A Proximal Point Approach for Distributed System State Estimation

Marco Fabris * **Giulia Michieletto** * **Angelo Cenedese** *

* *Dept. of Information Engineering, University of Padova, Italy*
(corresponding author e-mail: marco.fabris.7@phd.unipd.it)

Abstract: System state estimation constitutes a key problem in several applications involving multi-agent system architectures. This rests upon the estimation of the state of each agent in the group, which is supposed to access only relative measurements w.r.t. some neighbors state. Exploiting the standard least-squares paradigm, the system state estimation task is faced in this work by deriving a distributed Proximal Point-based iterative scheme. This solution entails the emergence of interesting connections between the structural properties of the stochastic matrices describing the system dynamics and the convergence behavior toward the optimal estimate. A deep analysis of such relations is provided, jointly with a further discussion on the penalty parameter that characterizes the Proximal Point approach.

Keywords: state estimation, convergence analysis, dynamic systems, network topologies

1. INTRODUCTION

In the last twenty years, the advancements in pervasive computing and ambient intelligence have led to a fast development of smart networks characterized by the presence of various interacting cyber-physical systems (as, e.g., smart grids and robotic networks) whose effectiveness rests upon the cooperation of multiple autonomous entities under constraints of real-time high-level performance. In this perspective, the role of multi-agent systems is evident, due to the capability of these architectures involving large sets of interactive devices to solve complex tasks by exploiting local computation and communication (Chen et al., 2019).

One of the main challenging aspects of the embedded multi-agent systems consists in the wise management and exploitation of the local information since they are typically involved in the solution of large-scale logically/spatially distributed optimization problems. Along this line, *system state estimation* constitutes a popular task in several control and learning smart networks applications, requiring to estimate the state of each agent through the adoption of a distributed paradigm exploiting local information and interactions (see, e.g., (Cintuglu and Ishchenko, 2019) wherein the task is faced accounting for network microgrids).

Many existing works explore the properties of the distributed state estimation solution w.r.t. the topological features of the graph modeling the network in terms of available measurements. For instance, in (Barooah and Hespanha, 2009), the variance of the estimation error for a node variable is shown to increase proportionally to the distance of the node itself w.r.t an arbitrary reference one, hence a topological classification is provided resting upon the graph sparsity. In (Carron et al., 2014), instead, the system state problem boils downs to the agent position es-

timation exploiting some relative noisy distances measurement. This task is tackled through a consensus-based algorithm whose rate of convergence in expectation of the estimation error is determined for regular graphs, accounting for the network size. Recently, in (Ravazzi et al., 2018), the problem of distributed estimation is studied accounting for relative measurements characterized by heterogeneity and uncertainty. In this case a Least-Square (LS) approach is adopted. Similarly, in (Fabris et al., 2019), the problem is solved proposing a decentralize iterative solution whose convergence features are strictly related to the spectral properties of the update matrix.

The major contribution of this work consists in the design of a distributed iterative procedure for system state estimation. Its novelty relies on the use of Proximal Point (PP) method, nowadays widely employed in different research areas (such as statistical learning (Bai et al., 2019)) and recently exploited in the development of algorithms for the solution of estimation tasks resting upon the LS approach (Aster et al., 2018). A deep insight is provided on the role of topological links that define the communication constraints in the system w.r.t. the convergence properties of the proposed method, along with a convergence analysis that guarantees the scalability of the proposed solution. Moreover, a discussion on the role of the penalty parameter – as usual in many related works, see also (Olfati-Saber et al., 2007) – characterizing the PP approach is proposed.

The rest of the paper is organized as follows. In Sec. 2 the system state estimation is formally stated as a minimization problem. In Sec. 3 its distributed PP-based solution is derived, while Sec. 4 is devoted to the convergence analysis of the proposed solution. The theoretical observations are then validated via numerical simulations for different network topologies in Sec. 5. Finally, in Sec. 6 conclusions are drawn and future work are envisaged.

2. SYSTEM STATE ESTIMATION PROBLEM

This section is devoted to the formalization of the system state estimation task as an optimization problem based on the set of the available relative measurements. In doing this, a graph-based representation is introduced for multi-agent systems that model the devices interactions.

