An exponential-rate consensus-based algorithms for estimation from relative measurements: implementation and performance analysis

Ruggero Carli Andrea Carron Luca Schenato Marco Todescato

***Abstract*— In this work we address the problem of optimal estimating the position of each agent in a network from relative noisy vectorial distances with its neighbors. Although the problem can be cast as a standard least-squares problem, the main challenge is to devise scalable algorithms that allow each agent to estimate its own position by means of only local communication and bounded complexity, independently of the network size and topology. We first propose a synchronous consensus-based algorithm that is guaranteed to have exponentially convergence rate to the optimal centralized least-squares solution. We then extend this algorithm to more realistic asynchronous implementations via the use of local memory variables. We show that exponential convergence is still guaranteed under both uniform persistent communication and randomized persistent communication protocols. We finally complement the analytical results with some numerical simulations.**

I. Introduction

The proliferation of relatively inexpensive devices capable of communicating, computing, sensing, interacting with the environment and storing information is promising an unprecedented number of novel applications throughout the cooperation of these devices toward a common goal. These applications include swarm robotics, wireless sensor networks, smart energy grids, smart traffic networks, smart camera networks. These applications also pose new challenges, of which *scalability* is one of the major ones. Scalability is intended as the ability for an application to continue functioning without any dramatic performance degradation even if the number of devices involved keep increasing. In particular, an application is scalable if it is not necessary to increase HW resources or to adopt a more complex SW algorithms in each device even if the total number of devices increases.

In this work we address the problem of designing algorithms that are capable to reconstruct the optimal estimate of the location of a device from noisy relative measurements from its neighbors in a connected network. In particular, we want to design distributed algorithms that allows each device to reconstruct its own position only from exchanging information with its neighbors, regardless of the size of the network. Moreover, these algorithms must be scalable, i.e. their computational complexity, bandwidth and memory requirements should be independent of the network size. These algorithms will be obtained as the solution of an optimization problem.

Distributed optimization has being attracting ever growing attention in the past years since many problems in large scale network have been cast as convex optimization problem. In particular, a large class of problem can be cast as the solution of the following optimization problem

$$x^* = \begin{array}{ll} \text{argmin}_{x_1,\dots,x_N} & \sum_{i=1}^N f_i(x_i) \\ s.t. & x_i = x_j, \forall i,j \end{array} \tag{1}$$

where N is the number of nodes, $x_i \in \mathbb{R}^m$, and f_i are convex functions. The function f_i represents local cost of each agent i, but each agent must compute the minimizer x^* of the sum all local costs. The local cost functions are separable but the additional contraint that all local variables x_i must be the same, makes the optimization problem coupled. Many problems can be cast in this terms such as distributed least squares [1], [2], map building [3], network utility maximization [4], distributed learning and support vector machines [5]. In these context several approaches have been proposed such as the distributed subgradient methods (SDMs) [6], the alternating direction method of multipliers (ADMM)[5], the Newton-Raphson consensus [7], the control-based methods [8], and the Fast-Lipshitz [9], and the distributed LP [10].

The problem at hand in this work is of a different type and can be cast as the following unconstrained optimization problem:

$$\min_{x_1,\dots,x_N} \sum_{(i,j)\in\mathcal{E}}^{|\mathcal{E}|} f_{ij}(x_i - x_j) \tag{2}$$

where $x_i \in \mathbb{R}^\ell$, $\mathcal{E}$ represents all the pair of nodes for which are available relative measurements and f_{ij} are convex functions. Differently from the previous optimization problem, the local variables are unconstrained, but the cost functions are now not separable, and therefore the problem is once again coupled. Many problems can be cast in this framework such as sensor localization [11], [12], sensor calibration [2], clock synchronization [13] and camera localization [14], [15]. For example, in the context of localization from vectorial relative distance in a plane, the cost function f_{ij} are given by:

$$f_{ij}(x_i - x_j) = ||x_i - x_j - z_{ij}||^2$$

where $z_{ij} \in \mathbb{R}^\ell$ is the noisy measurement of the relative (vector) distance of node i from node j. As a consequence, the optimization problem in Eqn. (2) becomes a distributed least-square problem that in principle could be cast as the

This work is supported by the European Community's Seventh Framework Programme [FP7/2007-2013] under grant agreement n. 257462 HYCON2 Network of excellence.

R. Carli, A. Carron, L. Schenato and M. Todescato are with the Department of Information Engineering, University of Padova, Via Gradenigo 6/a, 35131 Padova, Italy {carlirug|carronan|schenato|todescat}@dei.unipd.it.

optimization problem in Eqn. (2). However, in this case, each node will need to compute the location of all other nodes, i.e. the size of the local variable becomes of size $m = N\ell$, which according to our objective is not scalable. There exist alternative approaches that try to exploit the special structure of the problem, but they either show oscillatory behavior as in consensus-based with constant weights [2], or the convergence rate decreases only as $1/k$ as in the Randomized Kaczmarz with Under-Relaxation [16]. This slow convergence rate is mainly due to the fact that these algorithms adopt asynchronous gradient-based algorithms whose step-size decreases to zero as progress.

The contribution of this work is to provide both synchronous and asynchronous algorithms which are scalable and have proven exponential rate of convergence under mild assumptions. The main idea is to cast the problem as a consensus problem under some suitable change of coordinates, and then solve the problem by applying synchronous broadcast consensus iterations. In a more realistic asynchronous broadcast communication, the problem is solved similarly by adding some extra memory variables at each node which keep track of the estimated location of their neighbors, i.e. the nodes from which they collected the relative distance measurements. Estimates of these local variables eventually converge to the estimates of the neighbors, thus guaranteeing the convergence of the whole algorithm, at the price of some delay.

II. Problem Formulation

The problem we deal with is that of estimating N variables $x_1, \dots, x_N$ from noisy measurements of the form

$$z_{ij} := x_i - x_j + n_{ij}, \qquad i, j \in \{1, \dots, N\}, \tag{3}$$

where n_{ij} is zero-mean measurement noise. Though the variables are often vector-valued, for simplicity, in this paper we assume that $x_i \in \mathbb{R}$, $i \in \{1, \dots, N\}$.

This estimation problem can be naturally associated with a *measurement graph* $\mathcal{G} = (V; \mathcal{E})$. The vertex set V of the measurement graph consists of the set of nodes $V = \{1, \dots, N\}$ where N is the number of nodes, while its edge set $\mathcal{E}$ consists of all of the ordered pairs of nodes (i, j) such that a noisy measurement of the form (3) between i and j is available to node i. The measurement errors on distinct edges are assumed uncorrelated. The measurement graph $\mathcal{G}$ is a directed graph since $(i, j) \in \mathcal{E}$ implies the measurement z_{ij} is available to node i, while $(j, i) \in \mathcal{E}$ implies the measurement z_{ji} is available to node j, and these two are in general distinct.

To formally state the problem we aim at solving we need some preliminary definitions.

Assume $|\mathcal{E}| = M$ and let us introduce the *incidence matrix* $A \in \mathbb{R}^{M \times N}$ of $\mathcal{G}$ defined as $A = [a_{ei}]$, where $a_{ei} = 1, -1, 0$, if edge e is incident on node i and directed away from i, is incident on node i and directed toward it, or is not incident on node i, respectively.

Now, let $\mathbf{x} \in \mathbb{R}^N$ be the vector obtained stacking together all the variables $x_1, \dots, x_N$, i.e., $\mathbf{x} = [x_1, \dots, x_N]^T$, where given a vector v with v^T we denote its transpose, and let $\mathbf{z} \in \mathbb{R}^M$ and $\mathbf{n} \in \mathbb{R}^M$ be the vectors obtained stacking together all the measurements z_{ij} and the noises n_{ij}, respectively. Additionally, let $R_{ij} > 0$ denote the covariance of the zero mean error n_{ij}, i..e, $R_{ij} = \mathbb{E}[n_{ij}^2]$, where $\mathbb{E}$ denotes the expectation operator, and let $R \in \mathbb{R}^{M \times M}$ be the diagonal matrix collecting in its diagonal the covariances of the noises n_{ij}, $(i, j) \in \mathcal{E}$, i.e., $R = \mathbb{E}[\mathbf{n}\mathbf{n}^{\mathbf{T}}]$. Finally let $\mathbf{1}$ be the column vector with all components equal to one.

Observe that equation (3) can be rewritten in a vector form as

$$\mathbf{z} = A\mathbf{x} + \mathbf{n}$$

Define the set

$$\chi := \underset{\mathbf{x} \in \mathbb{R}^N}{\operatorname{argmin}} (\mathbf{z} - A\mathbf{x})^T R^{-1} (\mathbf{z} - A\mathbf{x}).$$

The goal is to construct an optimal estimate x^* of $\mathbf{x}$ in a least square sense, namely, to compute

$$x^* \in \chi \tag{4}$$

Assume the measurement graph $\mathcal{G}$ to be *weakly connected*[1], then it is well known (see [12]) that

$$\chi = \left\{ \left(A^T R^{-1} A\right)^\dagger A^T R^{-1} \mathbf{z} + \alpha \mathbf{1} \right\}.$$

Moreover let

$$x^*_{\text{opt}} = \left(A^T R^{-1} A\right)^\dagger A^T R^{-1} \mathbf{z},$$

then x^*_{opt} is the minimum norm solution of (4), i.e.,

$$x^*_{\text{opt}} = \min_{x^* \in \chi} \parallel x^* \parallel$$

The matrix $A^T R^{-1} A$ is called in literature the *Weighted Generalized Grounded Laplacian* ([12]).

Remark II.1 Observe that, just with relative measurements, determining the x_i's is only possible up to an additive constant. This ambiguity might be avoided by assuming that a node (say node 1) is used as reference node, i.e., $x_1 = 0$.

