Distributed Estimation from Relative and Absolute Measurements

Distributed Estimation from Relative and Absolute Measurements

Wilbert Samuel Rossi, Paolo Frasca, and Fabio Fagnani

***Abstract*—This note defines the problem of least-squares distributed estimation from relative and absolute measurements, by encoding the set of measurements in a weighted undirected graph. The role of its topology is studied by an electrical interpretation, which easily allows distinguishing between topologies that lead to "small" or "large" estimation errors. The least-squares problem is solved by a distributed gradient algorithm: the computed solution is approximately optimal after a number of steps that does not depend on the size of the problem or on the graph-theoretic properties of its encoding. This fact indicates that only a limited cooperation between the sensors is necessary.**

***Index Terms*—Sensor Networks; Distributed Estimation; Optimization algorithms; Cooperative control.**

I. Introduction

Multi-agent and network-based problems often involve measurements and estimation of unknown quantities. In this work, we are interested in problems where a state is distributed over the agents–that is, the state is a vector and one component is assigned to each agent–and each agent has access to noisy measurements of its state and of pairwise differences between its own state and the states of some other agents.

We briefly describe two motivating applications. The first one is self-localization in mobile robotic networks [1], *e.g.*, autonomous teams of road vehicles [2]. Here the agents can be equipped with GPS location sensors as well as with radars to sense the relative positions of their neighbors. These two different kinds of measurements can be combined into an improved estimate of the vehicle location. The second one is statistical ranking in databases, like in the "Netflix problem" [3]. Here items (*e.g.*, movies) have to be ranked according to their "quality", which can be assessed by users in either absolute or comparative way. Other applications include, for instance, clock synchronization [4].

In this note, we define the problem of distributed estimation from relative and absolute measurements and we encode the set of measurements by using a weighted graph. We observe that finding the optimal estimate in least-squares sense is equivalent to solving a network of resistors [5]. This classical electrical interpretation highlights the role of the topology of the measurement graph: for instance, we show that on complete graphs the estimation error decreases to zero as the number of nodes grows, whereas on cycle graphs the estimation error is bounded away from zero.

In previous works on estimation from relative measurements [1], [6], the nodes have access to relative measurements only, possibly with the exception of one anchor node that serves as reference (and can thus be seen as having perfect absolute information). In this literature, the available relative measurements are described by a weighted graph. It is known that the dimension of the graph is a crucial parameter [7] in distinguishing whether the optimal estimator scales well with an increasing number of nodes [8], [9]. In fact, the mean square error is determined by the effective resistance of the graph [1], [10]–[12]: this interpretation suggests intuitive criteria to optimize the acquisition of data [13]. Our work extends this graph-based description to include absolute measurements, so that every node can access the reference value, albeit corrupted by noise. The reference also plays the role of *a priori* regularization for the relative estimation problem, ruling out certain pathological behaviors observed in [14]. From a mathematical perspective, our problem can be rewritten into the classical setup by adding a virtual reference node: we take advantage of this transformation in our analysis.

We propose to solve our estimation problem by a distributed gradient descent algorithm. This algorithm has a remarkable feature: it approximates the optimal solution up to a given tolerance within a number of iterates that does not depend on the number of nodes or on the measurement graph topology. This feature contrasts with other network estimation algorithms, such as averaging algorithms that compute a common parameter from distributed measurements: in the latter, obtaining a given precision of the estimate can require a number of iterates that grows with the number of nodes [15]. Even if the least-squares problem can be solved by a simple gradient algorithm, the literature contains a variety of methods tailored to the distributed solution of the "pure" relative estimation problem, including Jacobi methods [7], Kaczmarz iterates [16], randomized *gossiping* techniques [17], and asynchronous algorithms exploiting memory [18]. The extension of these methods to the case with absolute measurements can be a topic of future work.

Preliminary versions of some of our results appeared in [19]. In comparison with this conference paper, in the present note we make more general assumptions, which allow for heterogeneous measurements, and we discuss the electrical interpretation of the estimation problem, establishing tight

W. S. Rossi is with the Department of Applied Mathematics, University of Twente, 7500 AE Enschede, The Netherlands (e-mail: w.s.rossi@utwente.nl).

Paolo Frasca is with Univ. Grenoble Alpes, CNRS, Inria, GIPSA-lab, F-38000 Grenoble, France (e-mail: paolo.frasca@gipsa-lab.fr).

F. Fagnani is with the Dipartimento di Scienze Matematiche, Politecnico di Torino, corso Duca degli Abruzzi 24, 10129 Torino, Italy (e-mail fabio.fagnani@polito.it).

bounds on the error of the optimal estimate.

Paper organization: The formal statement of the problem and its least-squares solution are given in Section II. Section III describes the electrical interpretation and Section IV its application to highlight the role of the measurement graph topology. Next, the gradient algorithm and its mean square error are studied in Section V. Some conclusions are drawn in the final section.

Notation: Vectors are 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 A, we denote by $\mathrm{tr}(A)$ its trace, by $A^\top$ its transpose, and by A^{-1} its inverse. Given a vector $\mathbf{v} \in \mathbb{R}^N$ and a positive diagonal matrix Q, $\|\mathbf{v}\|_Q^2$ stands for the weighted norm $\mathbf{v}^\top Q \mathbf{v}$ and $\mathrm{diag}(v_1, \dots, v_N)$ is a diagonal matrix in $\mathbb{R}^{N\times N}$ with the the entries of the vector in the main diagonal. Given a set S, we denote its cardinality by $|S|$. Given two positive number a and b, we use $\oplus$ to denote their harmonic sum $a \oplus b = \left(a^{-1} + b^{-1}\right)^{-1}$.

II. Relative and absolute measurements

We consider a set of N agents and we endow each of them with a scalar quantity $\bar{x}_i \in \mathbb{R}$, for $i \in \{1, \dots, N\}$. The ith agent does not know the value $\bar{x}_i$ and wants to estimate it. We shall denote by $\bar{\mathbf{x}}$ the N-dimensional vector whose components are $\bar{x}_i$.

We assume that each agent i can perform a noisy absolute measurement of $\bar{x}_i$, denoted by $x_i^0 = \bar{x}_i + a_i$, where the a_is are independent real-valued random variables with $\mathbb{E}[a_i] = 0$ and $\mathbb{E}\left[a_i^2\right] = \nu_i^2 > 0$. The absolute measurements and the corresponding noises can be stacked together in the vectors $\mathbf{x}^0$, $\mathbf{a} \in \mathbb{R}^N$ to get

$$\mathbf{x}^0 = \bar{\mathbf{x}} + \mathbf{a}$$

with $\mathbb{E}[\mathbf{a}] = \mathbf{0}$. The covariance matrix of $\mathbf{a}$ is the positive definite diagonal matrix $\mathcal{N} \in \mathbb{R}^{N\times N}$ with $\mathcal{N} = \mathbb{E}\left[\mathbf{a}\mathbf{a}^\top\right] = \mathrm{diag}(\nu_1^2, \dots, \nu_N^2)$.