2.1 Graph-Based Network Model

An n-agent system can be modeled through a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ so that each element in the vertex set $\mathcal{V} = \{v_1, \ldots, v_n\}$ is related to an agent in the group, while the edge set $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ characterizes the agents interactions in terms of both sensing and communication capabilities. Hereafter, bidirectional agents interactions are supposed, therefore $\mathcal{G}$ is assumed to be undirected and such that $e_{ij} = \{v_i, v_j\} \in \mathcal{E}$ if and only if the i-th and j-th agents can sense each other and are able to mutually exchange information according to some communication protocol.

Agent links are often described by the *adjacency matrix* $\mathbf{A} \in \mathbb{R}^{n \times n}$ such that $[\mathbf{A}]_{ij} = 1$ if $e_{ij} \in \mathcal{E}$ and $[\mathbf{A}]_{ij} = 0$ otherwise. The set $\mathcal{N}_i = \{v_j \in \mathcal{V}, j \neq i \mid [\mathbf{A}]_{ij} = 1\}$ thus identifies the *neighborhood* of the node v_i, i.e., the set of agents interacting with the i-th one. The cardinality $\deg(v_i)$ of $\mathcal{N}_i$ is generally referred to as *degree* of the i-th agent, and $d_m = \min_{i \in \{1, \ldots, n\}} \{\deg(v_i)\}$ and $d_M = \max_{i \in \{1, \ldots, n\}} \{\deg(v_i)\}$ respectively identify the *minimum* and *maximum* degree of the graph. Furthermore, the degree of the i-th agent corresponds to the i-th element of the (diagonal) *degree matrix* $\mathbf{D} = \mathrm{diag}(\mathbf{A}\mathbb{1}_n) \in \mathbb{R}^{n \times n}$, where $\mathbb{1}_n \in \mathbb{R}^n$ indicates a n-dimensional (column) vector whose entries are all ones. In particular, matrix $\mathbf{D}$ contributes to the definition of the *Laplacian matrix*, $\mathbf{L} = \mathbf{D} - \mathbf{A} \in \mathbb{R}^{n \times n}$, that summarizes the topology of the networked system.

2.2 Problem Statement

According to the introduced graph-based model, let us consider a multi-agent system composed of n devices whose interactions are represented by an undirected *connected* graph $\mathcal{G}$. Each i-th agent, $i \in \{1, \ldots, n\}$, in the network is assumed to be characterized by a scalar attribute $x_i \in \mathbb{R}$ (the *i-th agent state*) corresponding to a physical quantity, and by a set of relative measurements of the same quantity, namely $\{\tilde{x}_{ij} \in \mathbb{R}, \ v_j \in \mathcal{N}_i\}$ where $\tilde{x}_{ij}$ represents a noisy measurement of the difference between x_j and x_i.

Given these premises, the system state estimation problem consists in the determination of the agents state set $\{x_1^*, \ldots, x_n^*\}$ that allows to best approximate (and be consistent with) the set of the existing measurements. Adopting the standard LS approach, this estimation issue can be formalized in the optimization framework

Problem 1. For a given n-agent networked system modeled by the graph $\mathcal{G}$, let $\cup_{i=1,\ldots,n}\{\tilde{x}_{ij} \in \mathbb{R}, \ v_j \in \mathcal{N}_i\}$ be the set of all the available relative measurements. Then, the system state estimation task consists in the resolution of the following convex minimization problem

$$\underset{\{x_1, \ldots, x_n\}}{\arg\min} \frac{1}{2} \sum_{v_i \in \mathcal{V}} \sum_{v_j \in \mathcal{N}_i} (x_i - x_j + \tilde{x}_{ij})^2, \qquad (1)$$

whose cost function is hereafter denoted by $h(\mathbf{x})$, with $\mathbf{x} = [x_1 \ \cdots \ x_n]^\top \in \mathbb{R}^n$ identifying the *system state vector*.

3. DISTRIBUTED PP-BASED SOLUTION

In this section, the main contribution of this work is provided presenting a distributed PP-based solution for Problem 1 inspired by the state-of-art centralized and distributed approaches described in (Fabris et al., 2019).