III. A synchronous distributed consensus based solution

To compute an optimal estimate x^* directly, one needs all the measurements and their covariances $(\mathbf{z}, R)$, and the topology of the measurement graph $\mathcal{G}$. In this section the goal is to compute the optimal solution in a distributed fashion, employing only local communication. In particular we assume that a node i and another node j can communicate with each other if either $(i, j) \in \mathcal{E}$ or $(j, i) \in \mathcal{E}$. Accordingly a node i is said to be a neighbor of another node j (and vicecersa) if either $(i, j) \in \mathcal{E}$ or $(j, i) \in \mathcal{E}$. For $i \in$

[1]Given a directed graph, a path from a node to another node that does not respect the orientation of the edges is called an *undirected path*. A directed graph is said to be *weakly connected* if there is a undirected path form any node to any other node.

$\{1,\dots,N\}$, by $\mathcal{N}_i$ we denote the set of neighbors of node i, namely

$$\mathcal{N}_i = \{j \in V \text{ such that either } (i,j) \in \mathcal{E} \text{ or } (j,i) \in \mathcal{E}\}.$$

In what follows we introduce a distributed solution which is based on standard linear consensus algorithm. A discussion of the linear consensus algorithm can be found in the review paper [17], hence we refrain from describing it here. Instead we make the presentation of the algorithm self-contained. At first we assume that the communications among the nodes are synchronous, namely the nodes perform their transmissions' actions at the same instant times, and design the algorithm for that scenario. We refer to this algorithm as the *synchronous consensus-based localization* algorithm (denoted hereafter as s-CL algorithm). In section IV we modify the s-CL algorithm to make it suitable to asynchronous communications.

We assume that before running the s-CL algorithm, the nodes exchange with their neighbors their relative measurements as well as the associated covariances. So every node has access to the measurements on the edges that are incident to it, whether the edge is directed to or away from it. Each node uses the measurements obtained initially for all future computations.

The s-CL algorithm is formally described as follows.

Processor states: For $i \in \{1,\dots,N\}$, node i stores in memory the measurements $\{z_{ij}, (i,j) \in \mathcal{E}\}$, $\{z_{ji}, (j,i) \in \mathcal{E}\}$, and the associated covariances $\{R_{ij}, (i,j) \in \mathcal{E}\}$, $\{R_{ji}, (j,i) \in \mathcal{E}\}$. Moreover node i stores in memory also an estimate $\hat{x}_i$ of x_i.

Initialization: Every nodes initializes its estimate to an arbitrary value.

Transmission iteration: For $k \in \mathbb{N}$, at the start of the $(k+1)$-th iteration of the algorithm, node i transmits its estimate $\hat{x}_i(k)$ to all its neighbors. It also gathers the k-th estimates of its neighbors, $\hat{x}_j(k)$, $j \in \mathcal{N}_i$.

Update iteration: For $k \in \mathbb{N}$, node i, $i \in \{1,\dots,N\}$, based on the information received from its neighbors, updates its estimate as follows

$$\hat{x}_i(k+1) := p_{ii}\,\hat{x}_i(k) + \sum_{j \in \mathcal{N}_i} p_{ij}\,\hat{x}_j(k) + b_i$$

where

$$b_i = \epsilon \sum_{(i,j)\in\mathcal{E}} R_{ij}^{-1} z_{ij} - \epsilon \sum_{(j,i)\in\mathcal{E}} R_{ji}^{-1} z_{ji}$$

and where

$$p_{ij} = \begin{cases} \epsilon(R_{ij}^{-1} + R_{ji}^{-1}) & \text{if } (i,j) \in \mathcal{E} \text{ and } (j,i) \in \mathcal{E} \\ \epsilon R_{ij}^{-1} & \text{if } (i,j) \in \mathcal{E} \text{ and } (j,i) \notin \mathcal{E} \\ \epsilon R_{ji}^{-1} & \text{if } (j,i) \in \mathcal{E} \text{ and } (i,j) \notin \mathcal{E} \end{cases}$$

and

$$p_{ii} = 1 - \sum_{j \in \mathcal{N}_i} p_{ij}$$

being ϵ a positive constant *a-priori* assigned to the nodes.

Now, let $P \in R^{N\times N}$ be the matrix defined by the weights p_{ij} above introduced. One can see that such matrix P is equal to

$$P = I - \epsilon A^T R^{-1} A$$

Moreover let

$$b = \epsilon A^T R^{-1} \mathbf{z}$$

Then the s-CL algorithm can be written in a compact form as

$$\hat{x}(k+1) = P\hat{x}(k) + b$$

To characterize the convergence properties of the s-CL algorithm, we introduce some definitions. Firstly, for $i \in \{1,\dots,N\}$, let $d_{max} = \max\{|\mathcal{N}_i|, i \in \{1,\dots,N\}\}$. Secondly let $R_{min} = \min\{R_{ij}, (i,j) \in \mathcal{E}\}$. Observe that if $0 < \epsilon < 1/(2d_{\max}R_{min}^{-1})$ then the matrix P is stochastic. Moreover if the graph $\mathcal{G}$ is weakly connected then the matrix P is primitive, i.e., it satisfies the following two properties

(i) P has only one eigenvalue equal to 1, $P\mathbf{1} = \mathbf{1}$ and $\mathbf{1}^T P = \mathbf{1}^T$, since P is double stochastic;

(ii) all the other eigenvalues of P are all strictly inside the unitary circle.

We recall that, based on standard results on nonnegative matrices and linear consensus algorithms, the above two properties imply $\lim_{k\to\infty} P^k = \frac{\mathbf{1}\mathbf{1}^T}{N}$. Moreover let ρ_{ess} denote the essential spectral radius of P, namely, the second largest in absolute value eigenvalue of P (see [18]). Clearly $0 \le \rho_{ess} < 1$. We have the following Proposition.

Proposition III.1 *Let ϵ be such that $0 < \epsilon < 1/(2d_{max}R_{min}^{-1})$. Moreover let $\hat{x}_i$, $i \in \{1,\dots,N\}$, be initialized to any real number. Let $\mathcal{G}$ be* weakly connected. *Consider the s-CL algorithm running over $\mathcal{G}$. Then the following two facts hold true*

(i) *the evolution $k \to \hat{x}(k)$ asymptotically converges to an optimal estimate $x^* \in \chi$, i.e., there exists $\alpha \in \mathbb{R}$, such that*

$$\lim_{k\to\infty} \hat{x}(k) = x^*_{opt} + \alpha\mathbf{1};$$

where α linearly depends on $x(0)$.

(ii) *the convergence is exponential, namely, there exists $C > 0$ such that*

$$\|\hat{x}(k) - \left(x^*_{opt} + \alpha\mathbf{1}\right)\| \le C\rho_{ess}^k \|\hat{x}(0) - \left(x^*_{opt} + \alpha\mathbf{1}\right)\|.$$

Proof: We start by proving item (i). Let us define the change of variable $\xi = \hat{x} - x^*_{\text{opt}}$. Since

$$x^*_{\text{opt}} = Px^*_{\text{opt}} + b$$

it is possible to write

$$\begin{aligned} \hat{x}(k+1) - x^*_{\text{opt}} &= P\hat{x}(k) + b - x^*_{\text{opt}} \\ &= P\hat{x}(k) + b - (Px^*_{\text{opt}} + b) \\ &= P(\hat{x}(k) - x^*_{\text{opt}}) \end{aligned}$$

and, in turn,

$$\xi(k+1) = P\xi(k).$$

The above equation describes the iteration of the classical consensus algorithm. Since P is a primitive doubly stochastic matric, we have that

$$\xi(k) \to \frac{\mathbf{1}\mathbf{1}^T}{N}\xi(0)$$

where $\xi(0) = x(0) - x^*_{\text{opt}}$. This implies that

$$x(k) \to x^*_{\text{opt}} + \frac{\mathbf{1}\mathbf{1}^T}{N}x(0) - \frac{\mathbf{1}\mathbf{1}^T}{N}x^*_{\text{opt}}$$

Being $\frac{\mathbf{1}\mathbf{1}^T}{N}x^*_{\text{opt}} = 0$ this prove the result.

Concerning item (ii) it is well known ([18]) that the convergence rate of a consensus algorithm ruled by a primitive matrix P, is exponential and is upper bounded by the essetial spectral radius ρ_{ess}. ■

Remark III.2 The s-CL algorithm is similar to the algorithm proposed in [19]. However in [19], the measurement graph is assumed to be undirected, namely, both measurements z_{ij} and z_{ji} are available to node i and j under the additional assumption that $z_{ij} = -z_{ji}$.

Remark III.3 The authors in [20] solved the problem formulated in (4) proposing a synchronous algorithm that implements the Jacobi iterative method. The performance of this algorithm, in terms of rate of convergence to the optimal solution, is similar, for many families of measurement graphs, to the one of the synchronous consensus-based algorithm introduced in this section. However, to the best of our knowledge, there is no an asynchronous implementation of the Jacobi-based algorithm proposed in [20], which has been shown to be provably convergent to the optimal solution.

IV. An asynchronous implementation of distributed consensus based solution

The distributed algorithm illustrated in the previous section, has an important limitation: it is applicable only to sensors' networks with synchronized and reliable communication. Indeed, the s-CL algorithm requires that there exists a predetermined common communication schedule for all nodes and, at each communication round, each node must simultaneously and reliably communicates its information. The aim of this section is to reduce the communication requirements of the s-CL algorithm, in particular in terms of synchronization. To do so, we next introduce the *asynchronous Consensus-based Localization* algorithm (denoted as a-CL hereafter). This algorithm is based on an asymmetric broadcast communication protocol. Differently from the s-CL, at each iteration of the a-CL there is only one node transmitting information to all its neighbors.