We also assume that each agent i can take relative cooperative measurements of the quantity $\bar{x}_j - \bar{x}_i$ with respect to some neighbors j. An undirected graph $\mathbb{G} = (\{1, \dots, N\}, E)$ is used to represent the available relative measurements. The set of vertices is constituted by the N agents, and the edges (*unordered* pairs of agents like $\{i, j\}$) in E correspond to the available measurements. The set N_i of neighbors of the node i, *i.e.,* $N_i = \{j \in \{1, \dots, N\} : \{i, j\} \in E\}$, contains the nodes with whom i took a relative measurement. We let $d_i = |N_i|$ and $d_{\max} = \max_i d_i$. In fact, we assume that both agents of a pair $\{i, j\} \in E$ know the measurement, even if we assign–without loss of generality–an orientation to each edge in the graph $\mathbb{G}$: an edge $\{i, j\}$ with $i < j$ is oriented *from i to j* and the corresponding measurement $b_{\{i,j\}}$ regards the quantity $\bar{x}_j - \bar{x}_i$. The quantities $b_{\{i,j\}}$ and $b_{\{j,i\}}$ coincide by definition, as $\{i, j\}$ and $\{j, i\}$ denote the same edge. Measurements are corrupted by errors that we model with independent random variables $n_{\{i,j\}}$ such that $b_{\{i,j\}} = \bar{x}_j - \bar{x}_i + n_{\{i,j\}}$, $\mathbb{E}\left[n_{\{i,j\}}\right] = 0$ and $\mathbb{E}\left[n_{\{i,j\}}^2\right] = \sigma_{\{i,j\}}^2 > 0$.

In order to encode the measurements in a vector, we define the incidence matrix $A \in \mathbb{R}^{|E|\times N}$ as follows

$$A_{e,i} = \begin{cases} 1 & \text{if } e = \{i, j\} \text{ and } i > j \\ -1 & \text{if } e = \{i, j\} \text{ and } i < j \\ 0 & \text{otherwise.} \end{cases}$$

Let $\mathbf{b} \in \mathbb{R}^{|E|}$ be the vector of the measurements and $\mathbf{n} \in \mathbb{R}^{|E|}$ 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 $\mathbb{E}\left[\mathbf{n}\mathbf{n}^\top\right] = \mathrm{diag}(\sigma_1^2, \dots, \sigma_{|E|}^2) = \Sigma$, where $\Sigma \in \mathbb{R}^{|E|\times|E|}$ is the positive definite diagonal covariance matrix. All components of $\mathbf{a}$ and of $\mathbf{n}$ are assumed to be independent of each other.

A natural way to define the best estimate, given the absolute and relative measurements, is solving the least squares problem

$$\min_{\mathbf{x}} \Phi(\mathbf{x}) \quad \text{with} \quad \Phi(\mathbf{x}) = \|A\mathbf{x} - \mathbf{b}\|_{\Sigma^{-1}}^2 + \|\mathbf{x} - \mathbf{x}^0\|_{\mathcal{N}^{-1}}^2.$$

The functional $\Phi(\mathbf{x})$ sums two terms that represent the estimation errors with respect to the relative and absolute measurements respectively, weighted according to their significance, *i.e.*, the inverse of their covariance matrices. The quadratic convex functional $\Phi(\mathbf{x})$ has gradient

$$\nabla\Phi(\mathbf{x}) = 2M\mathbf{x} - 2\left(A^\top\Sigma^{-1}\mathbf{b} + \mathcal{N}^{-1}\mathbf{x}^0\right), \tag{1}$$

where $M := A^\top\Sigma^{-1}A + \mathcal{N}^{-1}$ and the optimal solution $\mathbf{x}^* = \arg\min_{\mathbf{x}} \Phi(\mathbf{x})$ is the solution of the linear system

$$M\,\mathbf{x}^* = \left(A^\top\Sigma^{-1}\mathbf{b} + \mathcal{N}^{-1}\mathbf{x}^0\right). \tag{2}$$

The matrix M is invertible and positive definite, hence

$$\mathbf{x}^* = M^{-1}\left(A^\top\Sigma^{-1}\mathbf{b} + \mathcal{N}^{-1}\mathbf{x}^0\right), \tag{3}$$

which by [20, p. 66], can be rewritten as

$$\mathbf{x}^* = \mathbf{x}^0 + \mathcal{N}A^\top\left(A\mathcal{N}A^\top + \Sigma\right)^{-1}\left(\mathbf{b} - A\mathbf{x}^0\right).$$

This formula highlights that the optimal solution and the absolute measurements have the same weighted average $\frac{1}{N}\mathbf{1}^\top\mathcal{N}^{-1}\mathbf{x}^* = \frac{1}{N}\mathbf{1}^\top\mathcal{N}^{-1}\mathbf{x}_0$, because $A\mathbf{1} = \mathbf{0}$.

III. Electrical Networks

The estimation problem of Section II is intimately related to linear circuit theory and its solution has an intuitive electrical interpretation.

Given $\mathbb{G} = (\{1, \dots, N\}, E)$, we consider an *augmented* graph $\mathbb{H} = (\{0, \dots, N\}, E \cup F)$ of which $\mathbb{G}$ is an induced subgraph: the graph $\mathbb{H}$ contains an additional *reference* node 0 and has edge set $E \cup F$ where $F = \bigcup_{i\in\{1,\dots,N\}}\{0, i\}$ contains the edges between the reference node and every other node. Notice that $\mathbb{H}$ is connected by construction. We build an electrical network on the graph $\mathbb{H}$ by substituting each edge $\{i, j\} \in E \cup F$ with the series of an ideal voltage source and a resistor. The source has the positive terminal oriented toward the larger of i, j. On each edge $\{0, i\} \in F$, the voltage of the source is x_i^0 and the resistance is ν_i^2; on each edge $\{i, j\} \in E$ the source's voltage is $b_{\{i,j\}}$ and the resistance is $\sigma_{\{i,j\}}^2$. The matrix $M = A^\top\Sigma^{-1}A + \mathcal{N}^{-1}$ is the *reduced*

conductance matrix of the electrical network built over $\mathbb{H}$, obtained by *grounding* the reference node 0: indeed, M is also known as *grounded Laplacian* of the weighted graph. We refer the reader to [10] for a more complete introduction to these concepts.