3.1 Preliminary Observations

Problem 1 can be solved in a centralized way through the computation of the gradient of the cost function $h(\mathbf{x})$ w.r.t. $\mathbf{x}$, whose i-th component, $i \in \{1, \ldots, n\}$ results in

$$[\nabla_\mathbf{x} h(\mathbf{x})]_i = 2\deg(v_i)x_i - 2\sum_{v_j \in \mathcal{N}_i} x_j - \sum_{v_j \in \mathcal{N}_i} (\tilde{x}_{ji} - \tilde{x}_{ij}).$$

In particular, imposing the equilibrium condition $\nabla_\mathbf{x} h(\mathbf{x}) = \mathbf{0}_n$, with $\mathbf{0}_n \in \mathbb{R}^n$ indicating the n-dimensional zero (column) vector, the following linear system is derived

$$2\mathbf{L}\mathbf{x} = \tilde{\mathbf{x}}, \qquad (2)$$

where $\tilde{\mathbf{x}} = \left[\sum_{v_j \in \mathcal{N}_1}(\tilde{x}_{j1} - \tilde{x}_{1j}) \ \cdots \ \sum_{v_j \in \mathcal{N}_n}(\tilde{x}_{jn} - \tilde{x}_{nj})\right]^\top \in \mathbb{R}^n$ is the *relative measurements vector* and $\mathbf{L} \in \mathbb{R}^{n \times n}$ is the Laplacian matrix associated to $\mathcal{G}$. In the light of (2), it is straightforward that, if $\tilde{\mathbf{x}} \notin \ker(\mathbf{L}) \backslash \{\mathbf{0}_n\}$, the (minimum norm) *centralized* solution $\mathbf{x}^\star \in \mathbb{R}^n$ of the minimization (1) can be computed accounting for the pseudo-inverse $\mathbf{L}^\dagger \in \mathbb{R}^{n \times n}$ of the Laplacian matrix, namely

$$\mathbf{x}^\star = \frac{1}{2}\mathbf{L}^\dagger \tilde{\mathbf{x}}. \qquad (3)$$

On the other hand, Problem 1 can be solved also adopting the distributed paradigm. Indeed, each equation of system (2) is related to single agent information in terms of local topology and measurements. In particular, by setting $[\nabla_\mathbf{x} h(\mathbf{x})]_i = 0$, $i \in \{1, \ldots, n\}$, it follows that

$$x_i = \frac{2\sum_{v_j \in \mathcal{N}_i} x_j + \sum_{v_j \in \mathcal{N}_i}(\tilde{x}_{ji} - \tilde{x}_{ij})}{2\deg(v_i)}, \qquad (4)$$

namely, at the equilibrium, the i-th agent state depends only on the state of its neighbors and on its set of relative measurements. Thus, to solve Problem 1, the following iterative update rule can be defined for the whole system:

$$\Sigma_0 : \quad \mathbf{x}(k+1) = \mathbf{F}_0 \mathbf{x}(k) + \mathbf{u}_0, \qquad (5)$$

where the state matrix $\mathbf{F}_0 \in \mathbb{R}^{n \times n}$ is derived by (pre)-normalizing the adjacency matrix through the degree matrix, i.e.,

$$\mathbf{F}_0 = \mathbf{D}^{-1}\mathbf{A}.$$

Similarly, the input $\mathbf{u}_0 \in \mathbb{R}^n$ in (5) is determined by the normalizing relative measurements vector, i.e.,