Since the actual value of neighboring estimates are not available at each iteration, we assume that each node stores in its local memory a copy of the neighbors' variables recorded from the last communication they performed. For $j \in \mathcal{N}_i$, we denote by $\hat{x}_j^{(i)}(k)$ the estimate of x_j kept in i's local memory at the end of the k-th iteration. If node j performed its last transmission to node i during h-th iteration, $h \leq k$, then $\hat{x}_j^{(i)}(k) = \hat{x}_j(h)$.

The a-CL algorithm is formally described as follows.

Processor states: For $i \in \{1, \dots, N\}$, node i stores in memory the measurements z_{ij}, z_{ji} and the covariances R_{ij}, R_{ji} for all $j \in \mathcal{N}_i$. Moreover node i stores in memory also the estimate $\hat{x}_i$ of x_i and, for $j \in \mathcal{N}_i$ an estimate $\hat{x}_j^{(i)}$ of $\hat{x}_j$,

Initialization: Every node i initializes its estimate $\hat{x}_i$ and the variables $\hat{x}_j^{(i)}$, $j \in \mathcal{N}_i$, to arbitrary values.

Transmission iteration: For $k \in \mathbb{N}$, at the start of the $(k+1)$-th iteration of the algorithm, there is only one node, say i, which transmits information to its neighbors; precisely, node i sends the value of its estimate $\hat{x}_i(k)$ to node j, $j \in \mathcal{N}_i$.

Update iteration: For $j \in \mathcal{N}_i$, node j performs the following actions in order

(i) it sets $\hat{x}_i^{(j)}(k+1) = \hat{x}_i(k)$, while for $s \in \mathcal{N}_j \setminus \{i\}$, $\hat{x}_s^{(j)}$ is left unchanged, i.e., $\hat{x}_s^{(j)}(k+1) = \hat{x}_s^{(j)}(k)$;

(ii) it updates $\hat{x}_j$ as

$$\hat{x}_j(k+1) := p_{jj}\hat{x}_j(k) + \sum_{h \in \mathcal{N}_j} p_{jh}\hat{x}_h^{(j)}(k+1) + b_j. \tag{5}$$

Clearly for $s \notin \mathcal{N}_i$, $\hat{x}_s$ is left unchanged during the $(k+1)$-th iteration of the algorithm, i.e, $\hat{x}_s(k+1) = \hat{x}_s(k)$.

We characterize now the convergence properties of the a-CL. We consider two different scenarios which we introduce in the following definition.

Definition IV.1 *A network of N nodes is said to be a* uniformly persistent communicating network *if there exists a positive integer number τ such that, for all $k \in \mathbb{N}$, each node transmits the value of its estimate to its neighbors at least once within the time interval $[k, k+\tau)$.*

A network of N nodes is said to be a randomly persistent communicating network *if there exists a N-upla of probabilities $(\beta_1, \dots, \beta_N)$ such that $\beta_i > 0$, for all $i \in \{1, \dots, N\}$, and $\sum_{i=1}^N \beta_i = 1$, and such that, for all $k \in \mathbb{N}$,*

$$\mathbb{P}\left[\textit{the transmitting node at iteration } k \textit{ is node } i\right] = \beta_i.$$

A. Convergence of the a-CL algorithm under uniform persistent communications

The following result characterizes the convergence properties of the a-CL when the network is uniformly persistent communicating.

Proposition IV.2 *Consider a* uniformly persistent communicating network *of N nodes with associated a weakly connected measurement graph $\mathcal{G}$ running the a-CL algorithm. Let ϵ be such that $0 < \epsilon < 1/(2d_{max}R^{-1}_{min})$. Moreover let $\hat{x}_i$, $i \in \{1, \dots, N\}$, $\hat{x}_j^{(i)}$, $j \in \mathcal{N}_i$, be initialized to any real number. Then the following facts hold true*

(i) *the evolution $k \to \hat{x}(k)$ asymptotically converges to an optimal estimate $x^* \in \chi$, i.e., there exists $\alpha \in \mathbb{R}$ such that*

$$\lim_{k\to\infty} \hat{x}(k) = x^*_{opt} + \alpha \mathbf{1};$$

(ii) *the convergence is exponential, namely, there exists $C > 0$ and $0 \leq \rho < 1$ such that*

$$\|\hat{x}(k) - \left(x^*_{opt} + \alpha\mathbf{1}\right)\| \leq C\rho^k \|\hat{x}(0) - \left(x^*_{opt} + \alpha\mathbf{1}\right)\|.$$

Proof: Observe that, for $i \in \{1, \ldots, N\}$ and $j \in \mathcal{N}_i$, we have $\hat{x}_i^{(j)}(k) = \hat{x}_i(\tau_i(k))$, where $\tau_i(k) < k$ denotes the iteration during which node i has perfomed its last transmission up to iteration k. Hence Equation (5) can be rewritten as

$$\hat{x}_j(k+1) := p_{jj}\hat{x}_j(k) + \sum_{h\in\mathcal{N}_j} p_{jh}\hat{x}_h(\tau_h(k)) + b_j. \quad (6)$$

Observe that

$$x^*_{\text{opt}} = Px^*_{\text{opt}} + b$$

and, in particular,

$$\left[x^*_{\text{opt}}\right]_j = \left[Px^*_{\text{opt}}\right]_j + b_j = p_{jj}\left[x^*_{\text{opt}}\right]_j + \sum_{h\in\mathcal{N}_j} p_{jh}\left[x^*_{\text{opt}}\right]_h + b_j.$$

By subtracting $\left[x^*_{\text{opt}}\right]_j$ to both the left-hand side and the right-hand side of (6) one gets

$$\begin{aligned}
&\hat{x}_j(k+1) - \left[x^*_{\text{opt}}\right]_j \\
&\quad = p_{jj}\hat{x}_j(k) + \sum_{h\in\mathcal{N}_j} p_{jh}\hat{x}_h(\tau_h(k)) + \\
&\qquad + b_j - \left(p_{jj}\left[x^*_{\text{opt}}\right]_j + \sum_{h\in\mathcal{N}_j} p_{jh}\left[x^*_{\text{opt}}\right]_h + b_j\right) \\
&\quad = p_{jj}(\hat{x}_j(k) - \left[x^*_{\text{opt}}\right]_j) + \\
&\qquad + \sum_{h\in\mathcal{N}_j} p_{jh}\left(\hat{x}_h(\tau_h(k)) - \left[x^*_{\text{opt}}\right]_h\right).
\end{aligned}$$

Let us introduce the auxiliary variable $\xi(k) = \hat{x}(k) - x^*_{\text{opt}}$. From the above equation we can write

$$\xi_j(k+1) = p_{jj}\xi_j(k) + \sum_{h\in\mathcal{N}_j} p_{jh}\xi_h(\tau_h(k)) \quad (7)$$

Observe that (7) describes the evolution of a consensus algorithm in presence of uniformly bounded delays. Hence, by invoking Proposition 1 in [21], we can conclude that $\xi(k)$ converges exponentially to consensus, and, in turn, that $\hat{x}(k)$ converges exponentially to an optimal solution x^*. ■

B. Convergence of the a-CL algorithm under randomly persistent communications

The following result characterizes the convergence properties of the a-CL when the network is randomly persistent communicating.

Proposition IV.3 *Consider a* randomly persistent communicating network *of N nodes with associated a weakly connected measurement graph $\mathcal{G}$ running the a-CL algorithm. Let ϵ be such that $0 < \epsilon < 1/(2d_{max}R^{-1}_{min})$. Moreover let $\hat{x}_i$, $i \in \{1, \ldots, N\}$, $\hat{x}_j^{(i)}$, $j \in \mathcal{N}_i$, be initialized to any real number. Then the following facts hold true*

(i) *the evolution $k \to \hat{x}(k)$ converges almost surely to an optimal solution $x^* \in \chi$, i.e., there exists $\alpha \in \mathbb{R}$ such that*

$$\mathbb{P}\left[\lim_{k\to\infty} \hat{x}(k) = x^*_{opt} + \alpha\mathbf{1}\right] = 1.$$

(ii) *the evolution $k \to \hat{x}(k)$ is exponentially convengent in mean-square sense, i.e., there exist $C > 0$ and $0 \leq \rho < 1$ such that*

$$\begin{aligned}
&\lim_{k\to\infty} \mathbb{E}\left[\|\hat{x}(k) - (x^*_{opt} + \alpha\mathbf{1})\|^2\right] \\
&\qquad \leq C\rho^k \mathbb{E}\left[\|\hat{x}(0) - (x^*_{opt} + \alpha\mathbf{1})\|^2\right].
\end{aligned}$$

