1

3

5

TRANSLATION INVARIANT: $x(0) \to x(0) + b \Rightarrow x^* \to x^* + b$

$x(t+1) = P\,x(t)$

$\tilde{x}(t+1) = \ldots = P^t(x(0) + c\mathbb{1}) =$

$\tilde{x}(0) = x(0) + c\mathbb{1}$

$\tilde{x}(t+1) = P\tilde{x} \ldots = P^t \tilde{x}(0) =$

$= P^t(x(0) + c\mathbb{1}) =$

$= P^t x(0) + P^t c\mathbb{1}$

$= x^* + c\mathbb{1}$

$= P^t(x(0) + c\mathbb{1})$

$P^t \mathbb{1} = \mathbb{1}$

Chapter 3

Consensus in time-invariant networks

3.1 Rendez-vous and consensus

One of the simplest examples of coordinated control is the so-called *rendezvous problem*. Assume units have dynamics of type $x_v(t+1) = x_v(t) + u_v(t)$ with $x_v(t), u_v(t) \in \mathbb{R}^n$ 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 if 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 $x^* \in \mathbb{R}^n$ such that

$$\lim_{t\to+\infty} x_v(t) = x^*, \quad \forall v \in V \tag{3.1}$$

(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^* 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 "dead-beat 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, 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^* which automatically drops out the dead-beat 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 translate the same way $x^* \to x^* + b$. We will refer to this as to the *translation invariant 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

$$u_v(t) = \sum_{w\in V} S_{vw} x_w(t) \tag{3.2}$$

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

$$x_v(t+1) = \sum_{w\in V} P_{vw} x_w(t) \tag{3.3}$$

where $P = I + S$.

This type of models (3.3) have 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 a same value is a phenomenon known as *consensus*. Later on we will explicitly develop the details of such possible applicative contexts.

Notice that the dimension of the state does not play any particular role in the dynamics (3.3) 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, (3.3) can be rewritten in more compact form simply as

$$x(t+1) = Px(t) \tag{3.4}$$

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 (3.2) is adapted to G if S (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 possible solution:

Theorem 3.1 (Consensus). *Suppose G has a globally reachable vertex v^*. 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) $P\mathbf{1} = \mathbf{1}$; → req. translation invariance

STOCH MATRIX

grafo adapted

modulo self loop.

(Pc) *For every* $v \neq w$, $P_{vw} > 0 \Leftrightarrow (v,w) \in E$;

(Pd) $P_{v^*v^*} > 0$. v^* glob. reach, → grafo connesso.

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 chains theory. Property (Pc) says that G_P and G can only possibly differ for 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 (3.4) as $x(t+1) = x(t) - L(P)x(t)$, which becomes $x_v(t+1) = x_v(t) + \sum_w P_{vw}(x_w(t) - x_v(t))$ component-wise. 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 3.1 and some generalizations, we will work out a special case, which explains how matrices like P above come naturally into the picture and encompasses many interesting examples.

3.2 Consensus 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. 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 it gives us the idea of how to obtain an adapted version of it. Indeed, it is sufficient to replace the barycenter $\overline{x}(t)$ with a local version of it, namely each unit uses a 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

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

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

$$P = I + \tau(D_G^{-1}A_G - I) = I - \tau D_G^{-1}L(G). \tag{3.5}$$

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

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

Notice now that

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

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

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

which can be rewritten as

$$\|P^t x(0) - \overline{x}(0)\mathbf{1}\| \leq \rho_2^t \|(I - N^{-1}\mathbf{1}\mathbf{1}^T x(0)\| \tag{3.6}$$

This shows that if $\rho_2 < 1$, $x(t) = P^t x(0) \to \overline{x}(0)\mathbf{1}$ 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 $\overline{x}(0)$. Moreover, (3.6) formally shows that ρ_2 dictates the speed of convergence of the dynamics towards 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 2.17) $\lambda_N \leq 2d$, a sufficient condition which guarantees consensus is $\tau < 1$

We can summarize the above discussion in the following result.