$$\mathbf{u}_0 = \frac{1}{2}\mathbf{D}^{-1}\tilde{\mathbf{x}}.$$

Note that, since the graph $\mathcal{G}$ describing the multi-agent system is assumed to be connected, the spectrum $\Lambda(\mathbf{F}_0)$ of $\mathbf{F}_0$ accounts for n real eigenvalues $\lambda_0^{\mathbf{F}_0} \geq \cdots \geq \lambda_{n-1}^{\mathbf{F}_0}$ in the range $[-1, 1]$, with $\lambda_0^{\mathbf{F}_0} = 1$ having single algebraic multiplicity. In addition, in (Fabris et al., 2019), it is shown that the convergence of system (5) is ensured only if $\lambda_{n-1}^{\mathbf{F}_0} \neq -1$. This condition is fulfilled if and only if $\mathcal{G}$ is not bipartite, because the spectrum of the *normalized Laplacian matrix* $\mathcal{L} = \mathbf{I}_n - \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2} \in \mathbb{R}^{n \times n}$, corresponding to the eigenvalues $0 = \lambda_0^{\mathcal{L}} < \lambda_1^{\mathcal{L}} \leq \cdots \leq \lambda_{n-1}^{\mathcal{L}} \leq 2$ is related to $\Lambda(\mathbf{F}_0)$ so that

$$\lambda_i^{\mathcal{L}} = 1 - \lambda_i^{\mathbf{F}_0}, \quad i \in \{0, \ldots, n-1\}. \qquad (6)$$

3.2 Solution Derivation

In the following, an alternative iterative distributed solution is proposed for Problem 1. This rests on the PP approach illustrated in (Parikh et al., 2014) and entails the presence of the penalty parameter $\rho \geq 0$ whose tuning, depending on the network topology, may improve the convergence properties toward the optimal solution as compared to (4), as will be clarified in the next section.

Let us consider then the *regularized PP objective function* $h_\rho(k) = h_\rho(\mathbf{x}(k), \mathbf{x}(k+1))$ depending on the system state at the k-th and $(k+1)$-th step as follows

$$h_\rho(k) = h(\mathbf{x}(k+1)) + \frac{\rho}{2} \|\mathbf{x}(k+1) - \mathbf{x}(k)\|_2^2 ,$$

with $\rho \geq 0$. The gradient and Hessian of such a function w.r.t. $\mathbf{x}(k+1)$ result to be respectively

$$\nabla_{\mathbf{x}(k+1)} h_\rho(k) = \nabla_{\mathbf{x}(k+1)} h(\mathbf{x}(k+1)) + \rho(\mathbf{x}(k+1) - \mathbf{x}(k)),$$
$$\mathcal{H}_{\mathbf{x}(k+1)\mathbf{x}(k+1)} h_\rho(k) = 2\mathbf{L} + \rho \mathbf{I}_n. \tag{7}$$

Note that, for $\rho > 0$, the Hessian (7) is positive definite [1] $(\mathcal{H}_{\mathbf{x}(k+1)\mathbf{x}(k+1)} h_\rho(k) \succ 0)$, hence the cost function $h_\rho(k)$ turns out to be strictly convex and its minimization admits a unique global solution. This can be derived imposing the equilibrium condition $\nabla_{\mathbf{x}(k+1)} h_\rho(k) = \mathbf{0}_n$, which, for all $i \in \{1, \ldots, n\}$, yields

$$x_i(k+1) = \frac{\rho x_i(k) + 2 \sum_{v_j \in \mathcal{N}_i} x_j(k) + \sum_{v_j \in \mathcal{N}_i} (\tilde{x}_{ji} - \tilde{x}_{ij})}{2 \deg(v_i) + \rho}. \tag{8}$$

Exploiting (8), the following update rule can be designed:
$$\Sigma_\rho : \quad \mathbf{x}(k+1) = \mathbf{F}_\rho \mathbf{x}(k) + \mathbf{u}_\rho. \tag{9}$$
In this case, the state matrix $\mathbf{F}_\rho \in \mathbb{R}^{n \times n}$ is equal to the ρ-regularized adjacency matrix (pre)-normalized by the ρ-regularized degree matrix, namely

$$\mathbf{F}_\rho = \left(\mathbf{D} + \frac{\rho}{2} \mathbf{I}_n\right)^{-1} \left(\mathbf{A} + \frac{\rho}{2} \mathbf{I}_n\right). \tag{10}$$

Similarly, the input $\mathbf{u}_\rho \in \mathbb{R}^n$ in (9) is given by the vector of the ρ-regularized and normalized relative measurements, i.e.,

$$\mathbf{u}_\rho = \frac{1}{2} \left(\mathbf{D} + \frac{\rho}{2} \mathbf{I}_n\right)^{-1} \tilde{\mathbf{x}} = \left(\mathbf{D} + \frac{\rho}{2} \mathbf{I}_n\right)^{-1} \mathbf{D} \mathbf{u}_0.$$

