 0 \\ & \vdots & & \ddots & & \\ & \vdots & & & & \\ \beta & 0 & \dots & 0 & 0 & 1-\beta \end{bmatrix},$$

where the parameters α, β satisfy $0 \leq \alpha \leq \frac{1}{n}$ and $0 \leq \beta \leq 1$.

(i) Verify that the eigenvalues of P_n are 1, $1-\beta$, and $1-n\alpha-\beta$.
(ii) Find the values of α, β which give the fastest consensus algorithm.

(iii) Design a consensus algorithm, adapted to S_n, which converges in a finite number of steps. How many steps does it need to converge? Compare the required number of steps of the algorithm with the diameter of S_n. Which is the consensus value?

(iv) Compute $\rho_2(n)$ as a function of α, β on its domain.

(v) Compute the invariant probability measure of P_n. Verify that the consensus algorithm defined by P_n converges to the average of the initial states if and only if $\alpha = \beta$.

(vi) Prove that if $\alpha = \beta$, the rate of convergence $\rho_2(n)$ grows to 1 as $n \to +\infty$. Estimate the convergence time on large networks, as a function of n.

(vii) Prove that

(a) if P_n is democratic, then necessarily $\rho_2(n) \to 1$ as n diverges;
(b) conversely, if $\rho_2(n) \leq 1 - c$ for all n and some positive c, then necessarily P_n is not democratic.

Conclude that in optimizing P_n one necessarily needs to trade off the speed of convergence for the distance between the limit value and the average of initial states.

Reversible Matrices

Exercise 2.19 (*Convergence rate*) Suppose $P \in \mathbb{R}^{V \times V}$ is stochastic reversible with invariant probability measure π. Then, the result in Proposition 2.2 can be strengthened to claim that

$$||P^t x(0) - \mathbf{1}\pi^* x(0)||_2 \leq \frac{\max_v \pi_v^{1/2}}{\min_v \pi_v^{1/2}} ||x(0)||_2 \rho_2^t \quad \forall t \in \mathbb{N}.$$

Exercise 2.20 (*Speed in unbalanced sequence*) The goal of this exercise is to show that the rate of convergence of a Metropolis random walk goes to one on a sequence of graphs, if there is a node whose degree vanishes compared to the degree of its neighbors. Let there be a sequence of symmetric connected graphs of increasing size $G_n = (V_n, E_n)$ and a sequence of nodes $v_n \in V_n$ such that

$$\lim_{n \to +\infty} \frac{d_{v_n}}{\min\{d_w \ : \ w \in \mathcal{N}_{v_n}\}} = 0.$$

Let P_n be the Metropolis random walk associated with G_n and $\rho_2^{(n)}$ its second largest eigenvalue. Using Cheeger bound, show that the gap $1 - \rho_2^{(n)}$ goes to zero when n diverges.

Exercise 2.21 (*Matrices adapted to $K_{m,n}$*) Let α, β be real numbers, and let the $(m+n)$-dimensional square matrix $M^{(\alpha,\beta)}$ be

$$M^{(\alpha,\beta)} = \begin{bmatrix} n\alpha I_m & -\alpha \mathbf{1}_m \mathbf{1}_n^* \\ -\beta \mathbf{1}_n \mathbf{1}_m^* & m\beta I_n \end{bmatrix}.$$

Verify that $M^{(\alpha,\beta)}$ has eigenvalues

- 0 corresponding to eigenvector $\mathbf{1}_{m+n}$;
- $n\alpha + m\beta$ corresponding to eigenvector $\begin{bmatrix} n\alpha\mathbf{1}_m \\ -m\beta\mathbf{1}_n \end{bmatrix}$;
- $n\alpha$ corresponding to the $(m-1)$-dimensional eigenspace span $\{\begin{bmatrix} x \\ \mathbf{0}_n \end{bmatrix} : x^*\mathbf{1} = 0\}$;
- $m\beta$ corresponding to the $(n-1)$-dimensional eigenspace span $\{\begin{bmatrix} \mathbf{0}_m \\ y \end{bmatrix} : y^*\mathbf{1} = 0\}$.