Each node in $\mathbb{H}$ is endowed with a scalar quantity x_i to be interpreted as electrical potential, with the reference node set at conventional zero. Given an edge $\{i,j\}$ with $i<j$, the potential difference across that edge is measured as $x_j - x_i$ whereas $\iota_{\{i,j\}}$ denotes the current flowing in $\{i,j\}$ from j to i. The vector $\mathbf{x} \in \mathbb{R}^N$ collects the potentials of all the nodes except the reference for which $x_0 = 0$. Notice that, for $i \neq 0$, each x_i is itself the potential difference across the edge $\{0,i\}$.

With all the voltage sources of the network switched off (and substituted by short circuits), we define the *effective resistance* R_{ij}^{eff} between any two nodes i,j of the network as the potential difference $x_j - x_i$ obtained when an external source supplies a unit current to j and extracts it from node i.

The following Lemma provides the electrical interpretations of the quantities involved in the estimation problem and will be useful in the following sections.

Lemma 1 (Electrical interpretation)**.** *Consider the electrical network $\mathbb{H}$ and the estimation problem of Section II.*

- *The components of the vector $\mathbf{x}^*$ defined in* (3) *are the potentials of the nodes in the network.*
- *For any pair of nodes in the network (reference included) the variance of the optimal estimate of their relative location coincides with the effective resistance between the two nodes. Therefore,*

$$\mathbb{E}\left[(x_i^* - \bar{x}_i)^2\right] = (M^{-1})_{i,i} = R_{0i}^{eff}. \tag{4}$$

Proof. In order to systematically analyze the resistive electric circuit we shall express the current in each edge as function of the unknown node potentials, then apply the Kirchhoff's current law (KCL) to each non-reference node, and obtain the set of N independent *node equations*. The potential difference and current flowing in each edge of $\mathbb{H}$ are related via the Ohm laws. On edges like $\{0,i\} \in F$, where one node is the reference, we have $x_i - x_0 = x_i = x_i^0 + \nu_i^2\, \iota_{\{0,i\}}$ since $x_0 = 0$. For the remaining edges $\{i,j\} \in E$ we have $x_j - x_i = b_{\{i,j\}} + \sigma^2_{\{i,j\}}\, \iota_{\{i,j\}}$.

By KCL, the sum of the currents flowing *away* from a node needs to be zero. Consider a node $i \neq 0$, then,

$$\iota_{\{0,i\}} + \sum_{j\in N_i, j<i} \iota_{\{i,j\}} - \sum_{j\in N_i, j>i} \iota_{\{i,j\}} = 0$$

$$\frac{x_i - x_i^0}{\nu_i^2} + \sum_{j\in N_i, j<i} \frac{x_i - x_j - b_{\{i,j\}}}{\sigma^2_{\{i,j\}}} - \sum_{j\in N_i, j>i} \frac{x_j - x_i - b_{\{i,j\}}}{\sigma^2_{\{i,j\}}} = 0$$

$$\frac{x_i}{\nu_i^2} + \sum_{j\in N_i} \frac{x_i - x_j}{\sigma^2_{\{i,j\}}} = \frac{x_i^0}{\nu_i^2} + \sum_{j\in N_i} \frac{(\mathbb{1}_{j<i} - \mathbb{1}_{j>i})\, b_{\{i,j\}}}{\sigma^2_{\{i,j\}}}$$

In matrix form, the above equations coincide with (2).

With $\mathbf{b} = A\bar{\mathbf{x}} + \mathbf{n}$ and $\mathbf{x}^0 = \bar{\mathbf{x}} + \mathbf{a}$ in (3) the optimal estimate becomes $\mathbf{x}^* = \bar{\mathbf{x}} + M^{-1}\left(A^\top \Sigma^{-1}\mathbf{n} + \mathcal{N}^{-1}\mathbf{a}\right)$. Then $\mathbb{E}[\mathbf{x}^*] = \bar{\mathbf{x}}$ and $\mathbb{E}\left[(\mathbf{x}^* - \bar{\mathbf{x}})(\mathbf{x}^* - \bar{\mathbf{x}})^\top\right] = M^{-1}$. Let us introduce the ith unit vector $\mathbf{e}_i \in \mathbb{R}^n$, with 1 in the ith position and 0 elsewhere. The variance of x_i^* is $\mathbb{E}\left[(x_i^* - \bar{x}_i)^2\right] = \mathbf{e}_i^\top \mathbb{E}\left[(\mathbf{x}^* - \bar{\mathbf{x}})(\mathbf{x}^* - \bar{\mathbf{x}})^\top\right]\mathbf{e}_i = (M^{-1})_{i,i}$and, being $\mathbb{H}$ a connected graph with 0 as reference node, $(M^{-1})_{i,i} = R_{0i}^{eff}$ as proved in [10]. □

IV. The error of the optimal estimate

In order to evaluate the quality of the least-squares estimate, we define the estimation error as

$$H_\infty := \frac{1}{N}\mathbb{E}\left[\|\mathbf{x}^* - \bar{\mathbf{x}}\|_2^2\right],$$

where the expectation is taken over all measurement noise. This quantity depends only on the topology of the measurement network and on the noise variances. Indeed, by (4) we have

$$H_\infty = \frac{1}{N}\operatorname{tr}\left(M^{-1}\right) = \frac{1}{N}\sum_{i=1}^{N} R_{0i}^{eff}. \tag{5}$$

Thus, the error of the optimal estimator can be effectively computed by the electrical analogy. The electrical analogy also permits to derive the following estimates.

Proposition 2 (Error bounds)**.** *Consider the graph $\mathbb{G} = (\{1,\dots,N\}, E)$ and the augmented graph $\mathbb{H} = (\{0,\dots,N\}, E \cup F)$ with $F = \bigcup_{i\in\{1\dots,N\}}\{0,i\}$ that represents the electrical network. It holds:*

$$\min_{i\in\{1,\dots,N\}}\left(\nu_i^2 \oplus \bigoplus_{j\in N_i}\sigma_{ij}^2\right) \leq H_\infty \leq \max_{i\in\{1,\dots,N\}}\nu_i^2.$$

Proof. Recall (5) and observe that each R_{0i}^{eff} can be estimated by using Rayleigh's monotonicity law [21, Chapter 9]. For the upper bound, we substitute with open circuits all the edges $\{i,j\}$ with $j \in N_i$ (note that $0 \notin N_i$), obtaining that $R_{0i}^{eff} \leq \nu_i^2$ for every i. For the lower bound, we substitute with short circuits all the edges $\{0,j\}$, for $j \in N_i$, obtaining that

$$R_{0i}^{eff} \geq \nu_i^2 \oplus \bigoplus_{j\in N_i}\sigma_{ij}^2.$$

The result then follows from (5). □

To highlight the role of the topology, let us assume that $\nu_i^2 = \nu^2$ for every $i \in \{1,\dots,N\}$ and that $\sigma^2_{\{i,j\}} = \sigma^2$ for every $\{i,j\} \in E$. The homogeneous measurements condition is insightful for all those cases where the heterogeneity of the measurements is not pathological, i.e. the ν_i^2 and $\sigma^2_{\{i,j\}}$ enjoy uniform upper and lower bounds, which behave accordingly. Under these assumptions, the general bounds above simplify to

$$\nu^2 \oplus \frac{\sigma^2}{d_{\max}} \leq H_\infty \leq \nu^2.$$

Let us then consider the values of H_∞ for a sequence of graphs of increasing size with fixed ν^2 and σ^2. These bounds imply that H_∞ can not diverge and is bounded away from zero if the degrees of the nodes are bounded.

In the rest of this section, we present some examples of (sequences of) graphs where H_∞ can be explicitly computed. In order to improve the readability of some expressions, we

define the ratio $\gamma = \frac{\sigma^2}{\nu^2}$ and assume without loss of generality that $\nu^2 = 1$.

The first example is the *complete* graph $\mathbb{K}_N$ over N nodes and the corresponding augmented graph as in Figure 1.

Example 1 (H_∞ on complete graphs)**.** For symmetry reasons, H_∞ is equal to the effective resistance between node 1 and the reference 0, which we denote as $R^{eff}_{01} =: R^N_{\mathbb{K}}$. When we compute R^{eff}_{01}, all the remaining nodes $i \notin \{0,1\}$ have the same potential and we can substitute the resistors on the edges $\{i,j\}$ with $i,j \notin \{0,1\}$ by short circuits. By doing so,

$$H_\infty = R^N_{\mathbb{K}} = 1 \oplus \left(\frac{\gamma+1}{N-1}\right) = \left(1 + \frac{N-1}{\gamma+1}\right)^{-1} = \frac{1+\gamma}{N+\gamma}.$$

For any finite γ, the error H_∞ is decreasing in N and tends to zero as the number of agents grows large.

This example shows that the lower bound of Proposition 2 is asymptotically tight on the complete graphs topology with homogeneous measurement noise. The upper bound of Proposition 2 is achieved when the graph $\mathbb{G}$ is made by isolated nodes (i.e. the edge set E is empty).

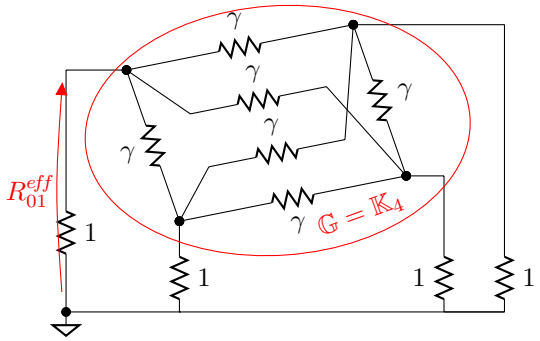

Fig. 1. The graph $\mathbb{G}$ is the complete graph $\mathbb{K}_4$.

Graphs with bounded degree have positive H_∞ as $N \to \infty$. An explicit computation can be carried out on the cycle graph. Preliminarily, we briefly consider the *line* graph as in Figure 2.

Lemma 3 (Resistances on line graphs)**.** *Let $\mathbb{L}_N$ be the line graph with N nodes and let $i = 1$ be one of the end-vertices of $\mathbb{L}_N$. Define $R^N_{\mathbb{L}} := R^{eff}_{01}$, the resistance between 1 and 0 on the corresponding augmented graph, and $R^\infty_{\mathbb{L}} = \lim_N R^N_{\mathbb{L}}$. Then,*

$$R^\infty_{\mathbb{L}} = \frac{-\gamma + \sqrt{\gamma^2 + 4\gamma}}{2}.$$

Proof. The sequence $R^N_{\mathbb{L}}$ can be computed recursively as

$$\begin{cases} R^1_{\mathbb{L}} &= 1 \\ R^{N+1}_{\mathbb{L}} &= 1 \oplus \left(\gamma + R^N_{\mathbb{L}}\right) = \frac{R^N_{\mathbb{L}} + \gamma}{R^N_{\mathbb{L}} + \gamma + 1}. \end{cases}$$

The nonnegative sequence $R^N_{\mathbb{L}}$ is thus decreasing and its limit satisfies

$$R^\infty_{\mathbb{L}} = \frac{R^\infty_{\mathbb{L}} + \gamma}{R^\infty_{\mathbb{L}} + \gamma + 1}.$$

Solving this equation gives the result. □

We are now ready to consider the *cycle* graph $\mathbb{C}_N$ with N nodes, that is, with $\{i,j\} \in E$ iff $|i-j| \in \{\pm 1, \pm(N-1)\}$.

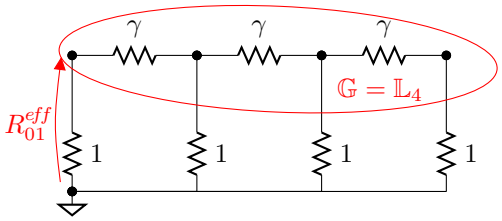

Fig. 2. The graph $\mathbb{G}$ is the line graph $\mathbb{L}_4$.

Proposition 4 (H_∞ on cycle graphs)**.** *Consider the cycle graph $\mathbb{C}_N$ and define $R^N_{\mathbb{C}} := R^{eff}_{01}$, the resistance between 1 and 0 on the corresponding augmented graph. Then, $H_\infty = R^N_{\mathbb{C}}$, $R^N_{\mathbb{C}}$ is a decreasing function of N, and*

$$R^\infty_{\mathbb{C}} := \lim_{N\to\infty} R^N_{\mathbb{C}} = 1 \oplus \left(\frac{1}{2}\left(\gamma + R^\infty_{\mathbb{L}}\right)\right) = \frac{\sqrt{\gamma}}{\sqrt{\gamma+4}}.$$

Proof. Symmetry implies that H_∞ is equal to the effective resistance between any node and the reference. In order to compute the asymptotics of $R^N_{\mathbb{C}} := R^{eff}_{01}$, we use Lemma 3. If N is odd, by symmetry, the two consecutive nodes $\frac{N+1}{2}$ and $\frac{N+3}{2}$ are at the same potential. We can then substitute the resistor on the edge $\{\frac{N+1}{2}, \frac{N+3}{2}\}$ by a open circuit and compute

$$R^N_{\mathbb{C}} = 1 \oplus \left(\frac{1}{2}\left(\gamma + R^{\frac{N-1}{2}}_{\mathbb{L}}\right)\right).$$

If N is even, we can use Rayleigh's monotonicity law to show that $R^{N+1}_{\mathbb{C}} \le R^N_{\mathbb{C}} \le R^{N-1}_{\mathbb{C}}$. To prove the lower bound, we substitute the resistor on the edge $\{0, \frac{N}{2}+1\}$ with an open circuit (this increases the effective resistance) and we observe that by symmetry no current flows on the edges $\{\frac{N}{2}, \frac{N}{2}+1\}$ and $\{\frac{N}{2}+1, \frac{N}{2}+2\}$. We thus put open circuits and compute

$$R^N_{\mathbb{C}} \le 1 \oplus \left(\frac{1}{2}\left(\gamma + R^{\frac{N}{2}-1}_{\mathbb{L}}\right)\right) = R^{N-1}_{\mathbb{C}}.$$

To prove the upper bound, we add a second resistance of value 1 in parallel to edge $\{0, \frac{N}{2}+1\}$ (this decreases the effective resistance). By symmetry, the current received by node $\frac{N}{2}+1$ from either node $\frac{N}{2}$ and $\frac{N}{2}+2$ is the same and is equally routed in the two parallel resistances. Therefore, we can split node $\frac{N}{2}$ into two distinct nodes and get

$$R^{N+1}_{\mathbb{C}} = 1 \oplus \left(\frac{1}{2}\left(\gamma + R^{\frac{N}{2}}_{\mathbb{L}}\right)\right) \le R^N_{\mathbb{C}}.$$

We conclude that the sequence $R^N_{\mathbb{C}}$ is decreasing and $R^\infty_{\mathbb{C}} = 1 \oplus \left(\frac{1}{2}\left(\gamma + R^\infty_{\mathbb{L}}\right)\right) = \frac{\sqrt{\gamma}}{\sqrt{\gamma+4}}$. □

V. Gradient descent algorithm

As Φ is convex, it is natural to consider gradient descent algorithms for its minimization. A gradient descent iterate can be defined from (1) as:

$$\begin{aligned} \mathbf{x}[t+1] &= \mathbf{x}[t] - \frac{\tau}{2}\nabla\Phi\left(\mathbf{x}[t]\right) \\ &= \mathbf{x}[t] - \tau\left[M\mathbf{x}[t] - \left(A^\top\Sigma^{-1}\mathbf{b} + \mathcal{N}^{-1}\mathbf{x}^0\right)\right] \\ &= \left(I - \tau M\right)\mathbf{x}[t] + \tau A^\top\Sigma^{-1}\mathbf{b} + \tau\mathcal{N}^{-1}\mathbf{x}^0 \end{aligned}$$

for a suitable $\tau > 0$. Equivalently, we may write the algorithm as

$$\begin{cases} \mathbf{x}[t+1] = Q\,\mathbf{x}[t] + \mathbf{w} \\ \mathbf{x}[0] = \mathbf{x}^0 \end{cases} \tag{6}$$

where $Q := I - \tau M$ and $\mathbf{w} := \tau A^\top \Sigma^{-1}\mathbf{b} + \tau \mathcal{N}^{-1}\mathbf{x}^0$. Remarkably, this algorithm is distributed, in the following sense. The matrix Q is adapted to the graph $\mathbb{G}$, *i.e.*, $Q_{ij} = 0$ if $\{i,j\} \notin E$: then, in order to update a component as $x_i[t+1] = \sum_j Q_{ij} x_j[t] + w_i$, the algorithm requires communication only with the nodes which are neighbors of i in the graph $\mathbb{G}$ and thus already share a measurement.

From here on in this section, we shall make the following standing assumption, which is sufficient to our results and permits a streamlined analysis of algorithm (6).

Assumption 1. *The parameter $\tau > 0$ is such that:*

$$\tau < \min_{i \in \{1,\dots,N\}} \nu_i^2 \oplus \bigoplus_{j \in N_i} \sigma_{ij}^2. \tag{7}$$

If $\nu_i = \nu$ for every vertex i and $\sigma_{\{i,j\}} = \sigma$ for every edge $\{i,j\}$ in the graph, then (7) reduces to $\tau < \nu^2 \oplus \frac{\sigma^2}{d_{\max}}$.

Before proceeding, let us observe that matrices M and $Q = I - \tau M$ are real symmetric and are diagonalizable with the same complete set of orthonormal eigenvectors $v^{(i)}$. The corresponding eigenvalues are μ_i and ξ_i respectively, with $\xi_i = 1 - \tau\mu_i$.

Lemma 5 (Stability)**.** *The eigenvalues of Q are such that $|\xi_i| \le 1 - \tau\hat{\nu}^{-2}$, with $\hat{\nu}^{-2} = \left(\max \nu_i^2\right)^{-1}$.*

Proof. From the definitions of Q and M, we obtain $Q = I - \tau A^\top \Sigma^{-1} A - \tau \mathcal{N}^{-1}$. Then

$$Q_{ii} = 1 - \tau\nu_i^{-2} - \tau \textstyle\sum_{j \in N_i} \sigma_{\{i,j\}}^{-2}$$
$$Q_{ij} = \begin{cases} \tau\sigma_{\{i,j\}}^{-2} & \text{if } \{i,j\} \in E \\ 0 & \text{if } \{i,j\} \notin E \end{cases} \quad i \neq j$$

and Assumption 1 implies that $\left[\nu_i^{-2} + \sum_{j \in N_i} \sigma_{\{i,j\}}^{-2}\right]\tau < 1$, for every i. Then, Gershgorin circle theorem implies that the eigenvalues of Q belong to the union of intervals

$$\bigcup_i \left[-1 + \tau\nu_i^{-2}, 1 - \tau\nu_i^{-2}\right]$$

and the result follows. □

Thanks to this lemma, we can easily prove the convergence of the proposed algorithm.

Proposition 6 (Convergence)**.** *The algorithm* (6) *converges at exponential rate $1 - \tau\hat{\nu}^{-2}$ to the optimal estimate $\mathbf{x}^*$ in* (3)*.*

Proof. By solving the recursion (6) we have

$$\mathbf{x}[t] = Q^t \mathbf{x}^0 + \sum_{n=0}^{t-1} Q^n \mathbf{w} \tag{8}$$

By Lemma 5, the matrix Q is asymptotically stable and the algorithm is exponentially convergent. Then, we can compute

$$\begin{aligned} \mathbf{x}[t] &= Q^t\mathbf{x}^0 + (I-Q)^{-1}(I-Q)\sum_{n=0}^{t-1} Q^n \mathbf{w} \\ &= Q^t\mathbf{x}^0 + (I-Q)^{-1}(I-Q^t)\mathbf{w} \end{aligned}$$

and $\lim_{t\to\infty} \mathbf{x}[t] = M^{-1}\left(A^\top\Sigma^{-1}\mathbf{b} + \mathcal{N}^{-1}\mathbf{x}^0\right) = \mathbf{x}^*$. □

A. Mean square error

To evaluate the algorithm performance, we follow the approach in [15] and define the performance metric as the mean square error between the current estimate $\mathbf{x}[t]$ and the true configuration $\bar{\mathbf{x}}$:

$$H_t := \frac{1}{N}\mathbb{E}\left[\|\mathbf{x}[t] - \bar{\mathbf{x}}\|_2^2\right],$$

where the expectation is taken on both the relative measurement noise $\mathbf{n}$ and the absolute measurement noise $\mathbf{a}$. Note that $H_\infty = \lim_{t\to\infty} H_t$. The performance metric can be computed in terms of the spectrum of the matrix Q and of the coefficients $\phi_i := {v^{(i)}}^\top \mathcal{N} v^{(i)}$.

Proposition 7 (Mean square performance)**.** *The following equality holds*

$$H_t = \frac{1}{N}\sum_{i=1}^{N}\left[\phi_i\,\xi_i^{2t} + \tau\frac{1-\xi_i^{2t}}{1-\xi_i}\right]. \tag{9}$$

Proof. We express $\mathbf{w}$ in terms of $\bar{\mathbf{x}}$, $\mathbf{n}$ and $\mathbf{a}$ as

$$\begin{aligned} \mathbf{w} &= \tau A^\top\Sigma^{-1}A\bar{\mathbf{x}} + \tau A^\top\Sigma^{-1}\mathbf{n} + \tau\mathcal{N}^{-1}\bar{\mathbf{x}} + \tau\mathcal{N}^{-1}\mathbf{a} \\ &= (I-Q)\,\bar{\mathbf{x}} + \tau A^\top\Sigma^{-1}\mathbf{n} + \tau\mathcal{N}^{-1}\mathbf{a} \end{aligned}$$

Given $\mathbf{w}$ and (8) we compute $\mathbf{x}[t] - \bar{\mathbf{x}}$ as

$$\mathbf{x}[t] - \bar{\mathbf{x}} = Q^t\mathbf{a} + \tau\sum_{n=0}^{t-1} Q^n\left(A^\top\Sigma^{-1}\mathbf{n} + \mathcal{N}^{-1}\mathbf{a}\right) \tag{10}$$

From the definition of H_t we have

$$H_t = \frac{1}{N}\mathbb{E}\left[\operatorname{tr}\left[(\mathbf{x}[t]-\bar{\mathbf{x}})(\mathbf{x}[t]-\bar{\mathbf{x}})^\top\right]\right]$$

By using (10) and through some (omitted) algebraic manipulations involving the properties of the trace operator, we get

$$\begin{aligned} H_t &= \frac{1}{N}\operatorname{tr}\left[\mathcal{N}Q^{2t} + \tau\left(I + Q^t\right)\sum_{n=0}^{t-1} Q^n\right] \\ &= \frac{1}{N}\operatorname{tr}\left[\mathcal{N}Q^{2t} + \tau\left(I + Q^t\right)\left(I - Q^t\right)(I-Q)^{-1}\right] \\ &= \frac{1}{N}\operatorname{tr}\left[\mathcal{N}Q^{2t} + \tau\left(I - Q^{2t}\right)(I-Q)^{-1}\right]. \end{aligned}$$

The result follows from the spectral decomposition of Q. □

The monotonicity property of H_t is stated in the next result.

Proposition 8 (Monotonicity of H_t)**.** *H_t is nonincreasing and is strictly decreasing if $\mathbb{G}$ contains at least two connected nodes.*

Proof. We will show that each of the terms in the sum (9) is nonincreasing in t. Let us compute the finite increment

$$\begin{aligned} H_{t+1} - H_t &= \frac{1}{N}\sum_{i=1}^{N}\xi_i^{2t}\left[\phi_i\left(\xi_i^2 - 1\right) + \tau\left(\xi_i + 1\right)\right] \\ &= \frac{1}{N}\sum_{i=1}^{N}\xi_i^{2t}\left(\xi_i + 1\right)\left[\phi_i\left(\xi_i - 1\right) + \tau\right] \\ &= \frac{1}{N}\sum_{i=1}^{N}\xi_i^{2t}\left(\xi_i + 1\right)\tau\left(1 - \phi_i\,\mu_i\right), \end{aligned}$$

where we used $\xi_i = 1 - \tau\mu_i$. Notice that, since $|\xi_i| < 1$ and $\tau > 0$, every term is nonincreasing if $\phi_i\,\mu_i \geq 1$. Using the definition of ϕ_i and expanding $\mu_i = {v^{(i)}}^\top M v^{(i)}$, we have

$$\begin{aligned}\phi_i\,\mu_i &= \left({v^{(i)}}^\top \mathcal{N} v^{(i)}\right)\left({v^{(i)}}^\top A^\top \Sigma^{-1} A v^{(i)} + {v^{(i)}}^\top \mathcal{N}^{-1} v^{(i)}\right)\\ &\geq \left({v^{(i)}}^\top \mathcal{N} v^{(i)}\right)\left({v^{(i)}}^\top \mathcal{N}^{-1} v^{(i)}\right)\end{aligned}$$

using ${v^{(i)}}^\top A^\top \Sigma^{-1} A v^{(i)} \geq 0$. Since $\mathcal{N}$ and $\mathcal{N}^{-1}$ are diagonal

$$\phi_i\,\mu_i \geq \left[\textstyle\sum_j |v_j^{(i)}|^2\,\nu_j^2\right]\left[\textstyle\sum_j |v_j^{(i)}|^2\,\nu_j^{-2}\right] \geq 1$$

where the last inequality holds since the weighted arithmetic mean is never smaller than the weighted harmonic mean. Furthermore, if there exists i such that $\phi_i\,\mu_i > 1$ and $\xi_i \neq 0$, then H_t is strictly decreasing. To get $\phi_i\,\mu_i > 1$, we need ${v^{(i)}}^\top A^\top \Sigma^{-1} A v^{(i)} > 0$, which happens if $\mathbb{G}$ contains at least two connected nodes. □

B. Limited need for cooperation

For every $\varepsilon > 0$, we can define a near-optimal stopping time, after which the estimation error is only a $(1+\varepsilon)$ factor larger than the optimal one:

$$t_\varepsilon^* = \inf\left\{t \,:\, H_t < (1+\varepsilon) H_\infty\right\}.$$

A consequence of the stability of the matrix Q, uniform in N and not dependent on the topology of the network (Lemma 5), is the following estimate of the near-optimal stopping time, illustrated by simulations in Figure 3.

Proposition 9 (Universal bound on stopping time)**.** *It holds*

$$t_\varepsilon^* \leq \frac{\hat\nu^2}{2\tau}\log\left(\frac{2\hat\nu^2}{\tau\varepsilon}\right), \tag{11}$$

where $\hat\nu^2 = \max \nu_i^2$.

Proof. Using (9) in the definition of t_ε^*, we immediately deduce that

$$t_\varepsilon^* = \inf\left\{t \,:\, \sum_{i=1}^N \left[\phi_i\,\xi_i^{2t} + \tau\frac{1-\xi_i^{2t}}{1-\xi_i}\right] < \tau\sum_{i=1}^N \frac{1+\varepsilon}{1-\xi_i}\right\}.$$

By taking an upper bound on the second term of the left-hand side of the inequality, we have

$$t_\varepsilon^* \leq \inf\left\{t \,:\, \sum_{i=1}^N \phi_i\,\xi_i^{2t} < \tau\varepsilon\sum_{i=1}^N \frac{1}{1-\xi_i}\right\}.$$

Recall that $\phi_i = {v^{(i)}}^\top \mathcal{N} v^{(i)}$ with the $v^{(i)}$s forming an orthonormal basis for $\mathbb{R}^N$. Therefore the ϕ_is are convex combinations of ν_i^2 which can be upper-bounded by $\hat\nu^2$. Recall also that Assumption 1 implies $|\xi_i| \leq 1 - \tau\hat\nu^{-2}$ and $\frac{1}{1-\xi_i} \geq \frac{1}{2-\tau\hat\nu^{-2}} > \frac{1}{2}$. Then, we can upper-bound each term of the summation of the left-hand side and lower-bound the terms of the sum of the right-hand side.

$$t_\varepsilon^* \leq \inf\left\{t \,:\, \hat\nu^2\left(1-\tau\hat\nu^{-2}\right)^{2t} < \frac{\tau\varepsilon}{2}\right\}.$$

By solving for t in the above inequality we get

$$t_\varepsilon^* \leq \frac{\log\left(\frac{2\hat\nu^2}{\tau\varepsilon}\right)}{2\log\frac{1}{1-\tau\hat\nu^{-2}}},$$

and then the result follows. □

This result shows that in order to achieve a certain accuracy (relative to the optimal estimator) the necessary number of steps does not depend on the topology of the measurement graph or even on the number of nodes. On a large graph of size N, this fact implies that only a limited portion of the graphs (independent of N) needs to be "explored". Thus, we may say that only a limited amount of cooperation is necessary to solve this estimation problem.

This phenomenon should be compared with other networked estimation problems. Let us briefly consider the following fundamental example: all nodes take noisy measurements of the same quantity θ and perform an average consensus algorithm to obtain the optimal estimate. The analysis in [15] shows that in this case $H_t := \frac{1}{N}\mathbb{E}\left[\|\mathbf{x}[t] - \theta\mathbf{1}\|\right] = \max\{1/\sqrt{t}, 1/N\}$ on cycle graphs. This fact implies that a given accuracy requires a time proportional to N^2. Hence, in the case of consensus, the cooperation of all nodes is required to meet the specification.

Remark 1 (Vanishing relative noise)**.** *Let us assume homogeneous measurements $\nu^2 = 1$ and consider the limit $\sigma^2 \to 0$. By the change of variable $z[t] = \mathbf{x}[t] - \bar{\mathbf{x}}$, algorithm* (6) *becomes equivalent to an average consensus algorithm with symmetric communication graph $\mathbb{G}$ and transition matrix $I - \frac{c}{d_{\max}} A^\top A$, where $c \in (0,1)$. Consistently with the discussion above, the bound* (11) *diverges in such a limit.*

Remark 2 (Limited benefit in Jacobi algorithm)**.** *As we did in Section III, the problem can be reformulated as the estimation from relative measurement only, by adding a virtual reference node. Within this reformulation, the results in the PhD thesis of Barooah [22, Thm 3.3.4] permit to deduce that also the Jacobi algorithm enjoys a limited benefit property similar to* (11) *when applied to system* (2)*.*

VI. Conclusion and future work

This note has studied the estimation of a distributed state based on absolute and relative measurements. Using an intuitive electrical interpretation of the problem, we have shown that the error of the least squares optimal estimator depends on the topology of the graph that encodes the measurements. Provided the heterogeneity between the measurements is bounded, the topology determines whether the error decreases to zero as the number of unknown variables grows to infinity: for instance, this happens on fully connected networks, but not on cycle networks.

We have assumed that all nodes have access to absolute information: it is then natural to ask what happens if only some of the nodes have access to it. In principle, our framework (including the electrical interpretation) can be extended to this more general case, but some parts of the analysis would be significantly different. For instance, Proposition 8 does not

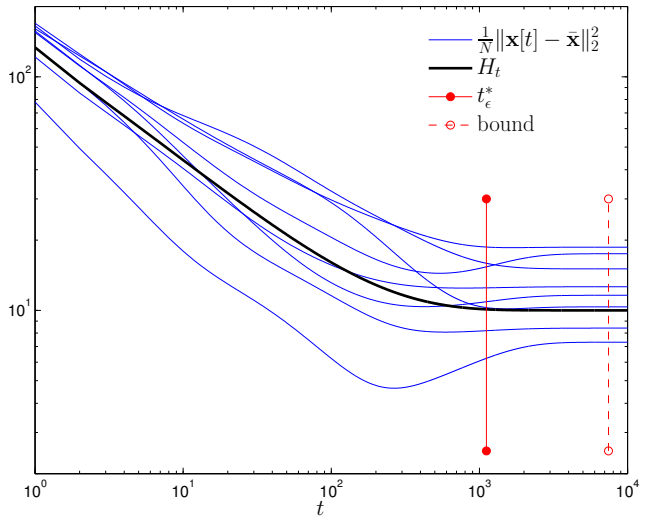

Fig. 3. Simulation of algorithm (6) on a cycle graph with $N = 160$, $\mathcal{N} = 400\, I$, $\Sigma = I$ (I is the $N \times N$ identity matrix), $\tau = \frac{1}{3}$, $\varepsilon = 0.01$. The plot shows the mean square error of the algorithm (6) starting from several initial conditions, together with the expected performance H_t. The vertical solid line indicates the near-optimal stopping time t_ε^* and the vertical dashed line the corresponding upper bound (11).

hold and the cost H_t is not decreasing. A detailed study is left to future research.

In this work we have also studied a gradient descent algorithm to compute the optimal estimator. Its analysis has brought an interesting insight: the ratio between the error of the current estimate and the optimal error can be made arbitrarily small within a number of iterates that does not depend on the number of unknowns, the number of measurements, or the topology of the graph that puts them in relation. This finding suggests that the estimate of a given node does not essentially benefit from measurements about nodes that are "far" in the graph. In this sense, the advantages of cooperation are limited in our problem, which regards the distributed estimation of a *distributed* parameter. We have argued that this property is not shared by other estimation problems, like the distributed estimation of a *common* parameter, where global cooperation improves the estimate. In fact, the literature has investigated the advantages of cooperation in distributed estimation and learning (*e.g.*, [23, Chapter 12] [24]), but the question of quantifying the right amount of cooperation has not been systematically approached yet.

REFERENCES

[1] P. Barooah and J. Hespanha, "Estimation on graphs from relative measurements," *IEEE Control Systems Magazine*, vol. 27, no. 4, pp. 57–74, 2007.

[2] A. Schindler, "Vehicle self-localization with high-precision digital maps," in *2013 IEEE Intelligent Vehicles Symposium (IV)*, Brisbane, Australia, June 2013, pp. 141–146.

[3] X. Jiang, L.-H. Lim, Y. Yao, and Y. Ye, "Statistical ranking and combinatorial Hodge theory," *Mathematical Programming*, vol. 127, no. 1, pp. 203–244, 2011.

[4] A. Giridhar and P. Kumar, "Distributed clock synchronization over wireless networks: Algorithms and analysis," in *IEEE 45th Conference on Decision and Control*, Dec. 2006, pp. 4915–4920.

[5] P. Doyle and J. Snell, *Random Walks and Electric Networks*, ser. Carus Monographs. Mathematical Association of America, 1984.

[6] P. Barooah and J. Hespanha, "Estimation from relative measurements: Electrical analogy and large graphs," *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2181–2193, 2008.

[7] ——, "Error scaling laws for linear optimal estimation from relative measurements," *IEEE Transactions on Information Theory*, vol. 55, no. 12, pp. 5661–5673, 2009.

[8] W. S. Rossi, P. Frasca, and F. Fagnani, "Average resistance of toroidal graphs," *SIAM Journal on Control and Optimization*, vol. 53, no. 4, pp. 2541–2557, 2015.

[9] B. Bamieh, M. Jovanovic, P. Mitra, and S. Patterson, "Coherence in large-scale networks: Dimension-dependent limitations of local feedback," *IEEE Transactions on Automatic Control*, vol. 57, no. 9, pp. 2235–2249, 2012.

[10] A. Ghosh, S. Boyd, and A. Saberi, "Minimizing effective resistance of a graph," *SIAM Review*, vol. 50, no. 1, pp. 37–66, 2008.

[11] E. Lovisari, F. Garin, and S. Zampieri, "Resistance-based performance analysis of the consensus algorithm over geometric graphs," *SIAM Journal on Control and Optimization*, vol. 51, no. 5, pp. 3918–3945, 2013.

[12] A. K. Chandra, P. Raghavan, W. L. Ruzzo, R. Smolensky, and P. Tiwari, "The electrical resistance of a graph captures its commute and cover times," *Computational Complexity*, vol. 6, no. 4, pp. 312–340, 1996.

[13] B. Osting, C. Brune, and S. J. Osher, "Optimal data collection for improved rankings expose well-connected graphs," *Journal of Machine Learning Research*, vol. 15, pp. 2981–3012, 2014.

[14] W. S. Rossi, P. Frasca, and F. Fagnani, "Transient and limit performance of distributed relative localization," in *IEEE 51st Annual Conference on Decision and Control*, Dec. 2012, pp. 2744–2748.

[15] F. Garin and S. Zampieri, "Mean square performance of consensus-based distributed estimation over regular geometric graphs," *SIAM Journal on Control and Optimization*, vol. 50, no. 1, pp. 306–333, 2012.

[16] N. Freris and A. Zouzias, "Fast distributed smoothing of relative measurements," in *IEEE 51st Annual Conference on Decision and Control*, Dec. 2012, pp. 1411–1416.

[17] C. Ravazzi, P. Frasca, R. Tempo, and H. Ishii, "Ergodic randomized algorithms and dynamics over networks," *IEEE Transactions on Control of Network Systems*, vol. 2, no. 1, pp. 78–87, 2015.

[18] A. Carron, M. Todescato, R. Carli, and L. Schenato, "An asynchronous consensus-based algorithm for estimation from noisy relative measurements," *IEEE Transactions on Control of Network Systems*, vol. 1, no. 3, pp. 283–295, 2014.

[19] W. S. Rossi, P. Frasca, and F. Fagnani, "Limited benefit of cooperation in distributed relative localization," in *IEEE 52nd Annual Conference on Decision and Control*, Dec. 2013, pp. 5427–5431.

[20] A. Tarantola, *Inverse problem theory and methods for model parameter estimation*. SIAM, 2005.

[21] D. A. Levin, Y. Peres, and E. L. Wilmer, *Markov Chains and Mixing Times*. American Mathematical Society, 2009.

[22] P. Barooah, "Estimation and control with relative measurements: Algorithms and scaling laws," Ph.D. dissertation, University of California, Santa Barbara, CA, USA, July 2007.

[23] A. H. Sayed, "Adaptation, learning, and optimization over networks," *Foundations and Trends in Machine Learning*, vol. 7, no. 4-5, pp. 311–801, 2014.

[24] A. P. Schoellig, J. Alonso-Mora, and R. D'Andrea, "Limited benefit of joint estimation in multi-agent iterative learning," *Asian Journal of Control*, vol. 14, no. 3, pp. 613–623, 2012.