Accounting for (9), the i-th agent state estimate at the $(k+1)$-th step turns out to be affected by the neighbors relative measurements and the neighbors state estimates at k-th step and by its relative measurements and the self estimate at k-th step. In particular, the self estimate is weighted through the penalty parameter ρ that thus allows to define a certain level of confidence w.r.t. the previous step computation. Note that, for $\rho = 0$, the proposed scheme Σ_ρ corresponds to Σ_0 in (5).

To conclude, because the graph $\mathcal{G}$ representing the multi-agent system is supposed to be connected, the spectrum $\Lambda(\mathbf{F}_\rho)$ of the (row-stochastic) matrix $\mathbf{F}_\rho$ consists in n real eigenvalues $\lambda_0^{\mathbf{F}_\rho} \geq \cdots \geq \lambda_{n-1}^{\mathbf{F}_\rho}$ belonging to the set

---

[1] Let $\mathbf{M} \in \mathbb{R}^{p \times p}$ be a matrix. Notation $\mathbf{M} \succeq 0$ and $\mathbf{M} \succ 0$ implies that $\mathbf{M}$ is positive semidefinite and positive definite in a generalized sense, respectively, i.e., all eigenvalues of $\mathbf{M} \succeq 0$ are real and nonnegative, even when $\mathbf{M} \neq \mathbf{M}^\top$ (bilinear forms in the latter case are not addressed). Moreover, given two $p \times p$ matrices $\mathbf{M}_1$ and $\mathbf{M}_2$, the notation $\mathbf{M}_1 \succeq \mathbf{M}_2$ stands for $\mathbf{M}_1 - \mathbf{M}_2 \succeq 0$.

$((\rho - 2d_M)/(\rho + 2d_M), 1] \subseteq (-1, 1]$, for $\rho > 0$, with $\lambda_0^{\mathbf{F}_\rho} = \lambda_0^{\mathbf{F}_0} = 1$ having single algebraic multiplicity.

4. CONVERGENCE ANALYSIS

Adopting the update rule (9), the convergence of the agent state estimate toward the equilibrium solution depends on the spectrum of the state matrix $\mathbf{F}_\rho$ and, in particular, the convergence rate depends on its second largest eigenvalue in modulus. Thus, this section aims at investigating the properties of $\Lambda(\mathbf{F}_\rho)$ and, subsequently, discussing the performance of the scheme Σ_ρ based on the topology characterizing the multi-agent system and on the selection of the penalty parameter ρ.

4.1 Insight on the Spectral Properties of $\mathbf{F}_\rho$

First, note that the spectrum of the matrix $\mathbf{F}_\rho$ in (10) trivially depends on the parameter ρ. In particular, property stated in the next lemma holds.

Lemma 1. Considering two values ρ_1, ρ_2 of the parameter ρ, it holds that $\lambda_i^{\mathbf{F}_{\rho_1}} > \lambda_i^{\mathbf{F}_{\rho_2}}$ for $i \in \{1, \ldots, n-1\}$ if and only if $\rho_1 > \rho_2 \geq 0$.

Proof. Consider $\mathbf{F}_\rho$ as the continuous function $\mathbb{R}_{>0} \to \mathbb{R}^{n \times n}, \rho \mapsto (2\mathbf{D} + \rho \mathbf{I}_n)^{-1}(2\mathbf{A} + \rho \mathbf{I}_n)$. By the Gershgorin circle theorem, the derivative of $\mathbf{F}_\rho$ w.r.t. ρ is positive semidefinite, namely $\mathbf{F}_\rho' = d\mathbf{F}_\rho/d\rho = 2 (2\mathbf{D} + \rho \mathbf{I}_n)^{-2} \mathbf{L} \succeq 0$. Hence, $\mathbf{F}_\rho$ is a nondecreasing function of the parameter ρ, i.e., $\rho_1 > \rho_2 \geq 0$ implies that $\mathbf{F}_{\rho_1} \succeq \mathbf{F}_{\rho_2}$ and viceversa. As a consequence, it follows that $\lambda_i^{\mathbf{F}_{\rho_1}} > \lambda_i^{\mathbf{F}_{\rho_2}}$ for $i \in \{1, \ldots, n-1\}$ if and only if $\rho_1 > \rho_2 \geq 0$. $\square$

Exploiting the results of Lem. 1, the next proposition provides a lower and an upper bound for each eigenvalue of $\mathbf{F}_\rho$, which turn out to be functions of the eigenvalues of $\mathbf{F}_0$, of the parameter ρ and of the maximum d_M and minimum d_m degree of $\mathcal{G}$.

Proposition 1. For any $\rho > 0$, all the eigenvalues $\lambda_i^{\mathbf{F}_\rho}$, $i \in \{1, \ldots, n-1\}$, of matrix $\mathbf{F}_\rho$ admit the following lower $\underline{\lambda}_i^{\mathbf{F}_\rho}$ and upper $\overline{\lambda}_i^{\mathbf{F}_\rho}$ bounds:

$$\underline{\lambda}_i^{\mathbf{F}_\rho} = \frac{\rho + 2\lambda_i^{\mathbf{F}_0} d_M}{\rho + 2d_M} \quad \text{and} \quad \overline{\lambda}_i^{\mathbf{F}_\rho} = \frac{\rho + 2\lambda_i^{\mathbf{F}_0} d_m}{\rho + 2d_m}. \tag{11}$$

Trivially, it holds that $\underline{\lambda}_i^{\mathbf{F}_\rho} = \lambda_i^{\mathbf{F}_\rho} = \overline{\lambda}_i^{\mathbf{F}_\rho}$ if and only if the graph $\mathcal{G}$ is regular, i.e., $\deg(v_i) = d \; \forall v_i \in \mathcal{V}$ with $d > 0$.

Proof. Let $\tilde{\mathbf{F}}_\rho(\mathbf{M}) : \mathbb{R}^{n \times n} \to \mathbb{R}^{n \times n}$ be a function that maps a diagonal matrix $\mathbf{M} \succ 0$ to the matrix $\tilde{\mathbf{F}}_\rho(\mathbf{M}) = (2\mathbf{M} + \rho \mathbf{I}_n)^{-1}(2\mathbf{M}\mathbf{F}_0 + \rho \mathbf{I}_n)$. Exploiting (6), criteria in Drazin and Haynsworth (1962) and the Gershgorin circle theorem, it is possible to prove that the condition $d_M \mathbf{I}_n \succeq \mathbf{M} \succeq d_m \mathbf{I}_n$, i.e. $\mathbf{M} - d_m \mathbf{I}_n \succeq 0$ and $d_M \mathbf{I}_n - \mathbf{M} \succeq 0$, implies

$$\begin{cases} 2\rho(\mathbf{M} - d_m \mathbf{I}_n)(\mathbf{I}_n - \mathbf{F}_0) \succeq 0 \\ 2\rho(d_M \mathbf{I}_n - \mathbf{M})(\mathbf{I}_n - \mathbf{F}_0) \succeq 0 \end{cases}$$

$$\begin{cases} (2\mathbf{M} + \rho \mathbf{I}_n)(2d_m \mathbf{F}_0 + \rho \mathbf{I}_n) \succeq (2d_m + \rho)(2\mathbf{M}\mathbf{F}_0 + \rho \mathbf{I}_n) \\ (2d_M + \rho)(2\mathbf{M}\mathbf{F}_0 + \rho \mathbf{I}_n) \succeq (2\mathbf{M} + \rho \mathbf{I}_n)(2d_M \mathbf{F}_0 + \rho \mathbf{I}_n) \end{cases}$$

and, thus, $\tilde{\mathbf{F}}_\rho(d_m \mathbf{I}_n) \succeq \tilde{\mathbf{F}}_\rho(\mathbf{M}) \succeq \tilde{\mathbf{F}}_\rho(d_M \mathbf{I}_n)$. The provided implications are valid for $\mathbf{M} = \mathbf{D}$. In this case, one has

$\tilde{\mathbf{F}}_\rho(\mathbf{D}) = \mathbf{F}_\rho$, hence the eigenvalue bounds in (11) can be derived for any $\rho > 0$ leveraging Lem. 1.