Exercise 2.22 (*Matrices adapted to wheels*) Let A_n be a normal[2] matrix of order n such that $A\mathbf{1}_n = d\mathbf{1}_n$, and denote by $x^{(k)}$ and $\lambda^{(k)}$ for $k \in \{1, \dots, n-1\}$ the remaining (orthonormal) eigenvectors of A_n with the corresponding eigenvalues. Consider matrix

$$M = \begin{bmatrix} 0 & \frac{1}{n}\mathbf{1}_n^* \\ \frac{1}{d+1}\mathbf{1}_n & \frac{1}{d+1}A_n \end{bmatrix}.$$

Verify that matrix M is stochastic and has eigenvalues 1 (simple), $-\frac{1}{d+1}$ (with eigenvector $\begin{bmatrix} -(d+1) \\ \mathbf{1}_n \end{bmatrix}$), and $\frac{\lambda^{(k)}}{d+1}$ for $k \in \{1, \dots, n-1\}$ (with eigenvector $\begin{bmatrix} 0 \\ x^{(k)} \end{bmatrix}$).

Exercise 2.23 (*Speed and democracy on $K_{m,n}$*) Let A, B be two sets such that $|A| = m \le n = |B|$ and consider the complete bipartite graph $K_{m,n} = (A \cup B, E)$ as defined in Example 1.2. Define on this graph

- the lazy simple random walk P by $P_{vw} = \frac{1}{2}\frac{1}{d_v}$ for all $(v, w) \in E$; and
- the lazy Metropolis random walk $\bar{P}$ by $\bar{P}_{vw} = \frac{1}{2}\min\{\frac{1}{d_v}, \frac{1}{d_w}\}$ for all $(v, w) \in E$.

Let π and ρ_2 be the invariant probability measure and the second largest eigenvalue of P, and correspondingly, let $\bar{\pi}$ and $\bar{\rho}_2$ be the invariant probability measure and the second largest eigenvalue of $\bar{P}$.

(i) Using Exercise 2.21, prove that the LSRW P is such that $\rho_2 = \frac{1}{2}$ and the invariant measure π is such that if $a \in A$, then $\pi_a = \frac{1}{2m}$, and if $b \in B$, then $\pi_b = \frac{1}{2n}$.

(ii) Prove that the LMRW $\bar{P}$ is such that $\bar{\pi}_v = \frac{1}{m+n}$ for all $v \in (A \cup B)$, and $\bar{\rho}_2 = 1 - \frac{m}{2n}$.

(iii) Let $m \in \mathbb{N}$ be fixed and $\{K_{m,n}\}_{n \ge m}$ a sequence of complete bipartite graphs of increasing size. Consider the consensus algorithms associated with these graphs by the above two definitions of random walks and compare them. Observe that the choice of either definition of the adapted stochastic matrix implies a trade-off between democracy and good convergence speed.

[2] A matrix A is said to be *normal* when $A^*A = AA^*$. Normal matrices are precisely those for which a complete basis of eigenvectors exists. Symmetric matrices are normal.

Exercise 2.24 (*Speed and democracy on wheels*) Let $n \geq 3$ and consider the n-wheel graph W_n, which is defined as the union graph of a cycle C_n having node set $\{1, \ldots, n\}$ and a star S_n having node set $\{0\} \cup \{1, \ldots, n\}$. Define on this graph

- the lazy simple random walk P by setting $P_{vw} = \frac{1}{2}\frac{1}{d_v}$ for all $(v, w) \in E$; and
- the lazy Metropolis walk $\bar{P}$ by setting $\bar{P}_{vw} = \frac{1}{2}\min\{\frac{1}{d_v}, \frac{1}{d_w}\}$ for all $(v, w) \in E$.