Proof: We start by rewriting the updating step of the a-CL in a more convenient way. Observe preliminarily that, under the assumption of reliable communications over the network, the broadcast protocol lets only two information about the estimate of x_i, $\forall i \in V$, to flow through the network. Specifically, $\hat{x}_i(k)$, that is the current value of the estimate $\hat{x}_i$ at iteration k, and $\hat{x}_i(t'_i(k))$, being $t'_i(k)$ the iteration during which node i has performed its last transmission up to iteration k of a-CL (that is, the value of $\hat{x}_i^{(i)}$ at its last communication round). Indeed notice that, for $j \in \mathcal{N}_i$, $\hat{x}_i^{(j)}(t'') = \hat{x}_i(t'_i(k))$ for all t'' such that $t'_i(k) < t'' \leq k$. Let us define $x'_i(k) = \hat{x}_i(k)$ and $x''_i(k) = \hat{x}_i(t'_i(k))$ and, accordingly, let $x'(k) = [x'_1(k), \ldots, x'_N(k)]^T$ and $x''(k) = [x''_1(k), \ldots, x''_N(k)]^T$. Morever let $Q_i \in \mathbb{R}^{2N\times 2N}$ be defined as

$$Q_i = \begin{bmatrix} Q_{11}^{(i)} & Q_{12}^{(i)} \\ Q_{21}^{(i)} & Q_{22}^{(i)} \end{bmatrix} \quad (8)$$

where

$$\begin{aligned}
Q_{11}^{(i)} &= \sum_{h\notin\mathcal{N}_i} e_h e_h^T + \sum_{j\in\mathcal{N}_i}\left(p_{jj}e_j e_j^T + p_{ji}e_j e_i^T\right) \\
Q_{12}^{(i)} &= \sum_{j\in\mathcal{N}_i} e_j\left(\sum_{h\in\mathcal{N}_j/i} p_{jh}e_h^T\right) \\
Q_{21}^{(i)} &= e_i e_i^T \\
Q_{22}^{(i)} &= I - e_i e_i^T
\end{aligned}$$

Observe that, for $i \in \{1, \ldots, N\}$, Q_i is a $2N$-dimensional stochastic matrix. Finally let

$$B_i = \begin{bmatrix} \sum_{j\in\mathcal{N}_i} e_j^T b \\ 0_N \end{bmatrix}$$

Assume, without loss of generality, that node i is the node performing the transmission during the $k+1$-th iteration of the a-CL. Hence the udpating step of the a-CL can be written in compact form as

$$\begin{bmatrix} x'(k+1) \\ x''(k+1) \end{bmatrix} = Q_i \begin{bmatrix} x'(k) \\ x''(k) \end{bmatrix} + B_i. \quad (9)$$

Now let us introduce the auxiliary variable

$$\xi(k) = \begin{bmatrix} x'(k) \\ x''(k) \end{bmatrix} - \begin{bmatrix} x^*_{opt} \\ x^*_{opt} \end{bmatrix}$$

By exploiting the fact that, for $i \in \{1, \dots, N\}$,

$$\begin{bmatrix} x^*_{opt} \\ x^*_{opt} \end{bmatrix} = Q_i \begin{bmatrix} x^*_{opt} \\ x^*_{opt} \end{bmatrix} + B_i \tag{10}$$

we have that the variable ξ satisfies the following $2N$-dimensional recursive equation

$$\xi(t+1) = Q_i \xi(t). \tag{11}$$

Observe that $\hat{x}(k) \to x^*_{\text{opt}} + \alpha\mathbf{1}$ if and only if $\xi(k) \to \alpha\mathbf{1}$. Moreover, since Q_i is a stochastic matrix for any $i \in \{1, \dots, N\}$, we have that (11) describes the evolution of a randomized consensus algorithm whose convergence properties can be analyzed following the treatment in [22].

Now let σ be the random process such that $\sigma(k)$ denotes the node performing the transmission action at the beginning of the $k+1$-th iteration. Clearly, in the randomized scenario we are considering, we have that, for $i \in \{1, \dots, N\}$, $\mathbb{P}[\sigma(k) = i] = \beta_i$ for all k. Let

$$S(k) = \prod_{h=0}^{k} Q_{\sigma(h)}.$$

Observe that $S(k)$ inherits the same block structure of the matrices $\{Q_i\}_{i=1}^N$, namely we can write

$$S(k) = \begin{bmatrix} S_{11}(k) & S_{12}(k) \\ S_{21}(k) & S_{22}(k) \end{bmatrix}$$

As consequence of Theorem 3.1 in [22] the a-CL reaches almost surely consensus if and only if, for every i and j in V

$$\mathbb{P}\left[\mathcal{E}_{ij}\right] = 1, \tag{12}$$

where

$$\mathcal{E}_{ij} = \left\{\exists \ell, \exists k \,|\, S_{i\ell}(k) S_{j\ell}(k) > 0\right\}.$$

Now observe that, since the measurement graph is weakly connected, then the communication graph is a connected undirected graph. This fact together with the fact the diagonal elements of $Q^{(i)}_{11}$ are all positive for any $i \in \{1, \dots, N\}$ implies that there exists almost surely $\bar{k}$ such that, for all $k' \geq \bar{k}$, all the elements of the matrix $S_{11}(k')$ are strictly greater than 0. Assume now, without loss of generality, that $\sigma(k') = i$, for $k' \geq k$. Then, since the i-th row of $S_{21}(k'+1)$ is equal to $e_i e_i^T S_{11}(k')$, it turns out that, all the elements of the i-th row of $S_{21}(k'+1)$ are strictly greater than 0. Moreover, it is easy to see that they will remain stricty greater than 0 also for any $k'' \geq k'$. Hence we can argue that, there exists almost surely, also a $\bar{k}'$ such that for all $k' \geq \bar{k}'$, all the elements of the matrix $S_{21}(k')$ are strictly greater than 0. It follows that the property stated in (12) is satisfied for any $k \geq \bar{k}'$ and for any $\ell \in \{1, \dots, N\}$. This concludes the proof of item (i).

Concerning item (ii), we again resort to the results in [22]. Let $\Omega = I - 1/2N \mathbf{1}\mathbf{1}^T$ where in this expression we assume that I is the $2N$-dimensional identity matrix and the vector $\mathbf{1}$ is $2N$-dimensional. From the results in [22], it follows that to study the rate of convergence of $\mathbb{E}\left[\|\xi(k) - \alpha\mathbf{1}\|^2\right]$ is equivalent to study the convergence rate of $\mathbb{E}\|\Omega\xi(k)\|^2$ and in particular of the linear recursive system

$$\Delta(t+1) = \mathbb{E}\left[Q^T_{\sigma(0)} \Delta(t) Q_{\sigma(0)}\right]$$

where $\Delta(0) = \Omega$. Observe that $\Delta(t)$ is the evolution of a linear dynamical system which can be written in the form

$$\Delta(t+1) = \mathcal{L}(\Delta(t))$$

where $\mathcal{L} : \mathbb{R}^{2N \times 2N} \to \mathbb{R}^{2N \times 2N}$ is given by

$$\mathcal{L}(M) = \mathbb{E}\left[Q^T_{\sigma(0)} M Q_{\sigma(0)}\right].$$

As highlighted in [22], the linear operator $\mathcal{L}$ can be represented by the matrix $\mathbf{L} = \mathbb{E}[\mathbf{Q}_{\sigma(\mathbf{0})} \otimes \mathbf{Q}_{\sigma(\mathbf{0})}]^{\mathbf{T}}$ where $\otimes$ denotes the Kronecker product of matrices. Following the proof of Proposition 4.3 of [22], one can see that $\mathbf{L^T}$ is a primitive stochastic matrix which, therefore, has the eigenvalue 1 with algebraic multiplicity 1. Moreover, $\mathbf{L^T}(\mathbf{1} \otimes \mathbf{1}) = (\mathbf{1} \otimes \mathbf{1})$ and $(\mathbf{1} \otimes \mathbf{1})(\Omega \otimes \Omega) = 0$, from which it follows that $\mathbb{E}\|\Omega\xi(k)\|^2 \leq C \rho_{ess}(\mathbf{L^T}) \mathbb{E}\|\mathbf{\Omega}\xi(\mathbf{0})\|^{\mathbf{2}}$ where $\rho_{ess}(\mathbf{L^T})$ denotes the essential spectral radius of $\mathbf{L^T}$. ■

Thanks to a result obtained in [22], the quantities $\mathbb{E}\left[\|\xi - \alpha\mathbf{1}\|^2\right]$ and $\mathbb{E}\left[\|\Omega\xi\|^2\right]$ have the same exponential convergence rate to zero, or, in other words, for any initial condition $\xi(0)$,

$$\limsup_{k\to\infty} \mathbb{E}\left[\|\xi(k) - \alpha\mathbf{1}\|^2\right]^{1/k} = \limsup_{k\to\infty} \mathbb{E}\left[\|\Omega\xi(k)\|^2\right]^{1/k}.$$

For this reason, in what follows we study the right-hand expression, which turns out to be simpler to analyze. In order to have a single figure not dependent on the initial condition, we focus on this worst case exponential rate of convergence

$$R = \sup_{\xi(0)} \limsup_{k\to\infty} \mathbb{E}\left[\|\Omega\xi(k)\|^2\right]^{1/k}$$

It has been proved in Proposition 4.4 of [22] that

$$esr(\bar{Q})^2 \leq R \leq sr(\mathbb{E}(Q_i^T \Omega Q_i)). \tag{13}$$

where $\bar{Q}$ denotes $\mathbb{E}[Q_i] = \sum_{i=1}^N \beta_i Q_i$ and where $sr(\mathbb{E}(Q_i^T \Omega Q_i))$ denotes the spectral radius of semidefinite positive matrix $\mathbb{E}(Q_i^T \Omega Q_i)$, i.e., its largest eigenvalue.

Unfortunately, it turns out from a numerical inspection over significant families of graphs, like Cayley graphs (see [18]), random geometric graphs, that the upper bound $sr(\mathbb{E}(Q_i^T \Omega Q_i))$ is greater than 1, that is, it is not informative for our analysis.

However we have run a number of MonteCarlo simulations randomized over graphs of different topology and size and over different initial conditions, and it always resulted that $\limsup_{k\to\infty} \mathbb{E}\left[\|\Omega\xi(k)\|^2\right]^{1/k} \leq esr(\bar{Q})$. Based on this experimental evidence we formulate the following conjecture.

Conjecture IV.4 *The quantity $esr(\bar{Q})$ is an upper bound for the exponential convergence rate R, i.e.,*

$$R \leq esr(\bar{Q})$$

The above conjecture and the fact that $esr(\bar{Q})^2 \leq R$ motivates to study $esr(\bar{Q})$.