Finally, when $\mathcal{G}$ is a regular graph, it trivially holds that $d_m = d_M$ implying $\underline{\lambda}_i^{\mathbf{F}_\rho} = \lambda_i^{\mathbf{F}_\rho} = \overline{\lambda}_i^{\mathbf{F}_\rho}$, $i \in \{1, ..., n-1\}$. $\square$

4.2 Convergence Properties of Scheme Σ_ρ

Exploiting the spectral properties of $\mathbf{F}_\rho$, the rest of the section is devoted to the analysis of the role of parameter ρ in the performance of scheme Σ_ρ, by developing through three steps: first (in Prop. 2), the solution of Σ_ρ is shown to converge to the centralized one; secondly (in Prop. 3), it will be discussed when to adopt either the Σ_ρ or the Σ_0 scheme and how to tune the ρ parameter; finally (in Prop. 4), convergence rate bounds for the two schemes are obtained.

Proposition 2. For any $\rho > 0$, the estimate $\hat{\mathbf{x}} = \lim_{k \to +\infty} \mathbf{x}(k)$ provided by scheme Σ_ρ converges in terms of relative differences to the centralized solution $\mathbf{x}^\star$ in (3), namely $\hat{x}_i - \hat{x}_j = x_i^\star - x_j^\star$ for all $i, j \in \{1, \ldots, n\}$.

Proof. Since at the equilibrium it holds that

$$\hat{\mathbf{x}} = \mathbf{F}_\rho \hat{\mathbf{x}} + \mathbf{u}_\rho$$
$$(2\mathbf{D} + \rho \mathbf{I}_n) \hat{\mathbf{x}} = (2\mathbf{A} + \rho \mathbf{I}_n) \hat{\mathbf{x}} + \tilde{\mathbf{x}}$$
$$2\mathbf{L}\hat{\mathbf{x}} = \tilde{\mathbf{x}},$$

the solution provided by (9) converges to the centralized solution (3) in terms of relative differences, that is, up to a vector $\mathbf{y} \in \ker(\mathbf{L})$ such that $\hat{\mathbf{x}} = \mathbf{x}^\star + \mathbf{y}$. $\square$

Given this premise, in the following it is shown how the parameter ρ can be tuned in order to optimize the convergence performance of the scheme Σ_ρ through the minimization of the convergence rate $\mathfrak{r}_\rho \in [0, 1]$ of Σ_ρ, which is related to the second largest eigenvalue (in modulus) of $\mathbf{F}_\rho$. In particular, because of the monotonicity of $\Lambda(\mathbf{F}_\rho)$ for a given ρ, the convergence rate corresponds to

$$\mathfrak{r}_\rho = \max_{i \in \{1,...,n-1\}} |\lambda_i^{\mathbf{F}_\rho}| = \max\left(|\lambda_1^{\mathbf{F}_\rho}|, |\lambda_{n-1}^{\mathbf{F}_\rho}|\right). \quad (12)$$

To minimize $\mathfrak{r}_\rho$ in (12), a good and viable strategy is thus to select the parameter ρ as

$$\rho^\star = \arg\min_{\rho \geq 0} \left\{ \max\left(|\lambda_1^{\mathbf{F}_\rho}|, |\lambda_{n-1}^{\mathbf{F}_\rho}|\right) \right\} \quad (13)$$

The next proposition focuses on the solution of the optimization problem (13). Note that $\rho^\star$ might be zero: this case corresponds the adoption of scheme (5).