Let π and ρ_2 be the invariant probability measure and the second largest eigenvalue of P and correspondingly let $\bar{\pi}$ and $\bar{\rho}_2$ be the invariant probability measure and the second largest eigenvalue of $\bar{P}$.

(i) Prove that the LSRW P is such that $\rho_2 \leq \frac{5}{6}$, and the invariant measure is such that $\pi_0 = \frac{1}{4}$ and $\pi_v = \frac{3}{4}\frac{1}{n}$ if $v \neq 0$.
(ii) Prove that the LMRW $\bar{P}$, which has uniform invariant measure, is such that $1 - \bar{\rho}_2 \leq \frac{7}{3}\frac{1}{n}$.
(iii) Now consider a sequence of wheel graphs of increasing size $\{W_n\}_{n\geq 3}$. Consider the consensus algorithms associated with these graphs by the above definitions of random walks. Remark that the choice of either definition of the adapted stochastic matrix implies a choice between democracy and good convergence speed.

Additional Topics

Exercise 2.25 (*Properties of Laplacians*) Let $G = (V, E, A)$ be a weighted graph of order n and L be the (weighted) Laplacian of G. Then, $\mathrm{rank}(L) = n - 1$ if and only if G contains a globally reachable vertex.

Exercise 2.26 (*Consensus in continuous time*) Consider a graph $G = (V, E)$ whose nodes are equipped with scalar dynamical systems $\dot{x}_v = u_v$, where $u \in \mathbb{R}^V$ is a control to be designed in order to achieve consensus. Let $A \in \mathbb{R}^{V\times V}$ be any nonnegative matrix such that $G = G_A$. Consider the feedback control law

$$u = -L(A)x.$$

(i) Verify that the control law $u = -L(A)x$ can be written componentwise as

$$u_v = \sum_w A_{vw}(x_w - x_v) \qquad \forall\, v \in V.$$

Consequently, it may be implemented by communicating with neighbors only and exchanging only relative information.

(ii) Show that, provided G has a globally reachable node, the closed loop system

$$\dot{x} = -L(A)x \tag{2.14}$$

yields consensus: For every initial condition $x(0)$, there exists a consensus value $\bar{x} \in \mathbb{R}$ such that $x_v(t) \to \bar{x}$ when $t \to +\infty$ for every $v \in V$. *Hint:* use Exercise 2.25.

(iii) Find the rate of convergence of (2.14).

Exercise 2.27 (*Node counting on a tree by message-passing*) Message-passing is a powerful approach to distributed computation, at least when the graph is a tree. Suppose then $G = (V, E)$ to be a tree and consider the following algorithm for the distributed computation of the number vertices N. Each unit $v \in V$ keeps in memory $d_v + 1$ scalar numbers, where d_v is the degree of v. We denote them as z_v^w with $w \in \mathcal{N}_v \cup \{v\}$. The algorithm is based on sending messages and updating z_v^w, according to the following rules.

- (Initialization): Set $z_v^w = 1$ for all $(v, w) \in E$.
- (Condition to send a message): Once unit v has received a message from all its neighbors except w, then v sends to w the following message: $z_{(v,w)} = z_v^w$.
- (Update upon receiving a message): When a node v receives $z_{(u,v)}$, the node does the following: $z_v^w = z_v^w + z_{(u,v)}$ for all $w \neq u$.
- (Termination): Once unit v has received message from all its neighbors, and updated z_w^v accordingly, node v sends messages $z_{(v,w)} = z_v^w$ to all neighbors w to whom no message has been send from v yet.

Verify that

(i) upon initialization, there is at least a node which satisfies the "condition to send a message";
(ii) on every directed edge (v, w), the message $z_{(v,w)}$ is transmitted exactly once, so that $2N - 2$ messages are exchanged over the network in total;
(iii) the algorithm terminates in finite time;
(iv) upon termination, $z_v^v = N$ for all $v \in V$.

Exercise 2.28 (*Consensus on a tree by message-passing*) Message-passing is a powerful approach to distributed computation, at least when G is a tree.

(i) Show that the procedure in Exercise 2.27—with a suitable initialization—may be used to compute the sum of N numbers given at the nodes, $x_v \in \mathbb{R}$ for $v \in V$.
(ii) Design a message-passing algorithm to compute the average of the x_v's.

Exercise 2.29 (*De Bruijn graphs and consensus* [14]) Consider De Bruijn graphs B_h^k on k symbols of dimension h.

(i) Design an algorithm, adapted to a De Bruijn graph, which converges in finite time to consensus.
(ii) How many steps does it take? Compare this value with the diameter and the degree of B_h^k.

Bibliographical Notes

Consensus problems have a very long history in social sciences [8, 13, 18], in statistics [10], and in computer science [40]. Their appearance in the field of control dates back to the thesis work of Tsitsiklis [49, 50] in the 80's, before a surge of interest about fifteen years ago, sparkled by the works [16, 28, 38]. Since then, the literature on the topic has grown enormously, motivated by the broad range of applications: rendezvous, deployment and formation control in robotic coordination [17], flocking of natural and artificial groups [52], load balancing in networks of processors and queues [12], clock synchronization [7, 32], optimal resource allocation [54], distributed optimization [37], distributed computation [24], distributed estimation and learning in sensor networks [44], social network analysis [27], and synchronization of interconnected systems [46, 53]. The averaging dynamics is ubiquitous to these problems (and many others): It is thus unsurprising that several books deal with the topic [5, 6, 31, 33, 41, 42].

This chapter gives a self-contained and comprehensive analysis of the "standard consensus algorithm" on time-invariant networks. In most prior works, its convergence properties are derived from the general theory of nonnegative matrices and in particular as corollaries of the Perron–Frobenius theorem (cf. [19, 45] for two classical references). This choice has two drawbacks. First, it is unnecessary because the needed results can be derived directly in an intuitive way. Second, since Perron–Frobenius theory does not extend to time-dependent networks, it hides the intimate connection between time-invariant and time-varying settings. Instead, the results from this chapter will be the foundation for the rest of the book. The main convergence principle in Lemma 2.1 is based on the presentation in Hendrickx's thesis [22, Sect. 9.2.1]. This principle is crucial and in this chapter we derived from it several properties of stochastic matrices that are central to our theory.

In probability theory, an important reason for the interest in stochastic matrices is the notion of Markov chain associated with it. While we refrain from introducing Markov chains in this text, we believe that the probabilistic interpretation of stochastic matrices is very useful. For instance, it motivates our discussion about the vector π. For these reasons, material on the theory of Markov chains can be a helpful additional reading: We recommend the textbook [30] and the monograph [1] that concentrates on reversible chains. More generally, the general theory of nonnegative matrices is an important background of our work.

Our analysis has highlighted the role of the second eigenvalue of the update matrix $\rho_2(P)$, which determines the speed of convergence of the average dynamics. In turn, the second eigenvalue is closely related to the spectral gap λ_2 of the associated graph. For this reason, graphs with a large spectral gap are of special interest to us. Graphs with large spectral gap are called *spectral expanders* and have been extensively studied in the last decades [2, 3, 43]. A serious study of expander graphs is outside the scope of this book. We only recall that De Bruijn graphs [55], defined in Chap. 1, have good expansion properties that, indeed, have been exploited in consensus problems [14].

In Sect. 2.6, we have shown how to construct stochastic matrices for a given topology. These constructions include the simple random walk and the Metropolis random walk, which are important examples all along the book. Actually, designing stochastic matrices according to certain performance criteria gives rise to a wide family of very interesting problems. For instance, one can look for doubly stochastic matrices, as we did specifically in Proposition 2.6: Paper [20] provides distributed algorithms to solve this design problem. Democracy is a milder requirement that is actually robust to perturbations of the graph, as shown in [11, 15].

Otherwise, one can seek to optimize the speed, that is, minimize the second eigenvalue. This problem is equivalent to finding the fastest mixing Markov chain and has been extensively studied, showing it to be a convex optimization problem [4]. Other (convex) performance metrics will be defined in Chap. 4. A related (more academic) question is finding the slowest possible convergence rate. For SRW matrices, the slowest rate is $1 - \gamma n^{-3}$, as proved in [29].

In a effort to overcome these slow convergence rates (recall also Exercise 2.12), some researchers have designed other distributed algorithms that guarantee a certain convergence speed, irrespective of the graph topology. For instance, the algorithm in [39] has a guaranteed $1 - cn^{-1}$ rate: We refer the reader to that paper also for several pointers to other "accelerated" consensus algorithms.

Our presentation of averaging dynamics has left aside a few topics which have attracted the interest of researchers and which we admit to be important: An incomplete list includes (i) consensus algorithms converging in finite time, (ii) consensus algorithms based on the "message-passing" approach, and (iii) consensus systems evolving in a continuous-time domain. We briefly discuss these natural issues here.

(i) The consensus algorithms presented in this chapter converge to consensus *asymptotically*. One can instead be interested in designing algorithms which converge in a finite number of step (necessarily, not smaller than the diameter). A trivial example is Example 2.6 for the complete graph, while other relatively simple examples can be constructed on De Bruijn graphs, see Exercise 2.29 and [14]. Actually, finite-time convergence can be obtained in more general graphs if we allow the update matrix to change with time (as we shall do in Chap. 3): A simple example are hypercubes [12, Sect. 4], but more general constructions are possible, see [23, 25, 47, 48].

(ii) Message-passing is a paradigm for distributed computation over networks, which we present through simple instances in Exercises 2.27 and 2.28. Nodes are thought of as objects with computational capabilities which can receive messages from their neighbors, elaborate them, and transmit them further. See, e.g., [34] for a general reference and [35] for an application to consensus. Message-passing has also recently found application in problems of leader selection, which will be defined in Chap. 5 [51].

(iii) In our work, we focus on discrete-time dynamics. However, much literature is concerned with continuous-time systems. For the time-invariant networks considered in this chapter, the analysis for continuous time and discrete time is closely related: Actually, the main results about the former can be derived as corollaries of our analysis, see Exercise 2.26. Instead, the analysis in continuous time can become trickier when the network is time-varying (see Chap. 3) or the interactions between nodes are nonlinear. We do not try to survey all the differences here: A few possible references, besides the books mentioned above, are [9, 26, 36].