V. Convergence Rate Analysis of a-CL algorithm for regular graphs

In this section we assume that the measurements graph $\mathcal{G} = (V, \mathcal{E})$ is a connected undirected regular graph[2] such that, for $i \in \{1, \dots, N\}$, $|\mathcal{N}_i| = \nu$. In other words ν represents the degree of the graph. Moreover we assume the following properties.

Assumption V.1 *We have that*

(i) *the error measurements covariances are all identical, i.e., $R_{ij} = R$ for all $(i, j) \in \mathcal{E}$;*
(ii) *$\epsilon = R/(2(\nu + 1))$;*
(iii) *the probabilities $\{\beta_1, \dots, \beta_N\}$ are uniform, i.e., $\beta_1 = \dots = \beta_N = 1/N$.*

Observe that, from properties (i) and (ii) of the above assumption, it turns out that the matrix $P = I - \epsilon A^T R^{-1} A$, associated to the s-CL algorithm, is a symmetric matrix such that $P_{ij} = 1/(\nu + 1)$ for all $j \in \mathcal{N}_i \cup \{i\}$.

Let $\lambda_1(P) = 1 > \lambda_2(P) \geq \dots \geq \lambda_N(P)$ the eigenvalues of P. Note that $esr(P) = \max\{|\lambda_2(p)|, |\lambda_N(P)|\}$. The goal is to show how the $2N$ eigenvalues of $\bar{Q}$ are related to those of P.

Lemma V.2 *Consider the a-CL algorithm running over an undirected regular graph $\mathcal{G} = (V, \mathcal{E})$ such that, for $i \in \{1, \dots, N\}$, $|\mathcal{N}_i| = \nu$. Assume Assumption V.1 holds true. Then the $2N$ eigenvalues of $\bar{Q}$ are the solutions of the following N second-order equations*

$$f(s, \lambda_i, N, \nu) = s^2 + (a + b)s + (ab + c) \tag{14}$$

where

$$a = -\left[\frac{N - \nu}{N} + \frac{\lambda_i}{N} + \frac{\nu - 1}{N(\nu + 1)}\right]$$
$$b = -\frac{N - 1}{N}$$
$$c = -\frac{\nu - 1}{N^2}(\lambda_i - \frac{1}{\nu + 1})$$

The proof is provided in appendix A. Now let $s_1^{(i)}$ and $s_2^{(i)}$ denote the two solutions of $f(s, \lambda_i, N, \nu)$. It easy to see that $s_1^{(1)} = 1$ and $s_2^{(1)} = 1 - \frac{\nu^2+1}{N(\nu+1)}$. The following result restricts the search of $esr(\bar{Q})$ among the values $|s_1^{(2)}|$ and $|s_2^{(2)}|$ and $1 - \frac{\nu^2+1}{N(\nu+1)}$.

[2]An undirected graph is said to be regular if all the nodes have the same number of neighbors

Theorem V.3 . *Consider the a-CL algorithm running on an undirected regular graph $\mathcal{G} = (V, \mathcal{E})$ such that, for $i \in \{1, \dots, N\}$, $|\mathcal{N}_i| = \nu$. Assume Assumption V.1 holds true. Moreover let $\gamma^* = \frac{\nu - 1 + N(\nu+1) - \sqrt{N^2(\nu+1)^2 - 2N(\nu^3+\nu+2) + (\nu-1)^2 + (\nu^2+1)^2}}{\nu+1}$ then*

(i) *if $1 - esr(P) \leq \gamma^* \Longrightarrow esr(\bar{Q}) = \max(|s_1^{(2)}|, |s_2^{(2)}|)$;*
(ii) *if $1 - esr(P) > \gamma^* \Longrightarrow esr(\bar{Q}) = s_2^{(1)} = 1 - \frac{\nu^2+1}{N(\nu+1)}$.*

The proof is reported in appendix B.

A. Asymptotic Behavior

Consider now a sequence of connected undirected regular graphs $\mathcal{G}_N$ of increasing size N, and fixed degree ν. Assume Assumption V.1 holds true for any $\mathcal{G}_N$. Then to any $\mathcal{G}_N$ we can associate a stochastic matrix P_N describing a s-CL algorithm, such that $(P_N)_{ij} = 1/(\nu + 1)$ for all $j \in \mathcal{N}_i \cup \{i\}$. In the following we analyze three different scenarios concerning three different spectral properties.

1) Cayley graphs: The following assumption holds true for family of graphs like the Cayley graphs and d-dimensional toruses. Moreover, it well describes the spectral behavior of Random Geometric graphs which is a family of graphs frequently used in practice.

Assumption V.4 *Consider the sequence of matrices P_N associated to the sequence of graphs $\mathcal{G}_N$ above described. Then we have*

$$esr\left(P_N\right) = 1 - \varepsilon(N) + o(\varepsilon(N)) \tag{15}$$

where $\varepsilon : \mathbb{N} \to \mathbb{R}$ is a positive function such that $\varepsilon(N) \to 0$ as $N \to \infty$. Let the matrix $\bar{Q}_N$ represent the average matrix associated to the a-CL algorithm running over $\mathcal{G}_N$. The following result characterizes the asymptotic behavior of $esr\left(\bar{Q}_N\right)$, with the respect to $esr\left(P_N\right)$.

Theorem V.5 *Consider the sequence of graphs $\mathcal{G}_N$ above described. Consider the a-CL algorithm running over $\mathcal{G}_N$. Assume Assumption V.1 and Assumption V.4 hold true. Then*

$$esr\left(\bar{Q}_N\right) = 1 - \frac{\nu(\nu + 1)}{N(\nu^2 + 1)}\varepsilon(N) + o\left(\frac{\varepsilon(N)}{N}\right)$$

The proof can be found in appendix C.

2) Ramanujan graphs: A similar analysis can be provided also for other relevant families of regular graphs like the Ramanujan graphs [23]. Let us recall the asymptotic lower bound proved by Alon and Boppana for doubly stochastic matrices built over ν-regular bidirected graphs. Specifically, if A denotes the adjacency matrix of a ν-regular bidirected graph, let P be the doubly stochastic matrix defined as $P = \nu^{-1}A$, then

$$\liminf_{N \to \infty} \rho_{\text{ess}}(P) \geq \frac{2\sqrt{\nu - 1}}{\nu}$$

where the lim inf is taken along the family of all ν-regular bidirected graphs having N vertices. Ramanujan graphs

are those ν-regular bidirected graphs which achieves the previous bound, i.e., such that $\rho_{\text{ess}}(P) = \frac{2\sqrt{\nu-1}}{\nu}$. Hence through the Ramanujan graphs it is possible to keep the essential spectral radius bounded away from 1, while keeping the degree fixed.

Theorem V.6 *Consider a sequence of Ramanujan graphs $\mathcal{G}_N$ of increasing size N and fixed degree ν. Consider the a-CL algorithm running over $\mathcal{G}_N$. Assume assumption V.1 holds true. Then, it holds*

$$\rho_{ess}(\bar{Q}_N) = 1 - \alpha(\nu)/N$$

where, respectively

$$\alpha(\nu) = \left[\frac{(\nu^2+1)+(1-\epsilon)(\nu+1)}{2(\nu+1)}\left(1 - \sqrt{1 - \frac{4\nu(1-\epsilon)(\nu+1)^2}{((\nu^2+1)+(1-\epsilon)(\nu+1))^2}}\right)\right]$$

$$\epsilon = \frac{2}{\nu}\sqrt{\nu-1}$$

The proof can be found in Appendix D.

3) Complete graphs: Concerning the family of complete graphs we have that $\rho_{\text{ess}}(P) = 0$ and $\nu = N - 1$ so, we can state the following

Corollary V.7 *Consider a sequence of Complete graphs $\mathcal{G}_N$ of increasing size N. Consider the a-CL algorithm running over $\mathcal{G}_N$. Assume assumption V.1 holds true. Then, it holds*

$$\rho_{ess}(\bar{Q}_N) = \frac{2(N-1)}{N}$$

The proof can be found in Appendix E.

VI. Numerical Results

In this section we firstly provide some simulations supporting the exponential rate of convergence of the different algorithms implemented. Secondly some simulations to support conjecture IV.4.

Example VI.1 In this example we consider a random geometric graph generated by choosing $N = 100$ points randomly placed in the interval $[0, 10]$, and pairs of nodes that are within a range of 3 took measurements of each others' relative positions, namely, both measurements z_{ij} and z_{ji} are available provided that $|x_i - x_j| \leq 3$. Every measurement was corrupted by Gaussian noise with unitary covariance. In this example we show the performance of a-CL with both the uniform persistent, assuming that each node transmitted at least once within each time interval $[k, k+\tau)$ where $\tau = 200$, and the random persistent communicating network.

In Figure 1 we plotted the behavior of the error

$$J(k) = \log\left(\|A(\hat{x}(k) - x^*)\|\right)$$

The plot reported is the result of the average over 1000 Monte Carlo runs, randomized with respect to both the measurement graph[3] and the initial conditions. Observe that the trajectories converge to 0 exponentially.

[3]In perfoming our average we kept only the random geometric graphs which resulted to be connected.

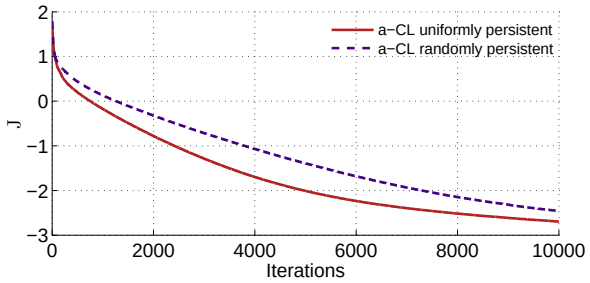


Fig. 1: Behavior of J for a uniformly persistent (red solid line) and a randomly persistent (blue dashed line) communicating network.

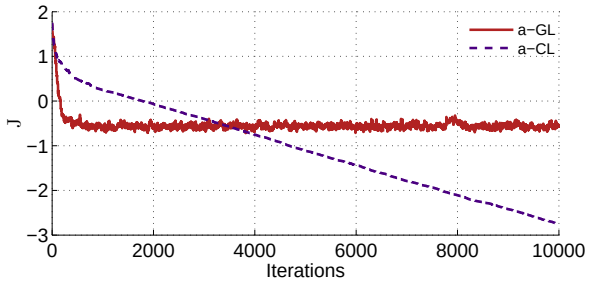


Fig. 2: Comparison between the a-CL algorithm and the randomized algorithm (a-GL) proposed in [24].

Example VI.2 In this example we provide a numerical comparison with the randomized algorithm proposed in [24], later on called a-GL, which, for the sake of the completeness, we shortly review next. Similarly to the a-CL algorithm, also during the k-th iteration in the randomized algorithm in [24], there is only one node, say node i, which transmits its variable $\hat{x}_i$ to all its neighbors. For $j \in \mathcal{N}_i$, node j, based on the information received from node i, performs the following update

$$\begin{aligned}\hat{x}_j(k+1) &= 1/2\left(\hat{x}_j(k) + \hat{x}_i(k) + 1/2(z_{ji} - z_{ij})\right) \\ &= \hat{x}_j(k) + 1/2\left(\hat{x}_j(k) - \hat{x}_i(k) + 1/2(z_{ji} - z_{ij})\right)\end{aligned} \tag{16}$$

while for $h \notin \mathcal{N}_i$, $\hat{x}_h(k+1) = \hat{x}_h(k)$. In this example we consider random geometric measurement graphs $\mathcal{G}$ built as in the previous example. For both strategies the transmitting node at each iteration is chosen with a uniform probability. In Figure 2 we plotted the behavior of J for both the a-CL algorithm (blue dashed line) and the a-GL proposed in [24] (red solid line).

As stressed in [24] the trajectory $\hat{x}(k)$ generated by the updating rule in (16) does not converge to x^* but oscillates within a neighborhood of x^*. This fact appears to be evident also in Figure 2. However, of note is the fact that the strategy in [24] seems to have a faster transient than the randomized a-CL algorithm.

Remark VI.3 The authors in [16] proposed an algorithm (denoted as Randomized Kaczmarz Smoothing) similar to

the one introduced in [24]. However also the Randomized Kaczmarz Smoothing algorithm does not converge to the optimal estimate x^* but oscillates within a neighborhood of it.

Remark VI.4 The convergence properties of the randomized algorithm in (16) might be improved by using the following modified updating rule

$$\hat{x}_j(k+1) = \hat{x}_j(k) + \gamma(k)\left(\hat{x}_j(k) - \hat{x}_i(k) + 1/2(z_{ji} - z_{ij})\right)$$

where $\{\gamma(k)\}$ is a sequence of receding step-sizes with $\gamma(k) \in (0,1)$ and $\lim_{k\to\infty}\gamma(k) = 0$. A similar idea has been proposed in [16] introducing the Randomized Kaczmarz algorithm with Under-Relaxation. However, in this case, since $\lim_{k\to\infty}\gamma(k) = 0$ and, in turn, the applied control action becomes smaller and smaller, the speed of convergence is not exponential but just sub-linear.

Example VI.5 In this example we consider a set of Cayley graphs $\mathcal{G}$ of increasing size N. What we show is a comparison between the empirical rate of convergence of the algorithm, its lower bound, represented by $esr(\bar{Q})^2$, and the $esr(\bar{Q})$. As a matter of fact, figure 3 shows that the rate is upper bounded by $esr(\bar{Q})$.

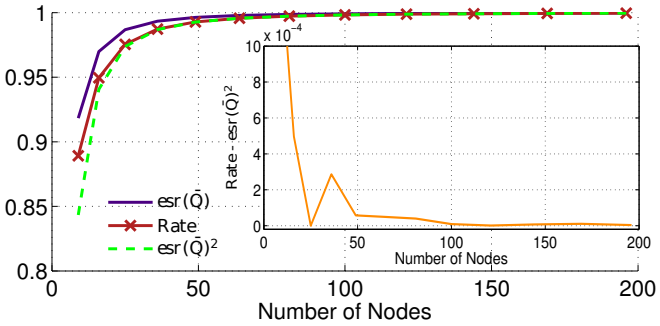


Fig. 3: Trend of the rate of convergence, of the $esr(\bar{Q})^2$ and of the $esr(\bar{Q})$ for Cayley graphs of increasing size N.

Example VI.6 In this final example we consider the family of random geometric graphs generated by choosing N nodes in the interval [0,1] and connecting those whose relative euclidean distance is less than 0.3. This meaning that node i and j perform the measurement z_{ij} and z_{ji} if $|x_i - x_j| \leq 0.3$. Figure 4 shows that, for random geometric graphs as well as for Cayley, the rate is upper bounded by $esr\bar{Q}$.

Remark VI.7 Note that, in both example presented, and in all the several simulations performed, the lower bound, $esr(\bar{Q})^2$, better approximates the behavior of the empirical rate of convergence.

VII. Conclusion

In this paper we considered the problem of optimal estimating the position of each agent in a network from relative noisy vectorial distances with its neighbors. We

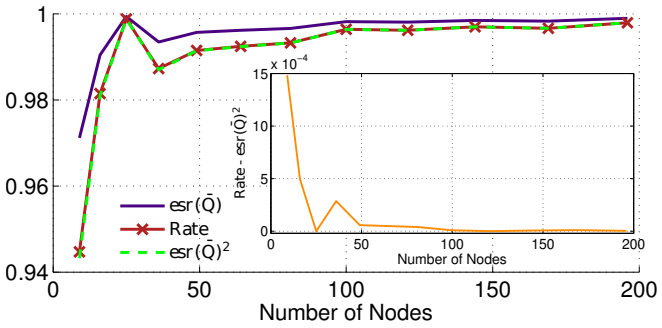


Fig. 4: Trend of the rate of convergence, of the $esr(\bar{Q})^2$ and of the $esr(\bar{Q})$ for random geometric graphs of increasing size N.

formulated the problem as a classical weighted least-squares problem and we reviewed its optimal centralized solution. The contribution we proposed is twofold. We first introduced a synchronous consensus-based algorithm that is guaranteed to have exponentially convergence rate to the optimal solution. We then extended this algorithm to more realistic asynchronous implementations via the use of local memory variables. We showed that exponential convergence is still guaranteed under both uniform persistent communication and randomized persistent communication protocols.

Moreover we studied the rate of convergence of the proposed randomized broadcast asynchronous algorithm. We provided closed form expressions for the rate in terms of convergence in expectation for regular graphs. We also conjectured based on extensive simulations that this provides also lower and upper bounds for the rate of convergence in mean square. We finally showed that asymptotically in the number of nodes N, this rate of convergence scales as standard memoryless consensus algorithms, thus implying a fast convergence rate.

Future work is directed in proving the conjecture that the rate of convergence in expectation can provide upper and lower bounds for the rate of convergence in expectation, and in providing extensive comparisons with the most popular alternative algorithms available in the literature to solve the same problem.

Appendix

A. Proof of Lemma V.2

Recalling equation (8), it is possible to write the average matrix $\bar{Q} = \mathbb{E}[Q_i]$ as

$$\bar{Q} = \begin{bmatrix} \bar{Q}_{11} & \bar{Q}_{12} \\ \bar{Q}_{21} & \bar{Q}_{22} \end{bmatrix}$$

where

$$\begin{aligned}
\bar{Q}_{11} &= \begin{bmatrix} \frac{N-|\mathcal{N}_1|}{N} & & \\ & \ddots & \\ & & \frac{N-|\mathcal{N}_N|}{N} \end{bmatrix} I + \frac{1}{N}P \\
&+ \begin{bmatrix} \frac{|\mathcal{N}_1|-1}{N} & & \\ & \ddots & \\ & & \frac{|\mathcal{N}_N|-1}{N} \end{bmatrix} diag(P) \\
\bar{Q}_{12} &= \begin{bmatrix} \frac{|\mathcal{N}_1|-1}{N} & & \\ & \ddots & \\ & & \frac{|\mathcal{N}_N|-1}{N} \end{bmatrix} (P - diag(P)) \\
\bar{Q}_{21} &= \frac{1}{N} I \\
\bar{Q}_{22} &= \frac{N-1}{N} I
\end{aligned}$$

Thanks to assumption (i) we have that

$$\mathcal{N}_i = \nu \qquad \forall i \in \{1, \dots, N\}$$

therefore we can write $\bar{Q}$ in a more compact form, where

$$\begin{aligned}
\bar{Q}_{11} &= \frac{N-\nu}{N} I + \frac{1}{N} P + \frac{\nu-1}{N(\nu+1)} I \\
\bar{Q}_{12} &= \frac{\nu-1}{N} (P - \frac{1}{\nu+1} I) \\
\bar{Q}_{21} &= \frac{1}{N} I \\
\bar{Q}_{22} &= \frac{N-1}{N} I
\end{aligned}$$

Since P and I are simultaneously diagonalizable, with an appropriate permutation is possible to obtain a block diagonal matrix where every block, of dimension two, is equal to

$$\begin{bmatrix} \frac{N-\nu}{N} + \frac{\lambda_i}{N} + \frac{\nu-1}{N(\nu+1)} & \frac{\nu-1}{N}(\lambda_i - \frac{1}{\nu+1}) \\ \frac{1}{N} & \frac{N-1}{N} \end{bmatrix}$$

where λ_i represents the generic eigenvalue of P. Thanks to [25] the matrix entire determinant can be obtain as the multilpcation of the single block determinant whose associated characteristic polynomial is equal to:

$$\begin{aligned}
f(s, \lambda_i, N, \nu) &= s^2 - \Big(\frac{N-\nu}{N} + \frac{\lambda_i}{N} + \frac{\nu-1}{N(\nu+1)} + \\
&+ \frac{N-1}{N} \Big) s + \\
&+ \frac{N-1}{N} \left(\frac{N-\nu}{N} + \frac{\lambda_i}{N} + \frac{\nu-1}{N(\nu+1)} \right) - \\
&- \frac{\nu-1}{N^2} \left(\lambda_i - \frac{1}{\nu+1} \right) \qquad (17)
\end{aligned}$$

Thanks to (17) it can be seen how every eigenvalue of P is mapped in two eigenvalues of $\bar{Q}$.

B. Proof of Theorem V.3

The proof of the theorem arises from a collection of preliminary results that we are now going to show. It is known that, being P a stochastic matrix, all its eigenvalues can be written like $\lambda_i = 1 - \gamma_i$, where $\gamma_i \in [0, 2]$. Now rewriting the second order equation (14) in two parts, one depending from γ_i and one not depending from it, we obtain:

$$f(s, \lambda_i, N, \nu) = d(s, N, \nu) + \gamma_i n(s, N, \nu) \qquad (18)$$

Note that, this can be considered as a "quantized" root locus, where the gain is given by the N values of γ_i. For sake of simplicity we will study its continuous time version in order to understand which are the possible eigenvalues of $\bar{Q}$.

Remark .1 It is known that the eigenvalues are discrete but, to the end of exploiting the root locus technique, we will consider their "continuous" version. We want to highlight that even if we have infinite values of γ_i in the root locus, we are interested in just 2N points on it.

As shown in equation (18) the two polynomials involved in the locus are

$$\begin{aligned}
d(s, N, \nu) &= s^2 - \frac{2N(\nu+1) - (\nu^2+1)}{N(\nu+1)} s + 1 - \frac{\nu^2+1}{N(\nu+1)} \\
n(s, N, \nu) &= \frac{s}{N} + \frac{\nu - N}{N^2}
\end{aligned}$$

whose roots are $d_1 = 1$, $d_2 = 1 - \frac{\nu^2+1}{N(\nu+1)}$ and $n_0 = 1 - \frac{\nu}{N}$. We recall from standard results that d_1 and d_2 are the starting point of the root locus[4], and n_0 its ending point, respectively for $\gamma = 0$ and $\gamma = \infty$. Additionally, the double points of the locus are

$$dp_{+,-} = 1 - \frac{1}{N} \left(\nu \pm \sqrt{\frac{\nu(\nu-1)}{\nu+1}} \right)$$

It is easy to see that

$$0 < n_0 \leq d_2 \leq d_1$$

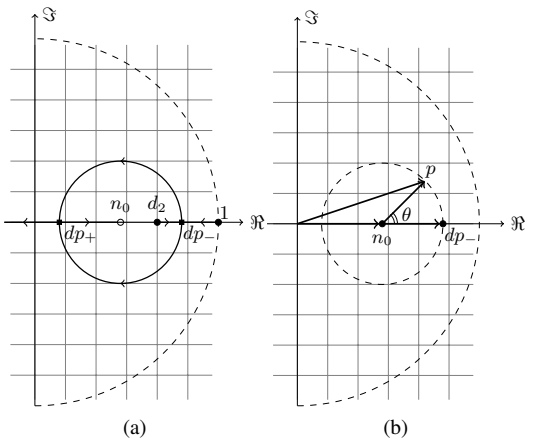
indeed, the first inequality is trivial being $\nu < N$, while the second leads to

$$\frac{N-\nu}{N} - 1 + \frac{\nu^2+1}{N(\nu+1)} \leq 0 \Rightarrow \nu \geq 1$$

Moreover is straightforward to see that n_0 is the mid point of dp_+ and dp_- being

$$\begin{aligned}
n_0 =& 1 - \frac{\nu}{N} = \frac{dp_- + dp_+}{2} = \\
=& \frac{1}{2} \left(1 - \frac{1}{N} \left(\nu - \sqrt{\frac{\nu(\nu-1)}{\nu+1}} \right) \right) + \\
&+ \frac{1}{2} \left(1 - \frac{1}{N} \left(\nu + \sqrt{\frac{\nu(\nu-1)}{\nu+1}} \right) \right)
\end{aligned}$$

[4] note that they coincide with the solution of equation (14), previously called $s_1^{(1)}$ and $s_2^{(1)}$



(a) (b)

Fig. 5: (a) Root Locus of $d(s) + \epsilon_i n(s) = 0$. (b) The figure shows that the length of the vector between the points 0 and dp_- is longer than the vector formed by the points 0, n_0 and p

It follows that the double points are centred on n_0 and that dp_- lays between d_1 and d_2.

Let us firstly show the following two corollaries.

Corollary .2 *Given a root locus function of the form*[5] $(s - p_1)(s - p_2) + K(s - z) = 0, \quad K \in [0, +\infty)$*, the complex roots lay on a circle centred on the* z*.*

Proof: The roots of $(s - p_1)(s - p_2) + K(s - z) = 0$ are:

$$s_{1,2} = \frac{(p_1 + p_2 - K) \pm \sqrt{(p_1 + p_2 - K)^2 - 4(p_1 p_2 - Kz)}}{2}$$

whose real and imaginary part, considering the complex case that holds when $(p_1 + p_2 - K)^2 - 4(p_1 p_2 - Kz) < 0$, are:

$$\Re(s_{1,2}) = \frac{p_1 + p_2 - K}{2}$$

$$\Im(s_{1,2}) = \pm\frac{\sqrt{(p_1 + p_2 - K)^2 - 4(p_1 p_2 - Kz)}}{2}$$

The square of the distance of these points from z can be computed as

$$(\Re(s_{1,2}) - z)^2 + (\Im(s_{1,2}) - 0)^2 = z^2 - zp_1 - zp_2 + p_1 p_2$$

that is independent on K and so is a constant. ■

Thank to corollary (.2), is easy to get that the absolute value of the complex roots is descending when γ_i increase.

In particular every pair of complex roots has module smaller that dp_-, indeed, being $|dp_-| - |n_0|$ the radius of the circonference centred in n_0, see figure 5b, it follows that

$$\begin{aligned} |p|^2 &= (|n_0| + (|dp_-| - |n_0|)\cos\theta)^2 + \\ &+ ((|dp_-| - |n_0|)\sin\theta)^2 \\ &= |dp_-|^2 + 2|n_0|(1 - \cos\theta)(|n_0| - |dp_-|) \leq |dp_-|^2 \end{aligned}$$

since $(1 - \cos\theta) \geq 0$ and $(|n_0| - |dp_-|) \leq 0$. Where θ is the angle of the vector $p - n_0$ wrt the real axis.

Corollary .3 *Consider polynomial in equation (18). For* $N > 1$ *and* $\nu \in \{1, \ldots, N-1\}$ *then* $\forall\ \gamma_i \in [0, 2]$*, the real parts of* $s_{1,2}^{(i)}$*, solution of* $f(s, \lambda_i, N, \nu)$*, are non negative.*

Proof: The generic solution of the characteristic polynomial (18) are equal to $s_{1,2}^{(i)} = \frac{1}{2}\left(\frac{2N(\nu+1)-(\nu^2+1)}{N(\nu+1)} - \frac{\gamma_i}{N}\right) \pm \frac{1}{2}\sqrt{\left(\frac{2N(\nu+1)-(\nu^2+1)}{N(\nu+1)} - \frac{\gamma_i}{N}\right)^2 - 4\left(1 - \frac{\nu^2+1}{N(\nu+1)} + \gamma_i\frac{\nu-N}{N^2}\right)}$ whose discriminant could be $\Delta \lesseqgtr 0$. Let us firstly consider $\Delta \geq 0$. In this case the value of γ_i such that $f(s, 1-\gamma_i, N, \nu)$ has one root equal to zero is

$$\gamma_i = \frac{N(\nu^2 + 1 - N\nu - N)}{(\nu + 1)(\nu - N)}$$

Imposing $\gamma_i \geq 2$, after some algebra, one gets

$$(N - 2)\nu^2 + (2N - 2 - N2)\nu + (3N - N^2) \leq 0$$

which is always satisfied for $N > 1$ and $\nu \in \{1, \ldots, N-1\}$.

Consider now, the second case with $\Delta < 0$. It is easy to see that the two pure imaginary roots of $f(s, 1-\gamma_i, N, \nu)$ are reached with a value of $\gamma_i \geq 2$. Indeed, imposing $\Re(s_{1,2}^{(i)}) = 0$ ones get

$$\gamma_i = 2N - \frac{\nu^2 + 1}{\nu + 1}$$

which, $\forall N \geq 2$, is a descending function of ν and assumes its minimum value in $\nu = 1$ which is

$$\gamma_i = 2N - 1 \geq 2$$

■

Thanks to corollary .2 and .3 we know that, $\forall\lambda_i(P)$, all the eigenvalues of $\bar{Q}$ have positive real part. Moreover analyzing the root locus form it is easy to understand that three main situations can occur, that are:

(i) the value of γ_i corresponding to $esr(P)$ is sufficiently small to individuate, on the root locus, two eigenvalue of $\bar{Q}$ that lie on the real axis between 1 and d_2. In this case the greater correspond to the $esr(\bar{Q})$;

(ii) the value of γ_i of the $esr(P)$ individuates, on the locus, two complex conjugate values that has module greater than d_2. In this case, again, the $esr(\bar{Q})$ correspond to them;

(iii) γ_i individuates values that have module smaller than d_2. In this case is obvious that d_2 becomes the $esr(\bar{Q})$

This is enough to conclude that, calling γ^* the value that individuates two complex conjugate roots on the locus, say $\mu_{1,2}$, such that

$$|\mu_1| = |\mu_2| = |d_2|$$

then, if $\gamma < \gamma^*$ we get that the $esr(P)$ maps the $esr(\bar{Q})$; otherwise the $esr(\bar{Q})$ is equal to d_2. The critical value of γ^* is

[5]This kind of equation results from the root locus of a transfer function with one zero and two poles.

$\gamma^* = \frac{\nu-1+N(\nu+1)-\sqrt{N^2(\nu+1)^2-2N(\nu^3+\nu+2)+(\nu-1)^2+(\nu^2+1)^2}}{\nu+1}$

This complete the proof.

C. Proof of Theorem V.5

Proof: Let us define the characteristic polinomial that identify the $2N$ eigenvalues of $\bar{Q}_N$ wrt those of P_N, denoted as $1-\gamma_i$, as

$$f(s,\lambda_i,N,\nu) = d(s,N,\nu) + \gamma_i n(s,N,\nu) \triangleq g(s,\gamma_i,N,\nu)$$

where g explicity explicitly shown the dependance on γ_i and where

$$d(s,N,\nu) = s^2 - \frac{2N(\nu+1)-(\nu^2+1)}{N(\nu+1)}s + 1 - \frac{\nu^2+1}{N(\nu+1)}$$
$$n(s,N,\nu) = \frac{s}{N} + \frac{\nu-N}{N^2}$$

It can be easily shown that the $sr(P_N)$ maps the $sr(\bar{Q}_N)$, i.e choosing $\gamma_i = 0$ one of the two solution of g is $s = 1$. Thank to assumption V.4 we have that

$$esr(P_N) = 1 - \varepsilon(N) + o(\varepsilon(N))$$

and since

$$\begin{cases} g(1,0,N,\nu) = 0 \\ \frac{\partial g}{\partial s}\Big|_{(1,0,N,\nu)} \neq 0 \end{cases}$$

it is possible to exploit the implicit function theorem that let us write, denoting with $s(\gamma_i)$ the eigenvalues of $\bar{Q}_N$ as function of the corresponding γ_i,

$$\begin{aligned} s(\varepsilon(N)+o(\varepsilon(N))) &= 1 - \frac{\partial g}{\partial \gamma_i}/\frac{\partial g}{\partial s}\Big|_{(1,0,N,\nu)} (\epsilon(N)+o(\epsilon(N))) \\ &= 1 - \frac{\nu(\nu+1)}{N(\nu^2+1)}\epsilon(N) + o\left(\frac{\epsilon(N)}{N}\right) \end{aligned}$$

which means that $esr(\bar{Q})$ can be expressed as

$$esr(\bar{Q}) = 1 - \frac{\nu(\nu+1)}{N(\nu^2+1)}\epsilon(N) + o\left(\frac{\epsilon(N)}{N}\right) \qquad (19)$$

■

D. Proof of Theorem V.6

Proof: Recalling that

$$\lim_{N\to\infty} \gamma^*(\nu,N) = \frac{\nu^2+1}{(\nu+1)^2}$$

and being $\nu > 2$, can be easily proved that $1-\rho_{\text{ess}}(\bar{P_N}) < \gamma^*$ so, thanks to Theorem V.3, it follows that $\rho_{\text{ess}}(\bar{Q}_N) = \max(|s_1^{(2)}|, |s_2^{(2)}|)$. Moreover, after straightforward algebraic calculations it is easy to see that $|s_1^{(2)}| > |s_2^{(2)}|$, therefore

$$\rho_{\text{ess}}(\bar{Q}_N) = s_1^{(2)} = 1 - \frac{\alpha(\nu)}{N} \qquad (20)$$

where, respectively
$\alpha = \left\lceil \frac{(\nu^2+1)+(1-\epsilon)(\nu+1)}{2(\nu+1)} \left(1 - \sqrt{1 - \frac{4\nu(1-\epsilon)(\nu+1)^2}{((\nu^2+1)+(1-\epsilon)(\nu+1))^2}}\right)\right\rceil$
$\epsilon = \frac{2}{\nu}\sqrt{\nu-1}$

■

E. Proof of Corollary V.7

Proof: Since all the eigenvalues of P_N different from 1 are equal to 0 the essential spectral radius $\rho_{ess}(P_N) = 0$ and the spectral gap $\gamma_2 = 1$. So it easy to see that

$$1 - \rho_{ess}(P_N) = 1 > \gamma^* = \frac{\nu^2+1}{(\nu+1)^2}$$

this meaning that, for theorem (V.6)

$$\rho_{ess}(\bar{Q}_N) = 1 - \frac{\nu^2+1}{N(\nu+1)} = \frac{2(N-1)}{N^2}$$

where the last equality follows from the fact that for a complete graph $\nu = N-1$. ■

REFERENCES

[1] L. Xiao, S. Boyd, and S. Lall, "A scheme for robust distributed sensor fusion based on average consensus," in *Proccedings of the Information Processing for Sensor Networks (IPSN'05)*, 2005, pp. 63– 70.

[2] S. Bolognani, S. D. Favero, L. Schenato, and D. Varagnolo, "Consensus-based distributed sensor calibration and least-square parameter identification in wsns," *International Journal of Robust and Nonlinear Control*, vol. 20, no. 2, January 2010.

[3] M. Schwager, D. Rus, and J. J. Slotine, "Decentralized, adaptive coverage control for networked robots," *International Journal of Robotics Research*, vol. 28, no. 3, pp. 357–375, March 2009.

[4] B. Johansson, P. Soldati, and M. Johansson, "Mathematical decomposition techniques for distributed cross-layer optimization of data networks," *IEEE Journal on Selected Areas in Communication*, vol. 24, no. 8, pp. 1535–1547, August 2006.

[5] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Mathematical decomposition techniques for distributed cross-layer optimization of data networks," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[6] A. Nedic and A. Ozdaglar, "Distributed subgradient methods for multi-agent optimization," *IEEE Transactions on Automatic Control*, vol. 54, no. 1, pp. 46–61, January 2009.

[7] F. Zanella, D. Varagnolo, A. Cenedese, G. Pillonetto, and L. Schenato, "Newton-raphson consensus for distributed convex optimization," in *IEEE Conference on Decision and Control (CDC 2011)*, 2011, pp. 5917–5922.

[8] J. Wang and N. Elia, "Control approach to distributed optimization," in *Annual Allerton Conference on Communication, Control, and Computing*, 2010, pp. 557–561.

[9] C. Fischione, "Fast-lipschitz optimization with wireless sensor networks applications," *IEEE Transactions on Automatic Control*, vol. 56, no. 10, pp. 2319 – 2331, October 2011.

[10] G. Notarstefano and F. Bullo, "Distributed abstract optimization via constraints consensus: Theory and applications," *IEEE Transactions on Automatic Control*, vol. 56, no. 10, pp. 2247–2261, October 2011.

[11] P. Barooah and J. P. Hespanha, "Distributed estimation from relative measurements in sensor networks," in *Proceedings of the 2nd International Conference on Intelligent Sensing and Information Processing*, Dec. 2005.

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

[13] R. Solis, V. Borkar, and P. R. Kumar, "A new distributed time synchronization protocol for multihop wireless networks," in *45th IEEE Conference on Decision and Control (CDC'06)*, San Diego, December 2006, pp. 2734–2739.

[14] D. Borra, E. Lovisari, R. Carli, F. Fagnani, and S. Zampieri, "Autonomous calibration algorithms for networks of cameras," in *Proceedings of the Americal Control Conference, ACC'12.*, July 2012.

[15] R. Tron and R. Vidal, "Distributed image-based 3-d localization of camera sensor networks," in *Proceedings of the 49th IEEE Conference on Decision and Control CDC'09*, 2009, pp. 901–908.

[16] N. M. Freris and A. Zouzias, "Fast distributed smoothing of relative measurements," in *to appear in IEEE Conference on Decision and Control (CDC 2012)*, 2012.

[17] R. Olfati-Saber, J. Fax, and R. Murray, "Consensus and cooperation in networked multi-agent systems," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 215–913, 2007.

[18] R. Carli, F. Fagnani, A. Speranzon, and S. Zampieri, "Communication constraints in the average consensus problem," vol. 44, no. 3, pp. 671–684, 2008.

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

[20] P. Barooah and J. Hespanha, "Estimation on graphs from relative measurements: Distributed algorithms and fundamental limith," *IEEE Control Systems Magazine*, vol. 27, no. 4, 2007.

[21] A. Nedic and A. Ozdaglar, "Convergence rate for consensus with delays," *Journal of Global Optimization*, vol. 47, no. 3, pp. 437–456, 2008.

[22] F. Fagnani and S. Zampieri, "Randomized consensus algorithms over large scale networks," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 4, pp. 634–649, 2008.

[23] M. R. Murthy, "Ramanujan graphs," *Journal of Ramanujan Mathematical society*, vol. 18, 2003.

[24] S. Bolognani, S. D. Favero, L. Schenato, and D. Varagnolo, "Consensus-based distributed sensor calibration and least-square parameter identification in wsns," *International Journal of Robust and Nonlinear Control*, vol. 20, no. 2, pp. 176–193, 2010.

[25] J. Silvester, "Determinants of block matrices," *The Mathematical Gazette*, vol. 84, no. 501, pp. 460–467, 2000.