# Chapter 4
# Performance and Robustness of Averaging Algorithms

**Abstract** This chapter has the goal of introducing more instruments for the study of consensus algorithms. We will define several performance metrics: Each proposed metric highlights a specific aspect of the algorithm, possibly in relation with a field of application. Namely, we shall consider the speed of convergence in Sect. 4.1, a quadratic control cost in Sect. 4.2, the robustness to noise in Sect. 4.3, and the estimation error in a distributed inference problem in Sect. 4.5. The metrics that we describe share the following feature: under suitable assumptions of symmetry of the update matrix, they can be evaluated as functions of the eigenvalues of the update matrix.

## 4.1 A Deeper Analysis of the Convergence to Consensus

We consider the usual time-invariant consensus dynamics

$$x(t+1) = Px(t), \tag{4.1}$$

where the matrix $P$ is adapted to a strongly connected aperiodic graph $G = (V, E)$ of order $N$. For simplicity, we assume that the matrix $P$ is *symmetric*, although this assumption can be relaxed to some extent. The eigenvalues of $P$ are denoted as $\mu_i$ for $i \in \{1, \dots, N\}$ and $\mu_1 = 1$. We recall that the second eigenvalue is defined as $\rho_2 = \max\{|\mu_i|,\ i = 2, \dots, N\}$. When convenient, we will also make suitable assumptions on the statistics of the initial condition.

In Chap. 2, the speed of convergence to the consensus value of dynamics (4.1) has been estimated in terms of the second eigenvalue $\rho_2$ of the matrix $P$. From Proposition 2.2 and Corollary 2.3, we can recall that the second eigenvalue determines the convergence rate according to the estimate

$$N^{-1}||P^t x(0) - N^{-1}\mathbf{1}\mathbf{1}^* x(0)||^2 \leq \rho_2^{2t} N^{-1}||x(0)||. \tag{4.2}$$


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_4

Notice that we have multiplied both sides of the inequality by $N^{-1}$. Indeed, in the large-scale limit $N \to +\infty$, it makes sense to consider the normalized version of the squared norm $N^{-1}||\cdot||^2$, because $|x(0)_v| \leq \epsilon$ for every $v \in V$ yields $N^{-1}||x(0)||^2 \leq \epsilon^2$.

There are applications, however, where estimate (4.2) turns out to be too loose or simply not adequate to the specific context. We consider the simplest such case, where the initial conditions $x(0)_v$ are assumed to be realizations of independent random variables with mean $m$ and variance $\sigma^2$. From (4.2), by taking the mean value we obtain:

$$N^{-1}\mathbb{E}||P^t x(0) - N^{-1}\mathbf{11}^* x(0)||^2 \leq \sigma^2 \rho_2^{2t}. \tag{4.3}$$

Actually, in this case, it is possible to work out an exact characterization of the mean distance:

$$\begin{aligned}
\frac{1}{N}\mathbb{E}||P^t x(0) - N^{-1}\mathbf{11}^* x(0)||^2 &= \frac{1}{N}\mathbb{E}[\|(P^t - N^{-1}\mathbf{11}^*)x(0)\|^2] \\
&= \frac{1}{N}\mathbb{E}[\mathrm{tr}((P^t - N^{-1}\mathbf{11}^*)x(0)x(0)^*(P^t - N^{-1}\mathbf{11}^*))] \\
&= \frac{1}{N}[\mathrm{tr}((P^t - N^{-1}\mathbf{11}^*)\mathbb{E}[x(0)x(0)^*](P^t - N^{-1}\mathbf{11}^*))] \\
&= \frac{\sigma^2}{N}\,\mathrm{tr}(P^{2t} - N^{-1}\mathbf{11}^*). \qquad (4.4)
\end{aligned}$$

This formula can be rewritten in terms of the Frobenius norm of a square matrix $A$, formally defined as $||A||_F := \sqrt{\mathrm{tr}(AA^*)}$. In our case, we have that

$$||P^t - N^{-1}\mathbf{11}^*||_F = \sqrt{\mathrm{tr}(P^{2t} - N^{-1}\mathbf{11}^*)} = \sqrt{\sum_{i=2}^{N} |\mu_i|^{2t}}.$$

Therefore,

$$N^{-1}\mathbb{E}||P^t x(0) - N^{-1}\mathbf{11}^* x(0)||^2 = \frac{\sigma^2}{N}||P^t - N^{-1}\mathbf{11}^*||_F^2 = \frac{\sigma^2}{N}\sum_{i=2}^{N}|\mu_i|^{2t}. \tag{4.5}$$

Notice how (4.2) can be directly obtained from (4.5) by simply upper bounding $||P^t - N^{-1}\mathbf{11}^*||_F \leq N\rho_2^t$. To illustrate the relation between these two estimates, we propose a few examples.

*Example 4.1* (*Complete graph*) If $P = N^{-1}\mathbf{11}^*$, we trivially have that $P^t - N^{-1}\mathbf{11}^* = 0$ for every $t$ so that $||P^t - N^{-1}\mathbf{11}^*||_F = 0$. On the other hand, we also have that $\rho_2 = 0$. The two estimates coincide in this case.

*Example 4.2* (*Disconnected graph*) Consider now the simple random walk associated with a disconnected graph consisting of two complete isolated graphs with $N/2$

nodes each ($N$ is assumed to be even):

$$P = \left[\begin{array}{c|c} \frac{2}{N}\mathbf{1}\mathbf{1}^* & 0 \\ \hline 0 & \frac{2}{N}\mathbf{1}\mathbf{1}^* \end{array}\right]. \tag{4.6}$$

The eigenvalues of $P$ are 1 with multiplicity 2 and 0 with multiplicity $N-2$. Therefore,

$$||P^t - N^{-1}\mathbf{1}\mathbf{1}^*||_F = 1, \quad \text{and} \quad \rho_2 = 1.$$

In this case, expressions (4.5) and (4.2) respectively become, for $t \geq 1$,

$$N^{-1}\mathbb{E}||P^t x(0) - N^{-1}\mathbf{1}\mathbf{1}^* x(0)||^2 = \sigma^2 N^{-1},$$
$$N^{-1}\mathbb{E}||P^t x(0) - N^{-1}\mathbf{1}\mathbf{1}^* x(0)||^2 \leq \sigma^2.$$

Clearly, they are significantly different in terms of $N$. Notice in particular that the first bound says that, for large $N$, the mean distance from consensus is small for every value of $t \geq 1$. This difference seems in contrast with the fact that the simple random walk on a disconnected graph does not lead to a consensus. However, notice that the consensus values on the two components are both small for large $N$ with high probability because of the law of large numbers, and therefore, the two consensus values are close to each other.

Perhaps more interestingly, this inconsistency between the two estimates is not limited to disconnected graphs, as we show in the following example, which is in fact a slight modification of the previous one.

*Example 4.3* (*Barbell graph*) A *barbell graph*, defined for even $N$, is a graph composed of two disjoint cliques connected by an edge. The SRW is now

$$\tilde{P} = P + \left[\begin{array}{c|c} & \\ -2/N & 2/N \\ \hline 2/N & -2/N \\ & \end{array}\right],$$

where $P$ is the SRW (4.6). Matrix $\tilde{P}$ has eigenvalue 1 with multiplicity 1, eigenvalue 0 with multiplicity $N-3$ and two simple eigenvalues $\frac{1}{2} - \frac{2}{N} \pm \frac{1}{2}\sqrt{1 + \frac{8}{N} - \frac{16}{N^2}}$. Here, we rare facing a bottleneck phenomenon due to the single edge connecting the two cliques and this results in a very slow convergence rate $\rho_2(P) = 1 - \frac{8}{N^2} + o(\frac{1}{N^2})$ as $N \to \infty$. Nevertheless, for all $t \geq 1$, it holds

$$||P^t - N^{-1}\mathbf{1}\mathbf{1}^*||_F^2 = \left(\frac{1}{2} - \frac{2}{N} + \frac{1}{2}\sqrt{1 + \frac{8}{N} - \frac{16}{N^2}}\right)^{2t} + \left(\frac{1}{2} - \frac{2}{N} - \frac{1}{2}\sqrt{1 + \frac{8}{N} - \frac{16}{N^2}}\right)^{2t}$$
$$\leq \frac{2}{N}.$$

As in the previous example, the estimation error becomes small already from the first iteration if $N$ is large, but this cannot be seen in the estimation that uses the second eigenvalue $\rho_2$.

## 4.2 Rendezvous and Linear-Quadratic Control

The mean convergence rate introduced in Sect. 4.1 is related to the analysis of various other cost functionals. In this section, we consider the consensus dynamics in the context of the rendezvous application, interpreting it as a closed-loop feedback control:

$$x(t+1) = x(t) + u(t) \quad \text{where} \quad u(t) = (P - I)x(t).$$

In this, setting a popular cost functional to measure the control performance of the systems is the quadratic cost defined as $J_{\text{LQ}} := J_x + \epsilon J_u$, where $\epsilon$ is a positive weight and

$$J_x := N^{-1} \sum_{t=0}^{\infty} \mathbb{E}||x(t) - N^{-1}\mathbf{1}\mathbf{1}^* x(0)||^2 \tag{4.7}$$

$$J_u := N^{-1} \sum_{t=0}^{\infty} \mathbb{E}||u(t)||^2. \tag{4.8}$$

Cost $J_x$ measures the speed of convergence to consensus, whereas $J_u$ measures the control effort needed to achieve it. The two functionals $J_x$ and $J_u$ can be expressed in terms of the eigenvalues $\{\mu_i\}$ as shown in the following result.

**Proposition 4.1** (LQ cost) *If the stochastic matrix $P$ is irreducible aperiodic and symmetric, then*

$$J_x = \frac{\sigma^2}{N} \sum_{t=0}^{\infty} ||P^t - N^{-1}\mathbf{1}\mathbf{1}^*||_F^2 = \frac{\sigma^2}{N} \sum_{i=2}^{N} \frac{1}{1 - \mu_i^2}$$
$$J_u = \frac{\sigma^2}{N} \sum_{t=0}^{\infty} ||P^{t+1} - P^t||_F^2 = \frac{\sigma^2}{N} \sum_{i=2}^{N} \frac{1 - \mu_i}{1 + \mu_i}$$

*Proof* The expression for $J_x$ is a straightforward consequence of (4.5). Regarding $J_u$, the first equality comes from a computation analogous to (4.4). The second one instead follows from the observation that the eigenvalues of $(P^{t+1} - P^t)^2$ are given by $\{(\mu_i^{t+1} - \mu_i^t)^2\} = \{\mu_i^{2t}(1 - \mu_i)^2\}$. □

The values of $J_x$ and $J_u$ can be effectively estimated or computed in several examples. In general, it is immediate to see that

$$\frac{N-1}{N} \leq \frac{J_x}{\sigma^2} \leq \frac{1}{1-\rho_2^2},$$

where both bounds are tight (take $P = N^{-1}\mathbf{1}\mathbf{1}^*$). The lower bound implies that $J_x$ is never infinitesimal in the number of nodes, while the upper bound implies that $J_x$ is limited if the second largest eigenvalue of $P$ is bounded away from one. Otherwise, $J_x$ may or may not diverge as $N$ goes to infinity, as shown in the example below.

Let us consider the lazy simple random walk matrix $P = (2d+1)^{-1}(I + A)$ on a $d$-dimensional torus $C_n^d$. Eigenvalues can easily be obtained from the eigenvalues of $L(A)$ computed in Example 1.8:

$$\mu_{(h_1,\dots,h_d)} = \frac{1}{2d+1}\left(1 + 2\sum_{i=1}^{d} \cos\frac{2\pi}{n}h_i\right), \quad h_1, \dots, h_d \in \{0, 1, \dots, n-1\}$$

Notice now that

$$J_x = \frac{\sigma^2}{n^d} \sum_{(h_1,\dots h_d)\neq 0} \frac{1}{1 - |\mu_{(h_1,\dots,h_d)}|^2}$$

can be interpreted as a Riemann sum of the function $f : [0,1]^d \setminus \{0\} \to \mathbb{R}$ given by

$$f(x) = \frac{\sigma^2}{1 - \left|\frac{1}{2d+1}\left(1 + 2\sum_{i=1}^{d}\cos 2\pi x_i\right)\right|^2}$$

Notice that $f(x)$ presents a singularity in 0: Precisely, we have that $f(x) = \Theta(||x||^{-2})$ for $x \to 0$. This implies that in dimension $d \geq 3$, function $f$ is (absolutely) integrable on $[0,1]^d$. This, combined with the fact that $f(x)$ is monotonic with respect to the each component of $x$ in a neighborhood of 0, implies that

$$\lim_{n\to+\infty} J_x = \int_{[0,1]^d} f(x)\mathrm{d}x < +\infty$$

In particular this shows that on a $d$-dimensional torus, with $d \geq 3$, $J_x$ is bounded in $N$. Instead, in dimension 1 and 2, $f$ is no longer integrable and previous argument cannot be applied. Indeed in both cases, $J_x$ turns out to be unbounded in $N$. In dimension one, an explicit computation shows that (letting $\sigma = 1$):

$$J_x = \frac{1}{N}\sum_{h=1}^{N-1} \frac{1}{1-\frac{1}{9}(1+2\cos\left(\frac{2\pi}{N}h\right))^2} \geq \frac{N^{-1}9/4}{2-\cos\left(\frac{2\pi}{N}\right)-\cos^2\left(\frac{2\pi}{N}\right)} \geq \frac{3}{8\pi^2}N. \tag{4.9}$$

A more detailed analysis including the two-dimensional case is provided in Exercise 5.11, by using the tools developed in that chapter.

On the contrary, $J_u$ shows better scaling properties: for the lazy SRW on the cycle,

$$J_u = \frac{1}{N}\sum_{h=1}^{N-1} \frac{1-\cos\left(\frac{2\pi}{N}h\right)}{2+\cos\left(\frac{2\pi}{N}h\right)},$$

which is clearly bounded in $N$. By interpreting it as a Riemann sum one can see that, more precisely,

$$\lim_{n\to+\infty} J_u = \int_0^1 \frac{1-\cos(2\pi x)}{2+\cos(2\pi x)}\mathrm{d}x = \sqrt{3}-1.$$

Other examples are given in Exercise 4.10. More generally, $J_u$ can be shown to be bounded under weak assumptions. To this goal, we recall a well-known property of the spectrum of a matrix.

**Lemma 4.1** (Gershgorin) *Let $A$ be an $n \times n$ matrix. Then,*

$$\mathrm{spec}(A) \subset \bigcup_{i\in\{1,\dots,n\}} \{z \in \mathbb{C} \,:\, |z-a_{ii}| \leq \sum_{j\neq i} |a_{ij}|\}.$$

An immediate application of this lemma yields the following result.

**Proposition 4.2** (Boundedness of $J_u$) *Let $P$ be such that $P_{vv} > 0$ for all $v \in V$, and denote $\alpha = \min_v P_{vv}$. Then,*

$$J_u \leq \frac{1-\alpha}{\alpha}.$$

## 4.3 Robustness Against Noise

In this section, we analyze the behavior and performance of consensus algorithms under the presence of noise in the dynamics. As we will see, cost functionals similar to those introduced above naturally come up in this case. Noise is unavoidable in many applications. Instances can be imprecisions in the motion of robots in the rendezvous problem or quantization errors in digital transmissions among the nodes of the network. In this section, we analyze the effects of noise in several models where the consensus dynamics is perturbed in different ways. We recall the standing assumption that $P$ is a symmetric stochastic matrix.

We start considering the case when noise enters additively in the update equation (this can be a model for the robots motion error):

$$x(t+1) = Px(t) + n(t) \tag{4.10}$$

We assume the $n_v(t)$ to be independent random variables with mean 0 and variance $\sigma^2$. Notice first that the mean value is governed by $\mathbb{E}[x(t+1)] = P\mathbb{E}[x(t)]$ so that, if $P$ is irreducible and aperiodic, we have convergence to the average consensus: $\mathbb{E}[x(t)] \to N^{-1}\mathbf{11}^*x(0)$ (here, $x(0)$ is seen as deterministic). If we define $m(t) = N^{-1}\sum_v x_v(t)$ and $\mu(t) = N^{-1}\sum_v n_v(t)$, Eq. (4.10) implies that

$$m(t+1) = m(t) + \mu(t). \tag{4.11}$$

Consequently,

$$m(t) = m(0) + \sum_{s=0}^{t-1}\mu(s)$$

Notice that each $\mu(t)$ is a r.v. with mean 0 and variance $\sigma^2/N$. Therefore, we can conclude that $m(t)$ is a process with

$$\mathbb{E}[m(t)] = m(0), \quad \mathrm{Var}[m(t)] = \sigma^2\frac{t}{N}. \tag{4.12}$$

This shows how noise accumulates into the linear dynamics (essentially because of its marginally stable structure) and creates such unbounded effects on the average dynamics. A similar phenomenon takes place if we measure the distance of the process from a consensus point. Consider indeed the following functional

$$J_{\mathrm{noise}} = \frac{1}{N}\lim_{t\to+\infty}\mathbb{E}||x(t) - N^{-1}\mathbf{11}^*x(t)||^2 \tag{4.13}$$

Remarkably, $J_{\mathrm{noise}}$ coincides with the functional $J_x$ introduced to describe the LQ cost functional.

**Proposition 4.3** (Noise cost) *Suppose that $P$ is a symmetric irreducible and aperiodic stochastic matrix. Then,* $J_{\mathrm{noise}} = J_x = \frac{\sigma^2}{N}\sum_{t=0}^{+\infty}||P^{2t} - N^{-1}\mathbf{11}^*||_F^2$.

*Proof* It follows from (4.10) that, for every time $t$, it holds

$$x(t) = P^t x(0) + \sum_{s=0}^{t-1} P^s n(t-s-1)$$

which yields

$$
\begin{aligned}
\mathbb{E}||x(t) - N^{-1}\mathbf{11}^* x(t)||^2 &= \mathbb{E}||(P^t - N^{-1}\mathbf{11}^*)x(0)||^2 \\
&+ \sum_{s=0}^{t-1}\sum_{s'=0}^{t-1} \mathbb{E}[(P^s - N^{-1}\mathbf{11}^*)n(t-s-1)]^*[(P^{s'} - N^{-1}\mathbf{11}^*)n(t-s'-1)] \\
&+ 2\sum_{s=0}^{t-1} \mathbb{E}[((P^t - N^{-1}\mathbf{11}^*)x(0))^*(P^s - N^{-1}\mathbf{11}^*)n(t-s-1)]
\end{aligned}
$$

Now, the first term converges to 0, when $t \to +\infty$, because of the assumptions made on $P$. The third term is 0 because all noises are zero mean. Finally, the second term can be rewritten as

$$
\sigma^2 \sum_{s=0}^{t-1} \mathrm{tr}[P^{2s} - N^{-1}\mathbf{11}^*] = \sigma^2 \sum_{s=0}^{t-1} ||P^s - N^{-1}\mathbf{11}^*||_F^2
$$

where we have used the independence assumption on the noises. By taking the limit $t \to +\infty$, we obtain the result. □

The example in Eq. (4.9) implies then that $J_{\text{noise}}$ is in general unbounded in $N$ for large-scale graphs. Moreover, Eq. (4.12) shows that the variance of the average value diverges with time. These considerations demonstrate that consensus dynamics is sensitive to additive noise: This sensitivity is stronger for matrices $P$ that have eigenvalues closer to the unit circle. Actually, when the dispersion of the initial condition is small with respect to the variance of the noise and to the number of nodes $N$, running a consensus algorithm may even be detrimental in terms of $\mathbb{E}||x(t) - N^{-1}\mathbf{11}^* x(t)||^2$. These circumstances are explored in Exercise 4.8. Even outside such extreme cases, sensitivity to noise can be a problem in practical applications and some countermeasures against noise have thus been proposed. A useful idea is replacing the time-invariant averaging dynamics with a time-varying version that smooths out the effects of noise by employing a "decreasing gain" strategy. We do not cover these more refined algorithms here, but some literature pointers are given at the end of this chapter.

## 4.4 Robustness Against Quantization Errors

If we assume that communication among units takes place through digital channels, then the communicated states will be affected by rounding (or quantization) errors. These errors are unavoidable because the state is real-valued, whereas the communicated values are discrete. These errors, which depend on the state and on the quantization rule, can be modeled as independent stochastic noises with zero mean and with variance $\sigma^2$ determined by the precision of the approximation. This modeling leads to consider a dynamics like

$$
x(t+1) = P\big(x(t) + n(t)\big). \tag{4.14}
$$

It is easy to realize that this dynamics suffers from the same drawbacks as dynamics (4.10). In particular, the average process $m(t)$ is governed by the same relation (4.11), and consequently, the same conclusions on the moments (4.12) can be drawn. Moreover, the functional defined as in (4.13) can be evaluated similarly to Proposition 4.3, as detailed in Exercise 4.7.

Notice, however, that in this case one can run, instead of (4.14), the alternative averaging dynamics

$$x(t+1) = P\big(x(t) + n(t)\big) - n(t). \tag{4.15}$$

In this dynamics, we subtract the noise $n(t)$: This operation is feasible as it is realistic to assume that each node $v$ knows $n_v(t)$, that is, the quantization error affecting its own value. This dynamics is chosen with the purpose of reducing the effect of the noise. Indeed, differently from (4.10), dynamics (4.15) deterministically preserves the average of the initial condition:

$$N^{-1}\mathbf{1}\mathbf{1}^* x(t+1) = N^{-1}\mathbf{1}\mathbf{1}^* x(t) + N^{-1}\mathbf{1}\mathbf{1}^* n(t) - N^{-1}\mathbf{1}\mathbf{1}^* n(t) = N^{-1}\mathbf{1}\mathbf{1}^* x(t).$$

The asymptotical dispersion around the average can be evaluated by using the functional

$$J_{\mathrm{q}} = \lim_{t\to+\infty} \mathbb{E}||x(t) - N^{-1}\mathbf{1}\mathbf{1}^* x(t)||^2,$$

which is formally defined as (4.13) but with the understanding that here $x(t)$ follows (4.15). Perhaps surprisingly, $J_{\mathrm{q}}$ coincides with the functional $J_u$ introduced to describe the LQ cost functional, as the reader can verify as an exercise.

**Proposition 4.4** (Quantization cost) *Suppose that $P$ is a symmetric irreducible and aperiodic stochastic matrix. Then,* $J_{\mathrm{q}} = J_u = \frac{\sigma^2}{N}\sum_{t=0}^{\infty} ||P^{t+1} - P^t||_F^2$.

By recalling the results of Sect. 4.2, the reader can see that the effect of noise is largely reduced in (4.15), compared to (4.14).

## 4.5 Distributed Inference

An important application of consensus is solving, in a distributed fashion, network inference problems. Below we discuss some basic examples and we show how the analysis of performance, also in this case, leads to functionals similar to those considered before.

Assume that each node $v \in V$ takes a measurement of the same unknown scalar quantity $\theta$. Each of these measurements, denoted by $y_v$, is affected by an (additive) measurement error $n_v$. Namely, $y_v = \theta + n_v$. The goal of each node is to estimate $\theta$. The node $v$ by itself could only estimate $\theta$ by the taken measurement $y_v$, whereas if it was possible to gather the measurements from all nodes, more efficient estimation could be performed. If we assume the measurement errors to be independent random variables with zero mean and variance $\sigma^2$, then the average $\hat{\theta} = \frac{1}{N}\sum_v y_v$ is the optimal estimator for $\theta$ and has estimation error $\mathbb{E}||\hat{\theta} - \theta||^2 = N^{-1}\sigma^2$. Conveniently, the network may collectively compute $\hat{\theta}$ by simply using the dynamics (4.1) with $x_v(0) = y_v$ for all nodes $v$. In this context, it is natural to define the time-dependent *estimation error* as

$$J_{\mathrm{e}}(t) = \frac{1}{N}\mathbb{E}[\|x(t) - \theta\mathbf{1}\|^2].$$

If we denote by $n \in \mathbb{R}^V$ the random vector collecting all noises and we repeat the computation as in (4.4), we see that $J_{\mathrm{e}}(t)$ can be rewritten as

$$J_{\mathrm{e}}(t) = \frac{1}{N}\mathbb{E}[\|P^t n\|^2] = \sigma^2 N^{-1}||P^t||_F^2 = \sigma^2 N^{-1}\sum_{i=1}^{N}\mu_i^{2t} \tag{4.16}$$

Notice the difference with respect to (4.5), where we had the Frobenius norm of $P^t - N^{-1}\mathbf{1}\mathbf{1}^*$. Indeed, differently from (4.5) that converges to 0 for $t \to +\infty$, we here have $J_{\mathrm{e}}(t) \to \frac{\sigma^2}{N}$ as $t \to \infty$. The asymptotic error is due to the intrinsic mean estimation error. The consensus algorithm (4.1) can also be used to solve more general inference problems, in which the measurements errors can have different variances $\sigma_v^2$. Again, the goal of each node is to estimate $\theta$. The node $v$ by itself could only estimate $\theta$ by the taken measurement $y_v$, whereas if it was possible to gather the measurements from all nodes, more efficient estimation could be performed. In the latter case, the best least squares estimator, defined as

$$\widehat{\theta} := \mathrm{argmin}_\theta \sum_v \frac{(y_v - \theta)^2}{\sigma_v^2},$$

can be computed as

$$\widehat{\theta} = \left(\sum_{w\in V}\frac{1}{\sigma_w^2}\right)^{-1}\sum_{v\in V}\frac{y_v}{\sigma_v^2}.$$

This estimator, which is a Maximum Likelihood estimator when the measurement noises are Gaussian (see Exercise 4.1), simply becomes the average of the measurements when all variances are equal. Clearly, to compute such an estimator, one needs to gather all the measurements $y_v$'s. However, rewriting it as

$$\widehat{\theta} = \left(\frac{1}{N}\sum_{w\in V}\frac{1}{\sigma_w^2}\right)^{-1}\frac{1}{N}\sum_{v\in V}\frac{y_v}{\sigma_v^2},$$

one can notice that it is the ratio between two arithmetic means. Then, consensus algorithms can be naturally applied to approximate it, provided each node knows the variance of its own measurement error. Consider two consensus algorithms built on the matrix $P$ and running in parallel:

$$x^{(1)}(t+1) = Px^{(1)}(t), \quad x_v^{(1)}(0) = \frac{y_v}{\sigma_v^2} \quad \forall v \in V$$
$$x^{(2)}(t+1) = Px^{(2)}(t), \quad x_v^{(2)}(0) = \frac{1}{\sigma_v^2} \quad \forall v \in V$$

and define $\widehat{\theta}_v(t) = x_v^{(1)}(t)/x_v^{(2)}(t)$. We know from the results of Chap. 2 (see also Exercise 4.2) that

$$\lim_{t\to+\infty} \widehat{\theta}(t)_v = \widehat{\theta} \quad \forall v \in V.$$

Hence, the estimator can be computed by running two consensus algorithms and computing the ratio of their states. In the exercises, we propose a few adaptations and variations of the above procedure.

## Exercises

**Exercise 4.1** (*Maximum Likelihood estimator*) For all $v \in V$, let $y_v = \theta + n_v$ and assume that each $n_v$ is a measurement error to be independent Gaussian random variables with zero mean and variance $\sigma_v^2$. Consider the density distribution of $y_v$ given that the unknown quantity is $\theta$

$$f(y_v \mid \theta) = \frac{1}{\sqrt{2\pi\sigma_v^2}} e^{\frac{(y_v-\theta)^2}{2\sigma_v^2}}$$

and the global density of the vector $y \in \mathbb{R}^V$

$$f(y \mid \theta) = \prod_v f(y_v \mid \theta) = \prod_v \frac{1}{\sqrt{2\pi\sigma_v^2}} e^{\sum_v \frac{(y_v-\theta)^2}{2\sigma_v^2}}$$

The ML estimator is defined to be $\widehat{\theta}^{ML} := \text{argmax}_{\theta\in\mathbb{R}} f(y \mid \theta)$. Verify that

$$\widehat{\theta}^{ML} = \left( \sum_{w\in V} \frac{1}{\sigma_w^2} \right)^{-1} \sum_{v\in V} \frac{y_v}{\sigma_v^2}$$

**Exercise 4.2** (*Consensus ratio*) Let $P \in \mathbb{R}^{V\times V}$ be a stochastic irreducible aperiodic matrix. Consider the dynamics

$$\begin{cases} x(t+1) = P\,x(t) & x(0) \in \mathbb{R}^V \\ y(t+1) = P\,y(t) & y(0) \in \mathbb{R}^V,\, y_v(0) > 0\ \forall v \in V. \end{cases}$$

Let $z(t) = \frac{x(t)}{y(t)}$. Determine $z(\infty) := \lim_{t\to+\infty} z(t)$.

**Exercise 4.3** (*Average consensus with nondoubly stochastic matrices*) Let $P \in \mathbb{R}^{V\times V}$ be a stochastic irreducible aperiodic matrix. Consider the dynamics

$$\begin{cases} x(t+1) = P^* x(t) & x(0) \in \mathbb{R}^V \\ y(t+1) = P^* y(t) & y(0) \in \mathbb{R}^V,\ y_v(0) > 0\ \forall v \in V. \end{cases}$$

(i) Let $z(t) = \frac{x(t)}{y(t)}$. Determine $z(\infty) := \lim_{t\to+\infty} z(t)$.
(ii) Given a vector $\bar{x} \in \mathbb{R}^V$, choose $x(0)$ and $y(0)$ in such way that $z(\infty) = \mathbf{1}N^{-1}\mathbf{1}^*\bar{x}$.

**Exercise 4.4** (*Weighted averages*) Using a doubly stochastic matrix $P$, design a consensus-based algorithm to compute any weighted average of values known to the nodes.

**Exercise 4.5** (*Number of nodes*) Design a consensus-based algorithm for a network to compute the number of its nodes.

**Exercise 4.6** (*Least Squares Regression*) We want to estimate a function $y = f(x)$ from a noisy data set $\{(x_v, y_v)\}_{v\in V}$ collected by the nodes.We parameterize $f(\cdot)$ according to a basis of functions $\{g_j(\cdot)\}_{j\in J}$, where $J$ a suitable index set, so that $f_\theta(x) = \sum_{j\in J} \theta_j g_j(x)$. The basis functions are known, and the $|J|$-dimensional vector $\theta$ is to be determined. We want to compute (distributely) the best estimate of $\theta$ in a least squares sense. Provided we define $G \in \mathbb{R}^{J\times V}$ to be a matrix such that $G_{jv} = g_j(x_v)$, the optimal estimator is defined as

$$\hat{\theta} = \operatorname{argmin}_\theta \|y - G\theta\|^2.$$

(i) Verify that $\hat{\theta} = (G^*G)^{-1}G^*y$.

Let $g^v$ denote a column of $G$ and define two consensus algorithms with initial conditions

$$\begin{aligned} z_v^{(1)}(0) =& g^v(g^v)^* \quad \forall v \in V \\ z_v^{(2)}(0) =& g^v y_v \quad \forall v \in V. \end{aligned}$$

Note that $z_v^{(1)} \in \mathbb{R}^{J\times J}$ and $z_v^{(2)} \in \mathbb{R}^J$. Since the states are non scalar, the update is performed independently on each component.

(ii) Remark that $\hat{\theta} = (\sum_v g^v(g_v)^*)^{-1} \sum_v g^v y_v$, and deduce that

$$\lim_{t\to+\infty} (z_v^{(1)}(t))^{-1} z_v^{(2)}(t) = \hat{\theta}.$$

**Exercise 4.7** (*Communication noise*) Consider a symmetric stochastic matrix $P$ on $V$ and the process $x(t)$ taking values in $\mathbb{R}^V$ and governed by equation

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

where $\{n_{vw})(t)\}$ is a family of independent 0 mean, $\sigma^2$ variance random variables. In this model, we assume that noises are independently generated in any pairwise transmission between units. For such process, study the behavior of the corresponding average process $m(t)$ and find an expression for the functional as defined in (4.13) in terms of the eigenvalues of $P$.