References

1. Aldous, D., Fill, J.A.: Reversible Markov Chains and Random Walks on Graphs (2002). Unfinished monograph, recompiled 2014
2. Alon, N.: Eigenvalues and expanders. Combinatorica **6**(2), 83–86 (1986)
3. Alon, N., Roichman, Y.: Random Cayley graphs and expanders. Random Struct. Algorithms **5**, 271–284 (1994)
4. Boyd, S., Diaconis, P., Xiao, L.: Fastest mixing Markov chain on a graph. SIAM Rev. **46**(4), 667–689 (2004)
5. Bullo, F.: Lectures on Network Systems. Version 0.95 (2017). With contributions by Cortés, J., Dörfler, F., Martínez, S
6. Bullo, F., Cortés, J., Martínez, S.: Distributed Control of Robotic Networks. Applied Mathematics Series. Princeton University Press (2009)
7. Carli, R., Chiuso, A., Schenato, L., Zampieri, S.: Optimal synchronization for networks of noisy double integrators. IEEE Trans. Autom. Control **56**(5), 1146–1152 (2011)
8. Cartwright, D., Harary, F.: Structural balance: a generalization of Heider's theory. Psychol. Rev. **63**(5), 277 (1956)
9. Ceragioli, F., De Persis, C., Frasca, P.: Discontinuities and hysteresis in quantized average consensus. Automatica **47**(9), 1916–1928 (2011)
10. Chatterjee, S., Seneta, E.: Towards consensus: some convergence theorems on repeated averaging. J. Appl. Probab. **14**(1), 89–97 (1977)
11. Como, G., Fagnani, F.: Robustness of large-scale stochastic matrices to localized perturbations. IEEE Trans. Network Sci. Eng. **2**(2), 53–64 (2015)
12. Cybenko, G.: Dynamic load balancing for distributed memory multiprocessors. J. Parallel Distrib. Comput. **7**(2), 279–301 (1989)
13. DeGroot, M.H.: Reaching a consensus. J. Am. Stat. Assoc. **69**(345), 118–121 (1974)
14. Delvenne, J.-C., Carli, R., Zampieri, S.: Optimal strategies in the average consensus problem. Syst. Control Lett. **58**(10–11), 759–765 (2009)
15. Fagnani, F., Delvenne, J.-C.: The robustness of democratic consensus. Automatica **52**, 232–241 (2015)
16. Fax, J.A., Murray, R.M.: Information flow and cooperative control of vehicle formations. IEEE Trans. Autom. Control **49**(9), 1465–1476 (2004)
17. Francis, B.A., Maggiore, M.: Flocking and Rendezvous in Distributed Robotics. Springer (2015)
18. French, J.R.P.: A formal theory of social power. Psychol. Rev. **63**, 181–94 (1956)
19. Gantmacher, F.R.: The theory of matrices. Chelsea, New York (1959)
20. Gharesifard, B., Cortés, J.: Distributed strategies for generating weight-balanced and doubly stochastic digraphs. Eur. J. Control **18**(6), 539–557 (2012)
21. Golub, B., Jackson, M.O.: Naïve learning in social networks and the hegsel of crowds. Am. Econ. J. Microecon. **2**(1), 112–149 (2010)
22. Hendrickx, J.M.: Graphs and Networks for the Analysis of Autonomous Agent Systems. Ph.D. thesis, Université catholique de Louvain (2008)
23. Hendrickx, J.M., Jungers, R.M., Olshevsky, A., Vankeerberghen, G.: Graph diameter, eigenvalues, and minimum-time consensus. Automatica **50**(2), 635–640 (2014)
24. Hendrickx, J.M., Olshevsky, A., Tsitsiklis, J.N.: Distributed anonymous discrete function computation. IEEE Trans. Autom. Control **56**(10), 2276–2289 (2011)
25. Hendrickx, J.M., Shi, G., Johansson, K.H.: Finite-time consensus using stochastic matrices with positive diagonals. IEEE Trans. Autom. Control **60**(4), 1070–1073 (2015)
26. Hendrickx, J.M., Tsitsiklis, J.N.: Convergence of type-symmetric and cut-balanced consensus seeking systems. IEEE Trans. Autom. Control **58**(1), 214–218 (2013)

27. Jackson, M.O.: Social and Economic Networks. Princeton University Press (2010)
28. Jadbabaie, A., Lin, J., Morse, A.S.: Coordination of groups of mobile autonomous agents using nearest neighbor rules. IEEE Trans. Autom. Control **48**(6), 988–1001 (2003)
29. Landau, H.J., Odlyzko, A.M.: Bounds for eigenvalues of certain stochastic matrices. Linear Algebra Appl. **38**, 5–15 (1981)
30. Levin, D.A., Peres, Y., Wilmer, E.L.: Markov Chains and Mixing Times. American Mathematical Society (2009)
31. Lewis, F.L., Zhang, H., Hengster-Movric, K., Das, A.: Cooperative Control of Multi-agent Systems: Optimal and Adaptive Design Approaches. Springer Science & Business Media (2013)
32. Li, Q., Rus, D.: Global clock syncronization in sensor networks. IEEE Trans. Comput. **55**(2), 214–226 (2006)
33. Mesbahi, M., Egerstedt, M.: Graph Theoretic Methods for Multiagent Networks. Applied Mathematics Series. Princeton University Press, 0 (2010)
34. Mézard, M., Montanari, A.: Information, Physics, and Computation. Oxford University Press (2009)
35. Moallemi, C.C., Van Roy, B.: Consensus propagation. IEEE Trans. Inf. Theory **52**(11), 4753–4766 (2006)
36. Moreau, L.: Stability of continuous-time distributed consensus algorithms. In: IEEE Conference on Decision and Control, Bahamas, pp. 3999–4003, December 2004
37. Nedich, A.: Convergence rate of distributed averaging dynamics and optimization in networks. Found. Trends® Syst. Control, **2**(1), 1–100 (2015)
38. Olfati-Saber, R., Murray, R.M.: Consensus problems in networks of agents with switching topology and time-delays. IEEE Trans. Autom. Control **49**(9), 1520–1533 (2004)
39. Olshevsky, A.: Linear time average consensus on fixed graphs and implications for decentralized optimization and multi-agent control, May 2016
40. Pease, M., Shostak, R., Lamport, L.: Reaching agreement in the presence of faults. J. Assoc. Comput. Mach. **27**(2), 228–234 (1980)
41. Ren, W., Beard, R.W.: Distributed Consensus in Multi-vehicle Cooperative Control Theory and Applications. Springer, New York (2007)
42. Ren, W., Cao, Y.: Distributed Coordination of Multi-agent Networks: Emergent Problems, Models, and Issues. Springer Science & Business Media (2011)
43. Sarnak, P.: What is...an expander? Not. Am. Math. Soc. **51**(7), 762–763 (2004)
44. Sayed, A.H.: Adaptation, learning, and optimization over networks. Found. Trends® Mach. Learn. **7**(4–5), 311–801 (2014)
45. Seneta, E.: Non-negative Matrices and Markov Chains. Springer Series in Statistics. Springer, Berlin (2006)
46. Strogatz, S.H.: From Kuramoto to Crawford: exploring the onset of synchronization in populations of coupled oscillators. Physica D **143**(1–4), 1–20 (2000)
47. Sundaram, S., Hadjicostis, C.N.: Finite-time distributed consensus in graphs with time-invariant topologies. In: American Control Conference, pp. 711–716, July 2007
48. Tran, T.D., Kibangou, A.Y.: Distributed design of finite-time average consensus protocols. In: IFAC Workshop on Estimation and Control of Networked Systems, Koblenz, Germany, pp. 227–233, September 2013
49. Tsitsiklis, J.N.: Problems in Decentralized Decision Making and Computation. Ph.D. thesis, Massachusetts Institute of Technology, November 1984
50. Tsitsiklis, J.N., Bertsekas, D.P., Athans, M.: Distributed asynchronous deterministic and stochastic gradient optimization algorithms. IEEE Trans. Autom. Control **31**(9), 803–812 (1986)
51. Vassio, L., Fagnani, F., Frasca, P., Ozdaglar, A.: Message passing optimization of harmonic influence centrality. IEEE Trans. Control Network Syst. **1**(1), 109–120 (2014)

52. Vicsek, T., Czirók, A., Ben-Jacob, E., Cohen, I., Shochet, O.: Novel type of phase transition in a system of self-driven particles. Phys. Rev. Lett. **75**(6), 1226–1229 (1995)
53. Wieland, P., Sepulchre, R., Allgöwer, F.: An internal model principle is necessary and sufficient for linear output synchronization. Automatica **47**(5), 1068–1074 (2011)
54. Xiao, L., Boyd, S.: Optimal scaling of a gradient method for distributed resource allocation. J. Optim. Theory Appl. **129**(3), 469–488 (2006)
55. Zhang, F.J., Lin, G.N.: On the de Bruijn-Good graphs. Acta Math. Sinica **30**(2), 195–205 (1987)