Proposition 3.2 (Consensus on symmetric regular graphs). *Let G be a symmetric, d-regular, and connected graph. Then the dynamics (3.5)-(3.4), 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 dropping the assumptions of symmetry, regularity, and –to a certain extent– also of connectivity of the underlying graph.

3.3 Stochastic matrices

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 *non-negative matrix*. A non-negative 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 (Sa), (Sb), and (Sc) above 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 the components x_w for those $w \in \mathcal{N}_v^{\text{out}}$. 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 random walk, a special stochastic process taking values on the nodes of a graph and where future only depends from the past through the present state. 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: 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 necessarily, 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 3.3. *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*

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

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:

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

□

The main result reads

Theorem 3.4. *Let $P \in \mathbb{R}^{V\times V}$ be a stochastic matrix such that $\mathcal{G}_P$ admits a globally reachable aperiodic vertex. Then, there exists a vector $\pi \in \mathbb{R}^V$ such that $\pi_v \geq 0$ for all v and $\sum_v \pi_v = 1$, and such that*

$$\lim_{t\to+\infty} P^t = \mathbf{1}\pi^* \tag{3.7}$$

(in other terms P^t converges to a matrix having all rows equal to the row vector π^*).

Proof. Let $s \in V$ be the aperiodic vertex which is reachable from all others. This means that there exists $t^* \in \mathbb{N}$ such that $Q := P^{t^*}$ 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 3.3 implies that

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

Fix now $x^0 \in \mathbb{R}^V$ arbitrarily, and let $x^t = P^t x^0$. Put $M_t = \max_{v\in V} x_v^t$ 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, non-increasing and non-decreasing (hence convergent). Hence also $\Delta_t = M_t - m_t$ converges. For the previous argument, moreover, it holds that $\Delta_{nt^*} \leq (1-\alpha)^n \Delta_0$. This implies that $\Delta_{nt^*} \to 0$ for $n \to +\infty$. Hence, all components of x^t will converge to the same limit. If we apply this result choosing $x_0 = e_w$, the w-th element of the canonical basis of $\mathbb{R}^V$, we obtain that all elements of the w-th column of P^t will converge to the same limit. This clearly yields the result. □

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 3.4 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 3.4.

Theorem 3.4 immediately yields Theorem 3.1. Notice precisely that, for $t \to +\infty$,

$$x(t) = P^t x(0) \to \mathbf{1}\pi^* x(0).$$

In other terms, all components $x_v(t)$ converge to the same consensus point $\pi^* x(0)$ which turns out to be a convex combination of the initial conditions $x_v(0)$.

Remark 3.1 (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 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 a consensus.

3.4 Convergence rate

Theorem 3.4 also contains further information useful to address issue (b) presented above. The next result is about the spectrum of the update matrices.

Corollary 3.5. *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 $\mathbf{1}$;

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

Proof. Suppose indeed that P satisfies the assumptions of Theorem 3.4 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 non trivial Jordan block relative to the eigenvalue 1 will imply that P^t would grow unbounded contrarily to what is asserted in Theorem 3.4. □

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

Corollary 3.6. *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 3.4. 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 above, 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 is unique under the assumptions of Corollary 3.6, but needs not to be unique in general (find an example as an exercise).

Example 3.1 (An irreducible, aperiodic stochastic matrix). Consider the stochastic matrix

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

From the picture of $\mathcal{G}_P$ given in Figure 3.1, 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

$$\pi = \left(\frac{2}{5}, \frac{3}{10}, \frac{3}{10}\right).$$

■

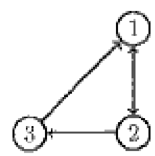

Figure 3.1: The graph of Example 3.1.

The speed of convergence of (3.7) is dictated by the magnitude of the eigenvalues of P. We start recalling the following result which is a standard fact of stability of linear dynamical systems: