# Transient and limit performance of distributed relative localization

Wilbert Samuel Rossi Paolo Frasca Fabio Fagnani

***Abstract*— In this paper, we consider the problem of relative localization in a network of sensors, according to the formulation of Barooah and Hespanha. We introduce a distributed algorithm for its solution, and we study the algorithm performance by evaluating a suitable performance metric as a function of the network eigenvalues. Remarkably, the performance analysis indicates that it is preferable to stop the algorithm before convergence is reached: an estimate of the optimal stopping time is provided.**

## I. Introduction

We study in this paper the distributed solution of a problem of relative localization in a network of sensors. We assume to have a group of agents organized in a graph and a vector, indexed over the agents and unknown to them: the agents are allowed to take relative noisy measurements of their vector entries with respect to their neighbors in the graph. The estimation problem consists in reconstructing the original vector, up to an additive constant. We refer to this problem as the problem of *relative localization*. Our formulation of the problem essentially adopts that of Barooah and Hespanha [1], [2], [3], although these authors assume to have an anchor node, in order to avoid the uncertainty about the additive constant, and they resort to infinite graphs as an analysis and modeling tool.

The main contribution of our paper is an analysis of the transient performance of a gradient algorithm solving the problem of relative localization. We define a time-dependent performance metric, which is then computed as a function of the eigenvalues of the graph encoding the problem. Remarkably, the performance analysis indicates that the expected performance is optimal at a certain finite time, instead than asymptotically: consequently we argue that the algorithm should not be run until convergence, but stopped at a suitable time. We provide an estimate for the optimal stopping time: our estimate does not depend on the measurement graph and is thus of immediate and general application.

### *Notation*

In the paper, we will use the following notation. Vectors will be denoted with boldface letters, and matrices with capital letters. By the symbols $\mathbf{1}$ and $\mathbf{0}$ we denote vectors having all entries equal to 1 and 0, respectively. Given a matrix $M$, we denote by $\mathrm{tr}(M)$ its trace, by $M^T$ its transpose and by $M^\dagger$ its Moore-Penrose pseudoinverse [4].

The authors are with the Dipartimento di Scienze Matematiche (DISMA), Politecnico di Torino, corso Duca degli Abruzzi 24, 10129 Torino, Italy. E-mail contacts: wilbertsamuel.rossi@polito.it, paolo.frasca@polito.it, fabio.fagnani@polito.it.
Their work has been partly supported by MIUR under grant PRIN-20087W5P2K. The authors wish to thank an anonymous reviewer for his remarks and Prof. Sandro Zampieri for introducing them to the problem and for several fruitful discussions about it.

## II. Problem Formulation and Basic Results

We consider a set of $N$ agents, and we endow each of them with a scalar quantity $\bar{x}_i \in \mathbb{R}$, for $i \in \{0, \dots, N-1\}$. The $i$th agent does not know the value $\bar{x}_i$, but has an estimate $x_i \in \mathbb{R}$. We shall denote by $\bar{\mathbf{x}}$ and $\mathbf{x}$ the $N$-dimensional vectors whose components are $\bar{x}_i$ and $x_i$, respectively. We suppose that each agent $i$ can take relative measurements $\bar{x}_i - \bar{x}_j$ with respect to some neighbors $j$. An undirected graph $\mathbb{G} = (\{0, \dots, N-1\}, E)$ is used to represent the available measurements. The set of vertices is constituted by the $N$ agents, and the edges (pairs of agents) in $E$ correspond to the available measurements. We assume that there are $M$ available measurements, and that measurements are symmetrical, meaning that both agents of a pair know the measurement, with a reversed sign. Furthermore, we assume that the graph $\mathbb{G}$ is connected. On each edge, we choose an orientation, that is, we define a starting node and an ending node, in order to encode the measurements by using the incidence matrix $A \in \mathbb{R}^{M \times N}$ defined as follows

$$(A)_{e,i} = \begin{cases} 1 & \text{if } i \text{ is the terminating edge of } e \\ -1 & \text{if } i \text{ is the starting edge of } e \\ 0 & \text{otherwise.} \end{cases}$$

Measurements are affected by errors, which can be modeled by independent and identically distributed noises. Let $\mathbf{b} \in \mathbb{R}^M$ be the vector of the measurements and $\mathbf{n} \in \mathbb{R}^M$ that of noises. Then, in matrix notation we have

$$\mathbf{b} = A\bar{\mathbf{x}} + \mathbf{n}$$

with $\mathbb{E}[\mathbf{n}] = \mathbf{0}$ and $\mathbb{E}\left[\mathbf{n}\mathbf{n}^T\right] = \sigma^2 I$ where $I \in \mathbb{R}^{M \times M}$ is the identity matrix. It is also useful to define the Laplacian of $\mathbb{G}$ as $L = A^T A$. The Laplacian $L$ is a symmetric matrix, and being $\mathbb{G}$ connected, $L$ has eigenvalues $\lambda_0 = 0$ and $0 < \lambda_i \leq 2d_{\max}$ for $i \in \{1, \dots, N-1\}$, with $d_{\max}$ denoting the maximum degree of the nodes.

### *A. Least Squares Performance*

We define the best estimate in a least squares sense by defining the quadratic error functional

$$\Psi(\mathbf{x}) = \|A\mathbf{x} - \mathbf{b}\|_2^2,$$

where $A\mathbf{x}$ are the measurements corresponding to a generic configuration $\mathbf{x}$. The set of minimizers $X = \arg\min_{\mathbf{x} \in \mathbb{R}^N} \Psi(\mathbf{x})$ can be computed by equating the gradient of $\Psi(\mathbf{x})$ to zero. The configurations in $X$ minimize the quadratic error with respect to the available measurements.

Since $A$ has rank $N-1$ and $A\mathbf{1}=\mathbf{0}$, the set $X$ is an affine space parallel to $\mathbf{1}$. This remark is consistent with the fact that from relative measurements the absolute configuration can only be inferred up to an additive constant. Then, among all the optimal configurations, we may define $\hat{\mathbf{x}} \in X$ as the one which satisfies $\frac{1}{N}\mathbf{1}^T\hat{\mathbf{x}} = \frac{1}{N}\mathbf{1}^T\bar{\mathbf{x}}$.

## III. Gradient Algorithm and Transient Performance

### A. Algorithm Definition and Convergence Properties

The gradient of the functional $\Psi(\mathbf{x})$ can be used to build an iterative distributed estimation algorithm, which is consistent with the requirements of local communication and computation. Since the Hessian $H\Psi(\mathbf{x})$ is positive-semidefinite, a gradient descent algorithm can be used. Let the nonnegative integer $t$ be the iteration, $\mathbf{x}[t] \in \mathbb{R}^N$ be the collection of node estimates at the iteration $t$, and $\mathbf{x}[0]$ be the initial condition. Then, we propose the following algorithm

$$\mathbf{x}[t+1] = \mathbf{x}[t] - \tau\nabla\Psi(\mathbf{x}[t]),$$

where the parameter $\tau$ is to be determined for convergence reasons. The recursive law can be rewritten as:

$$\begin{aligned}\mathbf{x}[t+1] &= \mathbf{x}[t] - \tau(A^TA\mathbf{x}[t] - A^T\mathbf{b}) \\ &= (I - \tau A^TA)\mathbf{x}[t] + \tau A^T\mathbf{b}\end{aligned}$$

or as

$$\mathbf{x}[t+1] = P\,\mathbf{x}[t] + \mathbf{y}, \tag{1}$$

provided $P = I - \tau A^TA$ and $\mathbf{y} = \tau A^T\mathbf{b}$.

It is of note that the matrix $P$ is inherently adapted to the measurement graph $\mathbb{G}$, in the sense that $P_{ij} > 0$ only if $(i,j)$ is an edge in $\mathbb{G}$. This observation is key as it implies that the algorithm is naturally distributed over the graph which describes the problem: there is *no need for communication between agents which do not share a measurement.*

The convergence properties of the algorithm are summarized in the following result.

*Proposition 1 (Convergence):* The algorithm (1) is such that

$$\lim_{t\to+\infty}\mathbf{x}[t] = \mathbf{x}$$

with $\mathbf{x} \in X$ and $\frac{1}{N}\mathbf{1}^T\mathbf{x} = \frac{1}{N}\mathbf{1}^T\mathbf{x}[0]$, provided $0 < \tau < \frac{1}{d_{\max}}$, where $d_{\max}$ denotes the largest degree in $\mathbb{G}$.

*Proof:* From the assumption on $\tau$ it follows that $P$ is a symmetric, irreducible, aperiodic stochastic matrix. It is a standard fact that 1 is a simple eigenvalue whose eigenspace is spanned by $\mathbf{1}$, while all other eigenvalues are, in modulus, strictly less than 1. Since $\mathbf{1}^T\mathbf{y} = 0$, it then easily follows that $\mathbf{x}[t]$ converges to a solution of the equation $\mathbf{x} = P\mathbf{x} + \mathbf{y}$. Finally, invariance of the barycenter simply follows by applying $\mathbf{1}^T$ to both sides of (1). ∎

We observe that, given an initial condition $\mathbf{x}[0]$ the algorithm converges to an optimal estimate $\mathbf{x} \in X$, specifically the estimate with the same average as $\mathbf{x}[0]$. Then, in order to converge to the best estimate $\hat{\mathbf{x}}$, it would be necessary to impose the same average of $\bar{\mathbf{x}}$ to the initial condition $\mathbf{x}[0]$.

### B. Transient Performance

To evaluate the algorithm performance, we follow the approach in [5] and define the performance metric as the mean square error between the current estimate $\mathbf{x}[t]$ and the true configuration $\bar{\mathbf{x}}$, as follows:

$$J_t(A) := \frac{1}{N}\mathbb{E}\|\mathbf{x}[t] - \bar{\mathbf{x}}\|_2^2,$$

where the expectation is taken on the noise $\mathbf{n}$. Interestingly, this metric can be computed in terms of the eigenvalues of the matrix $P$.

*Proposition 2 (Transient performance):* Under the standing assumptions on the noise $\mathbf{n}$ and on the graph $\mathbb{G}$, for algorithm (1) it holds

$$J_t(A) = \frac{1}{N}d_0^2 + \frac{1}{N}\sum_{i=1}^{N-1}\mu_i^{2t}d_i^2 + \sigma^2\frac{\tau}{N}\sum_{i=1}^{N-1}\frac{(1-\mu_i^t)^2}{1-\mu_i},$$

where $\mu_i$'s are the eigenvalues of $P$, $M$ is the matrix whose column are eigenvectors of $P$, and $\mathbf{d} = M^T(\mathbf{x}[0] - \bar{\mathbf{x}})$.

*Proof:* In order to prove the statement it is necessary to rewrite $\mathbf{y}$ as

$$\mathbf{y} = \tau A^T\mathbf{b} = \tau A^TA\bar{x} + \tau A^T\mathbf{n} = (I-P)\bar{\mathbf{x}} + \tau A^T\mathbf{n}$$

and derive

$$\begin{aligned}\mathbf{x}[t] - \bar{\mathbf{x}} =& P^t\mathbf{x}[0] + \sum_{n=0}^{t-1}P^n\mathbf{y} - \bar{\mathbf{x}} \\ =& P^t(\mathbf{x}[0] - \bar{\mathbf{x}}) + \sum_{n=0}^{t-1}P^n(\tau A^T\mathbf{n}).\end{aligned}$$

Plugging this formula into the definition of $J_t(A)$ we get

$$\begin{aligned}J_t(A) &= \frac{1}{N}\mathbb{E}\|\mathbf{x}[t] - \bar{\mathbf{x}}\|^2 \\ &= \frac{1}{N}\mathbb{E}\big[\,\mathrm{tr}\,\big[(\mathbf{x}[t] - \bar{\mathbf{x}})(\mathbf{x}[t] - \bar{\mathbf{x}})^T\big]\,\big] \\ &= \frac{1}{N}\,\mathrm{tr}\,\big[P^t(\mathbf{x}[0] - \bar{\mathbf{x}})(\mathbf{x}[0] - \bar{\mathbf{x}})^TP^t \\ &\qquad + \sigma^2\tau^2\sum_{n=0}^{t-1}\sum_{m=0}^{t-1}P^nA^TAP^m\big] \\ &= \frac{1}{N}\,\mathrm{tr}\,\big[P^t(\mathbf{x}[0] - \bar{\mathbf{x}})(\mathbf{x}[0] - \bar{\mathbf{x}})^TP^t \\ &\qquad + \sigma^2\tau\sum_{n=0}^{t-1}\sum_{m=0}^{t-1}P^n(I-P)P^m\big].\end{aligned}$$

If we diagonalize $P$ using the orthonormal matrix $M$, and we define $\tilde{P} = M^TPM$ and $\mathbf{d} = M^T(\mathbf{x}[0] - \bar{\mathbf{x}})$, we deduce that

$$J_t(A) = \frac{1}{N}\,\mathrm{tr}\,\big[\tilde{P}^{2t}\mathbf{d}\mathbf{d}^T\big] + \sigma^2\frac{\tau}{N}\,\mathrm{tr}\,\big[\sum_{n=0}^{t-1}\tilde{P}^n(I-\tilde{P})\sum_{m=0}^{t-1}\tilde{P}^m\big].$$

Plugging the identity $(I-\tilde{P}) = (I-\tilde{P})(I-\tilde{P})^\dagger(I-\tilde{P})$ into the previous expression we get

$$J_t(A) = \frac{1}{N}\,\mathrm{tr}\,\big[\tilde{P}^{2t}\mathbf{d}\mathbf{d}^T\big] + \sigma^2\frac{\tau}{N}\,\mathrm{tr}\,\big[(I-\tilde{P})^\dagger(I-\tilde{P}^t)^2\big]. \tag{2}$$

Since the eigenvalues of $(I-\tilde{P})^\dagger$ are 0 and $\frac{1}{1-\mu_i}$ for $i \neq 0$, we conclude

$$J_t(A) = \frac{1}{N}\sum_{i=0}^{N-1} \mu_i^{2t} d_i^2 + \sigma^2 \frac{\tau}{N} \sum_{i=1}^{N-1} \frac{(1-\mu_i^t)^2}{1-\mu_i},$$

which gives the result. ∎

A few remarks are due about the formula for $J_t(A)$ found in Proposition (2). First of all, $J_t(A)$ depends on the eigenvalues of the Laplacian $L$ via the relation $\mu_i = 1-\tau\lambda_i$, which implies

$$J_t(A) = \frac{1}{N} d_0^2 + \frac{1}{N}\sum_{i=1}^{N-1} (1-\tau\lambda_i)^{2t} d_i^2 + \\ + \sigma^2 \frac{1}{N} \sum_{i=1}^{N-1} \frac{(1-(1-\tau\lambda_i)^t)^2}{\lambda_i}.$$

This formula is helpful to compute the limit of $J_t(A)$ as $t \to +\infty$. If the initial condition has same average as the true configuration, namely $\frac{1}{N}\mathbf{1}^T\mathbf{x}[0] = \frac{1}{N}\mathbf{1}^T\bar{\mathbf{x}}$, then $d_0 = 0$ and at convergence we have:

$$J_\infty(A) := \lim_{t\to\infty} J_t(A) = \sigma^2 \frac{1}{N} \sum_{i=1}^{N-1} \frac{1}{\lambda_i}.$$

Remarkably, $J_\infty(A)$ does not depend on the algorithm and is equal to the mean square error of the Best Linear Unbiased Estimator studied by Barooah and Hespanha, namely

$$J_\infty(A) = \frac{1}{N}\mathbb{E}\|\hat{\mathbf{x}} - \bar{\mathbf{x}}\|_2^2.$$

If instead $d_0 \neq 0$, then the corresponding term is constant in time, that is, the difference between the average of $\mathbf{x}[0]$ and $\bar{\mathbf{x}}$ has an impact on the estimate at every time step. This is consistent with intuition, because indeed agents may not reconstruct the average of $\bar{\mathbf{x}}$. We note that, in order to fix this degree of freedom, an anchor node can be used as in the work [1]: our analysis carries on to such case with minor adaptations.

It is also interesting to discuss the role of the other terms in the expression of $J_t(A)$. Indeed, the first summation is due to the difference between the initial condition and the true configuration. The impact of this difference monotonically converges to zero along the iterations, since $\mu_i$s have modulus less than 1. The second sum, instead, encodes the diffusion of the noise in the estimates: its contribution converges to a constant.

### C. Finite-time Optimality

As $J_t(A)$ contains both decreasing and increasing terms, it is reasonable to ask whether $J_t(A)$ reaches a minimum at a finite time. The existence of such a minimum would be of significant applicative interest, since it implies an optimal stopping criterion for the algorithm. Intuitively, we expect a trade-off between increasing and decreasing terms controlled by the relative magnitude of the initial condition and the noise.

The analysis of $J_t(A)$ –see formula (2)– is complicated by the presence of the coefficients $d_i^2$, which depend on the initial condition. In the following we will make a simplification and average over the initial conditions assuming the initial condition $\mathbf{x}[0] - \bar{\mathbf{x}}$ to be a vector of i.i.d. random variables with zero mean and variance $\nu^2$. Since $\mathbf{d} = M^T(\mathbf{x}[0]-\bar{\mathbf{x}})$, it easy to show that also $\mathbf{d}$ is a vector of i.i.d. random variable with the same properties so that $\mathbb{E}[d_i^2] = \nu^2$ for every $i$ including 0. So we can consider the expected behavior of $J_t(A)$ by averaging on the initial condition:

$$J_t^{\mathbb{E}}(A) = \mathbb{E}[J_t(A)] = \frac{\nu^2}{N} + \frac{\nu^2}{N}\sum_{i=1}^{N-1} \mu_i^{2t} + \frac{\tau\sigma^2}{N}\sum_{i=1}^{N-1} \frac{(1-\mu_i^t)^2}{1-\mu_i}.$$

Note that $J_{t=0}^{\mathbb{E}}(A) = \nu^2$ and that we have

$$\lim_{t\to+\infty} J_t^{\mathbb{E}}(A) = J_\infty(A) + \frac{\nu^2}{N}.$$

This formula implies that the component of the error which is due to the incorrect estimate of the average goes to zero as $N$ grows.

In order to investigate our insights about the time dependence of $J_t(A)$ we shall make use of the rewriting

$$\begin{aligned} &J_t^{\mathbb{E}}(A) = \\ &= \tau\sigma^2 \int_{-1}^{1} \mathrm{d}\mu \left[\alpha\mu^{2t} + \frac{(1-\mu^t)^2}{1-\mu}\right] \cdot \left[\frac{1}{N}\sum_{i=0}^{N-1} \delta(\mu-\mu_i)\right] \\ &= \tau\sigma^2 \int_{-1}^{1} \mathrm{d}\mu \quad f_{t,\alpha}(\mu) \qquad \cdot \qquad g_N(\mu)\ , \end{aligned} \tag{3}$$

where $\delta(\mu-\mu_i)$ is the Dirac delta distribution and $\alpha = \dfrac{\nu^2}{\tau\sigma^2}$ is the *initial-condition-to-noise ratio*. The integrand in (3) is the product of two essentially independent terms. The first term is a continuous weight function $f_{t,\alpha}(\mu)$, which depends on the time step $t$ and on $\alpha$. The second term is the distribution of the eigenvalues $g_N(\mu)$. Indeed, the integral runs over $[-1,1]$, which contains the support of the eigenvalue distribution. Observe that $f_{t,\alpha}(\mu)$ and $g_N(\mu)$ respectively characterize the estimation problem and the graph. The algorithm parameter $\tau$, instead, influences both terms, as both $\alpha$ and $\mu_i$ depend on it.

By exploiting the above decomposition we can prove the following theorem, showing that $J_t^{\mathbb{E}}(A)$ is eventually increasing.

*Theorem 3 (Finite-time optimality):* For any graph topology and any size $N$, and provided $0 < \tau < \frac{1}{2d_{\max}}$, the index $J_t^{\mathbb{E}}(A)$ is increasing for all $t > \alpha$.

*Proof:* In order to show that the sequence $J_t(A)$ is eventually increasing, we study the discrete derivative $J_{t+1}^{\mathbb{E}}(A) - J_t^{\mathbb{E}}(A)$ and show that it is eventually positive. The discrete derivative reads

$$J_{t+1}^{\mathbb{E}}(A) - J_t^{\mathbb{E}}(A) = \tau\sigma^2 \int_{-1}^{1} \mathrm{d}\mu [f_{t+1,\alpha}(\mu) - f_{t,\alpha}(\mu)] \cdot g_N(\mu),$$

which is positive when the integrand is positive. Since $\mu_i = 1-\tau\lambda_i$ and $\lambda \in [0, 2d_{\max}]$, the assumption on $\tau$ implies that

$g_N(\mu) \geq 0$ in $[0,1]$. Next, if we compute

$$f_{t+1,\alpha}(\mu) - f_{t,\alpha}(\mu) = \alpha\mu^{2t+2} - \alpha\mu^{2t} + 2\mu^t - \mu^{2t} - \mu^{2t+1},$$

it is immediate to check that $f_{t+1,\alpha}(0) - f_{t,\alpha}(0) = 0$ and $f_{t+1,\alpha}(1) - f_{t,\alpha}(1) = 0$. Moreover, we claim that when $t \geq \alpha - \frac{1}{2}$, we have $f_{t+1,\alpha}(\mu) - f_{t,\alpha}(\mu) > 0$ for all $\mu \in (0,1)$. In order to prove this claim, we compute

$$f_{t+1,\alpha}(\mu) - f_{t,\alpha}(\mu) = \mu^{2t}\left(\frac{2}{\mu^t} - \mu - 1 - (\alpha - \alpha\mu^2)\right)$$

and we define $a(\mu) = \alpha - \alpha\mu^2$ and $b(\mu) = \frac{2}{\mu^t} - \mu - 1$ as functions on $(0,1]$. The following facts are then easy to verify:

$$\begin{aligned} a(1) &= 0 & b(1) &= 0 \\ a'(\mu) &= -2\alpha\mu < 0 & b'(\mu) &= -\frac{2t}{\mu^{t+1}} - 1 < 0 \\ a'(1) &= -2\alpha < 0 & b'(1) &= -2t - 1 < 0 \\ a''(\mu) &= -2\alpha < 0 & b''(\mu) &= \frac{2t(t+1)}{\mu^{t+2}} > 0. \end{aligned}$$

These observations imply that the functions $a$ and $b$ have a common zero and opposite convexity, as $a$ is convex and $b$ is concave. Moreover, since $\mu \leq 1$, it holds true that when $-2t - 1 < -2\alpha$, *i.e.*, when $t > \alpha - \frac{1}{2}$, then $b'(1) < a'(1)$. This fact implies that $a(\mu) < b(\mu)$ in $(0,1)$ thanks to the above discussion. We have then proved that if $t \geq \alpha - \frac{1}{2}$, then $f_{t+1,\alpha}(\mu) - f_{t,\alpha}(\mu) > 0$, for all $\mu \in (0,1)$. Overall, we have shown that both factors of the integrand are positive, and we can conclude that $J^{\mathbb{E}}_{t+1}(A) - J^{\mathbb{E}}_t(A) \geq 0$ when $t \geq \alpha - \frac{1}{2}$. So, clearly, the index $J^{\mathbb{E}}_t(A)$ is increasing for all $t \geq \alpha - \frac{1}{2}$. ■

Clearly, the theorem implies that $J^{\mathbb{E}}_t(A)$ has (at least) a minimum, and that the time step which achieves the minimum is not larger than $\alpha$: this result holds regardless of the topology and size of the network.

*Remark 4 (On the assumption on $\tau$):* The assumption on $\tau$ in Theorem 3 restricts the possible choice of $\tau$ to $(0, \frac{1}{2d_{\max}})$. The assumption has been made to simplify the proof and is slightly conservative: the result may actually be extended to every $\tau \in (0, \frac{1}{d_{\max}})$ by a similar argument. The argument is however more involved because the sign of the difference $f_{t+1,\alpha}(\mu) - f_{t,\alpha}(\mu)$ on $(-1,0)$ depends on the parity of $t$ (e.g. notice that $[f_{t+1,\alpha}(-1) - f_{t,\alpha}(-1)] = \pm 2$). Hence the reason for the simplifying assumption on $\tau$.

### D. Simulations and Conjectures

In this section we present simulations and numerical results, which validate our theoretical findings and confirm their practical interest. Our presentation below assumes the graph $\mathbb{G}$ to have a *ring* topology. Rings are defined as graphs having node set $\{0, \ldots, N-1\}$ and vertex set $E$ such that $(i,j) \in E$ if $|i - j| = 1 \mod N$. This simple topology has been chosen in order to provide useful examples: similar remarks can be obtained on other families of graphs.

Let us then investigate the dependence of $J^{\mathbb{E}}_t(A)$ on $\alpha = \frac{\nu^2}{\tau\sigma^2}$ and on the size $N$. In Figure 1 we fix $\sigma = 1$, $\tau = \frac{1}{4}$, while we let $N$ and $\nu$ vary. The plot clearly shows that for high initial-condition-to-noise ratio it is convenient to use the algorithm and the minimum, always present, is not very pronounced. As $\alpha$ decreases, the minimum becomes prominent, because the noise on the measurements is comparable with the true information contained in the initial condition (hereafter i.c.). The latter is thus disguised by the noise. For very low value of $\alpha$, the noise covers the i.c. so that running the algorithm is practically useless. Clearly, for what concerns i.c. and noise there is a sort of *compromise* regulated by $\alpha$ (namely between $\nu$ and $\sigma$), that shifts the location of the minimum in time and justifies or not the use of the algorithm.

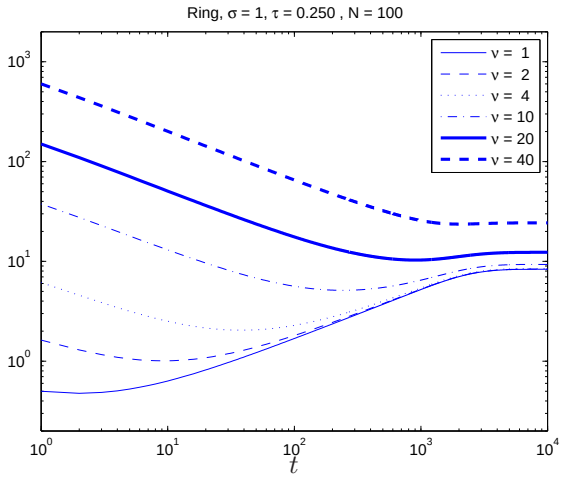


Fig. 1. The performance metric $J^{\mathbb{E}}_t(A)$ as a function of time. Plot assumes $\alpha$ varying and $N$ fixed.

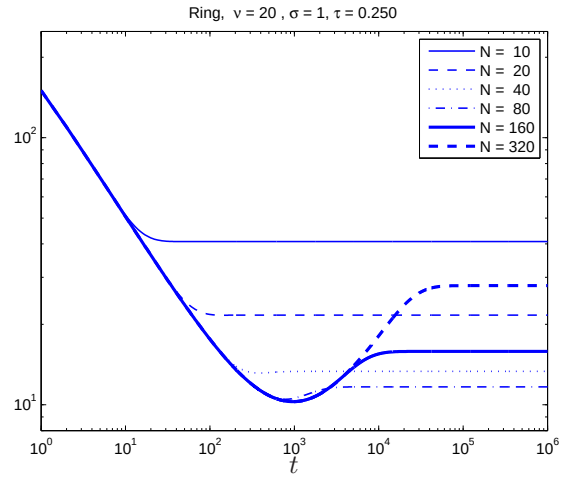


Fig. 2. The performance metric $J^{\mathbb{E}}_t(A)$ as a function of time. Plot assumes $\alpha$ fixed and $N$ varying.

Instead, in Figure 2 we fix $\alpha$ and let the size of the ring $N$ grow. In this plot, we notice that the curves converge, as $N$ grows and for every $t$, to a limit curve: we actually conjecture that this convergence is uniform on compact intervals. Interestingly, the optimal time $t_{\min}$ seems to converge to a finite

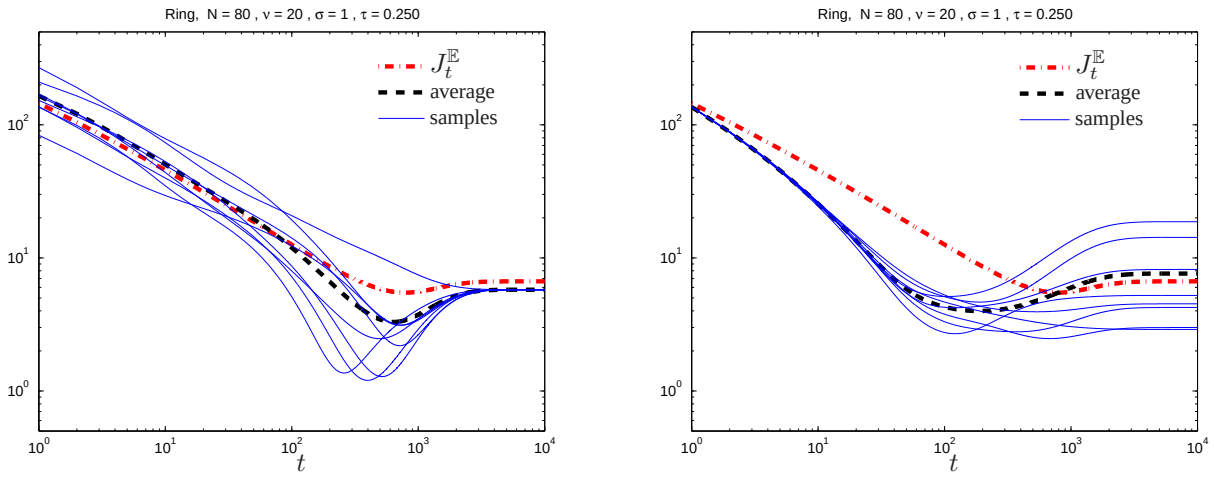


Fig. 3. Comparisons between the expected index $J_t^{\mathbb{E}}(A)$, a few simulated evolutions of $\frac{1}{N}\|\mathbf{x}[t] - \bar{\mathbf{x}}\|_2^2$, and the resulting average evolution. Left plot: same measurements $\mathbf{b}$, different initial conditions $\mathbf{x}[0]$. Right plot: same initial condition, different noise on the measurements.

value as $N$ grows. We sketch a possible approach to proving this convergence: detailed proofs are deferred to future work. In the limit of $N \to \infty$, the decomposition (3) is suitable for an interpretation with continuous eigenvalues: given a sequence of graphs, we postulate that the $g_N(\mu)$ converges to a continuous function $g(\mu)$. In rings, with $\tau = \frac{1}{3}$ we have:

$$g(\mu) = \frac{2}{\frac{4\pi}{3}\sqrt{1 - \frac{9}{4}\left(\mu - \frac{1}{3}\right)^2}}.$$

Suppose that $\alpha \gg 1$, so that the optimal value is not too small. Then $f_{t+1,\alpha}(\mu) - f_{t,\alpha}(\mu)$ is only positive for $\mu$ close to 1, and we just need to estimate the principal part of $g(\mu)$ in that neighborhood. Under these assumptions and with few additional approximations, the integral equation $J_{t+1}^{\mathbb{E}}(A) - J_t^{\mathbb{E}}(A) = 0$ can be solved analytically, leading us the following conjecture: on a family of ring graphs and for $\tau = \frac{1}{3}$, it holds

$$\lim_{N\to\infty} t_{\min} = \frac{1}{4(\sqrt{2}-1)}\alpha.$$

Finally, we want to stress that our analysis has focused on the mean index $J_t^{\mathbb{E}}(A)$, where the expectation is taken both over the noise and over the initial conditions. As a consequence, if we are interested in the behavior of the error $\|\mathbf{x}[t] - \bar{\mathbf{x}}\|_2$ for a given initial condition and measurement noise realization, Theorem 3 just suggests that the error reaches a minimum at a finite time for *some* initial conditions and measurements. This phenomenon is confirmed by simulations of algorithm (1): sample evolutions are plotted in Figure 3.

## IV. Open Problems

The analysis presented here opens up a wide range of research lines. On one hand, several questions are left unanswered in the current setting, including a finer graph-dependent analysis of the optimal stopping time, and a more extended study of the statistical distribution of $J_t$. On the other hand, the problem is suitable of useful extensions. The algorithm which has been proposed is synchronous, in the sense that all agents simultaneously update their estimates. A natural research line would thus look for asynchronous and gossip implementations of the gradient descent algorithm. Moreover, in this work the quantity of interest was assumed to be a time-invariant vector: a natural extension involves distributed filtering of time-dependent signals.

## References


[1] P. Barooah and J. P. Hespanha. Estimation from relative measurements: Algorithms and scaling laws. *IEEE Control Systems Magazine*, 27(4):57–74, 2007.

[2] P. Barooah and J. P. Hespanha. Estimation from relative measurements: Electrical analogy and large graphs. *IEEE Transactions on Signal Processing*, 56(6):2181–2193, 2008.

[3] P. Barooah and J. P. Hespanha. Error scaling laws for linear optimal estimation from relative measurements. *IEEE Transactions on Information Theory*, 55(12):5661–5673, 2009.

[4] A. Ben-Israel and T. N. E. Grevilee, *Generalized Inverses*, 2 ed., New York, Springer-Verlag, 2003

[5] F. Garin and S. Zampieri. Mean square performance of consensus-based distributed estimation over regular geometric graphs. *SIAM Journal on Control and Optimization*, 50(1):306–333, 2012.