**Exercise 4.8** (*Noise effects at finite times*) We define the following time-dependent version of the noise cost (4.13)

$$J_{\text{noise}}(t) = \frac{1}{N}\mathbb{E}||x(t) - N^{-1}\mathbf{1}\mathbf{1}^* x(t)||^2, \tag{4.17}$$

assuming that the noise components are iid random variables with zero mean and variance $\sigma_n^2$, while initial conditions are iid random variables with zero mean and variance $\sigma_x^2$ independent of the noise.

(i) By proceeding as in the proof of Proposition 4.3, show that

$$J_{\text{noise}}(t) = \frac{1}{N}\sum_{i=2}^{N}\frac{\sigma_n^2}{1-|\mu_i|^2} + \frac{1}{N}\sum_{i=2}^{N}\left(\sigma_x^2 - \frac{\sigma_n^2}{1-|\mu_i|^2}\right)|\mu_i|^{2t} \tag{4.18}$$

(ii) Observe that if $\sigma_x$ is small enough (for instance if $\sigma_x < \sigma_n$), then $J_{\text{noise}}(t)$ is increasing with time.

(iii) Verify that the second term of (4.18) is upper bounded by $\frac{1}{N}\left(N\sigma_x^2 - \frac{\sigma_n^2}{1-\rho_2^2}\right)\rho_2^{2t}$

(iv) Assume that $P$ is the lazy simple random walk matrix $P$ on the cycle graph $C_N$ (cf. Exercise 2.12). Verify that if $N > \frac{8\pi}{9}\frac{\sigma_x^2}{\sigma_n^2}$, then $J_{\text{noise}}(t)$ is increasing with time.

**Exercise 4.9** (*Normal update matrix*) Reconsider system (4.1) with the assumption that the irreducible and aperiodic matrix $P$ is doubly stochastic and normal (but not necessarily symmetric). Show that

$$J_{\text{e}}(t) = \frac{1}{N}\sum_{i}|\mu_i|^{2t} \quad J_x = \frac{1}{N}\sum_{i>1}\frac{1}{1-|\mu_i|^2} \quad J_u = \frac{1}{N}\sum_{i>1}\frac{|1-\mu_i|^2}{1-|\mu_i|^2}.$$

**Exercise 4.10** ($J_u$ *cost* [11]) Consider the cost $J_u$ defined in Sect. 4.2.

(i) Let $G = \mathbf{C}_N$ be a directed cycle graph and $P = \text{circ}(1/2, 1/2, 0, \dots, 0)$. Then, $J_u = 1 - \frac{1}{N}$.

(ii) Let $G = \mathbf{C}_n^d$ be a directed $d$-dimensional torus graph and $P = \frac{1}{d+1}(I + A_G)$. Then, $J_u = 1 - \frac{1}{n^d}$.

(iii) Let $G$ be a $d$-dimensional hypercube and $P = \frac{1}{d+1}(I + A_G)$. Then, $J_u = 1 - \frac{1}{2^d}$.

**Exercise 4.11** ($J_{\text{e}}$ *cost on toroidal grids* [14]) Let $G = C_n^d$ be a $d$-dimensional torus graph, $P = \frac{1}{2d+1}(I + A_G)$ and $N = n^d$.

(i) Verify that

$$c_1 \max\{\frac{1}{N}, \frac{1}{t^{d/2}}\} \leq J_{\mathrm{e}}(t) \leq c_2 \max\{\frac{1}{N}, \frac{1}{t^{d/2}}\} \quad \text{for some positive } c_1, c_2.$$

(ii) Estimate the time needed to achieve the best precision in the estimation for a given $N$.

**Exercise 4.12** (*Size optimization in distributed estimation* [14]) Based on the functional $J_{\mathrm{e}}(t)$, we consider the problem of optimizing the size of the graph in a specific family of consensus matrices. Let $A_n$ be the adjacency matrix of cycle graph $C_n$ of order $n$. Let $P_n = \frac{1}{3}(I + A_n)$.

(i) Verify that

(a) $J(P_n, t)$ is nonincreasing in $n$;
(b) $J(P_n, t)$ is nonincreasing in $t$;
(c) $J(P_n, t) = J(P_{2t+1}, t)$ for all $n \geq 2t + 1$.

(ii) Discuss the results above from the point of view of design, having the goal of efficient estimation of a parameter which is known via noisy measurements. Is there a "best" size of the network, if the available time for computation is limited?

## Bibliographical Notes

An introduction to the costs considered in this chapter is available in [13], which also provides useful pointers to the literature. These costs have been explicitly computed using the eigenvalues of $P$, thanks to the assumption of symmetry: However, one can generalize this analysis to normal matrices (see Exercise 4.9 and [7]) and to reversible matrices (see Exercise 5.10).

Specific references can be given for the different functionals considered. For instance, a thorough analysis of $J_{\mathrm{e}}$ on geometric graphs is given by [14]. Cost $J_{\mathrm{noise}}$ has been defined in the seminal paper [25] and later extensively studied with different interpretations and variations [17, 21]. Paper [11] has proposed dynamics (4.14) to cope with quantization errors and has studied cost $J_{\mathrm{q}}$. The statistical assumptions on the quantization errors can either be rigorously justified for certain randomized quantizers (for instance, quantizers with "dithering" [1]) or taken as a useful approximation for deterministic quantizers. Actually, several researchers have looked at quantization in the context of averaging algorithms, starting with [19]: A selection of the papers that are most closely related to our perspective includes [3, 6, 9, 18, 20, 22].

As we mentioned, the effects of noise entering the averaging system can be mitigated by using properly designed decreasing gains. This adaptation typically results in systems that almost surely converge to consensus, but such that convergence is not exponentially fast. Their analysis can benefit from the so-called stochastic approximation techniques [2]. Many papers have taken this approach to ensure robustness, including [5, 8, 16, 23, 24].

The application of distributed parameter estimation has also been presented in this chapter. In the literature, this problem has been extended in various directions, including least squares regression (see Exercise 4.6), distributed Kalman filtering [4], and estimation of parameters that are vector-valued and distributed over the nodes (see Sect. 5.4). The treatment given in this chapter assumes that each node knows the variance of its own measurement error. If this is not the case, a more complex algorithm is needed in order to estimate these quantities as well. For instance, paper [10] looks at a special case of this problem, where nodes are divided into two classes, having respectively small and large variance, and must identify to which class they belong to. The issue of parameter estimation can also be interpreted in the context of social networks. Empirical and theoretical evidences have shown that aggregate opinions may provide a good estimate of unknown quantities: Such phenomenon has been referred to in the literature as the wisdom of crowds [12, 15].

## References

1. Aysal, T.C., Coates, M.J., Rabbat, M.G.: Distributed average consensus with dithered quantization. IEEE Trans. Signal Process. **56**(10), 4905–4918 (2008)
2. Borkar, V.S.: Stochastic Approximation: A Dynamical Systems Viewpoint. Cambridge University Press (2008)
3. Carli, R., Bullo, F., Zampieri, S.: Quantized average consensus via dynamic coding/decoding schemes. Int. J. Robust Nonlinear Control **20**(2), 156–175 (2010)
4. Carli, R., Chiuso, A., Schenato, L., Zampieri, S.: Distributed Kalman filtering based on consensus strategies. IEEE J. Sel. Areas Commun. **26**(4), 622–633 (2008)
5. Carli, R., Como, G., Frasca, P., Garin, F.: Distributed averaging on digital erasure networks. Automatica **47**(1), 115–121 (2011)
6. Carli, R., Fagnani, F., Speranzon, A., Zampieri, S.: Communication constraints in the average consensus problem. Automatica **44**(3), 671–684 (2008)
7. Carli, R., Garin, F., Zampieri, S.: Quadratic indices for the analysis of consensus algorithms. In: Information Theory and Applications Workshop, San Diego, CA, pp. 96–104, February 2009
8. De Persis, C., Frasca, P.: Robust self-triggered coordination with ternary controllers. IEEE Trans. Autom. Control **58**(12), 3024–3038 (2013)
9. El Chamie, M., Liu, J., Başar, T.: Design and analysis of distributed averaging with quantized communication. IEEE Trans. Autom. Control **61**(12), 3870–3884 (2016)
10. Fagnani, F., Fosson, S.M., Ravazzi, C.: A distributed classification/estimation algorithm for sensor networks. SIAM J. Control Optim. **52**(1), 189–218 (2014)
11. Frasca, P., Carli, R., Fagnani, F., Zampieri, S.: Average consensus on networks with quantized communication. Int. J. Robust Nonlinear Control **19**(16), 1787–1816 (2009)
12. Galton, F.: Vox populi. Nature **75**(1949), 450–451 (1907)
13. Garin, F., Schenato, L.: A survey on distributed estimation and control applications using linear consensus algorithms. In: Bemporad, A., Heemels, M., Johansson, M. (eds.) Networked Control Systems. Lecture Notes in Control and Information Sciences. Springer (2010)
14. Garin, F., Zampieri, S.: Mean square performance of consensus-based distributed estimation over regular geometric graphs. SIAM J. Control Optim. **50**(1), 306–333 (2012)

15. Golub, B., Jackson, M.O.: Naïve learning in social networks and the hegsel of crowds. American Econ. J. Microecon. **2**(1), 112–149 (2010)
16. Huang, M., Manton, J.H.: Coordination and consensus of networked agents with noisy measurements: stochastic algorithms and asymptotic behavior. SIAM J. Control Optim. **48**(1), 134–161 (2009)
17. Jadbabaie, A., Olshevsky, A.: On performance of consensus protocols subject to noise: role of hitting times and network structure. In: IEEE Conference on Decision and Control, Las Vegas, NV, pp. 179–184, December 2016
18. Kar, S., Moura, J.M.F.: Distributed consensus algorithms in sensor networks: Quantized data and random link failures. IEEE Trans. Signal Process. **58**(3), 1383–1400 (2010)
19. Kashyap, A., Başar, T., Srikant, R.: Quantized consensus. Automatica **43**(7), 1192–1203 (2007)
20. Li, T., Fu, M., Xie, L., Zhang, J.-F.: Distributed consensus with limited communication data rate. IEEE Trans. Autom. Control **56**(2), 279–292 (2011)
21. Lovisari, E., Garin, F., Zampieri, S.: Resistance-based performance analysis of the consensus algorithm over geometric graphs. SIAM J. Control Optim. **51**(5), 3918–3945 (2013)
22. Nedic, A., Olshevsky, A., Ozdaglar, A., Tsitsiklis, J.N.: On distributed averaging algorithms and quantization effects. IEEE Trans. Autom. Control **54**(11), 2506–2517 (2009)
23. Olshevsky, A.: Consensus with ternary messages. SIAM J. Control Optim. **52**(2), 987–1009 (2014)
24. Rajagopal, R., Wainwright, M.J.: Network-based consensus averaging with general noisy channels. IEEE Trans. Signal Process. **59**(1), 373–385 (2011)
25. Xiao, L., Boyd, S., Kim, S.-J.: Distributed average consensus with least-mean-square deviation. J. Parallel Distrib. Comput. **67**(1), 33–46 (2007)