Proposition 3. The solution of problem (13) exists and is unique. Defining the function $\varsigma_{\mathbf{M}}$ that acts on matrix $\mathbf{M}$ with ordered spectrum such that $\varsigma_{\mathbf{M}} : \mathbb{R}^{n \times n} \to \mathbb{R}$, $\mathbf{M} \mapsto (\lambda_1^{\mathbf{M}} + \lambda_{n-1}^{\mathbf{M}})/2$, it holds that

$$\rho^\star = \begin{cases} \rho^+, & \varsigma_{\mathcal{L}} > 1 \\ 0, & \varsigma_{\mathcal{L}} \leq 1 \end{cases} \quad (14)$$

with

$$\underline{\rho}^+ = 2(\varsigma_{\mathcal{L}} - 1)d_m \leq \rho^+ \leq 2(\varsigma_{\mathcal{L}} - 1)d_M = \overline{\rho}^+. \quad (15)$$

Furthermore, both bounds $(\underline{\rho}^+, \overline{\rho}^+)$ in (15) coincide with ρ^+ if and only if $\mathcal{G}$ is regular.

Proof. Similarly to relation (6), for $\rho \geq 0$ it holds that

$$\frac{1}{2}\left(\left(1 - \lambda_{n-1}^{\mathbf{F}_\rho}\right) + \left(1 - \lambda_1^{\mathbf{F}_\rho}\right)\right) = \varsigma_{\mathcal{L}_\rho} = 1 - \varsigma_{\mathbf{F}_\rho}. \quad (16)$$

Thanks to Lem. 1, $\varsigma_{\mathbf{F}_\rho}$ is proven to be a continuous strictly increasing function of the parameter ρ, thus ensuring the existence of a unique solution for (13). This also implies that $\varsigma_{\mathbf{F}_\rho} > \varsigma_{\mathbf{F}_0}$ for $\rho > 0$.

Then, if $\varsigma_{\mathbf{F}_0} \geq 0$, namely $\varsigma_{\mathcal{L}} \leq 1$, it follows that $\varsigma_{\mathbf{F}_\rho} > 0$. Therefore, due to (16), the convergence rate results to be $\mathfrak{r}_\rho = |\lambda_1^{\mathbf{F}_\rho}|$ and it is trivially minimized for $\rho = 0$. Conversely, if $\varsigma_{\mathbf{F}_0} < 0$, namely $\varsigma_{\mathcal{L}} > 1$, the condition $\varsigma_{\mathbf{F}_\rho} = 0$, corresponding to the optimal scenario wherein $|\lambda_1^{\mathbf{F}_\rho}| = |\lambda_{n-1}^{\mathbf{F}_\rho}|$ may occur. In particular, for Bolzano's theorem, it follows that the condition $\varsigma_{\mathbf{F}_\rho} = 0$ is fulfilled only for one value of ρ (ρ^+).

Finally, resting upon Prop. 1, inequalities (15) can be derived by imposing

$$(\underline{\lambda}_1^{\mathbf{F}_{\rho^\star}} + \underline{\lambda}_{n-1}^{\mathbf{F}_{\rho^\star}})/2 \leq \varsigma_{\mathbf{F}_{\rho^\star}} \leq (\overline{\lambda}_1^{\mathbf{F}_{\rho^\star}} + \overline{\lambda}_{n-1}^{\mathbf{F}_{\rho^\star}})/2$$

and exploiting the fact that $\varsigma_{\mathbf{F}_{\rho^\star}} = 0$ only when $\varsigma_{\mathbf{F}_0} < 0$. Trivially, the bounds (15) are met for regular graphs since $d_m = d_M$ for these topologies. $\square$

Remark 1. Prop. 3 suggests when it is useful to adopt the PP-based solution (9) basing on the graph describing the given multi-agent system through $\varsigma_{\mathcal{L}}$. By definition, this quantity depends on the network connectivity and similarity to bipartite graph. Hence, when $\varsigma_{\mathcal{L}} > 1$, the corresponding topology tends to circulate the information between node groups. In this case, the choice $\rho^\star = \rho^+$ allows to improve the information management. On the other hand, when $\varsigma_{\mathcal{L}} \leq 1$, either $\rho^\star$ is set as an arbitrary relatively small value or the scheme Σ_0 has to be preferred, since the addition of self loop in the graph does not improve (and often worsen) the estimation speed.

To conclude, the next proposition discusses the bounds of the convergence rate for the optimal selection of ρ in (14)

Proposition 4. If $\varsigma_{\mathcal{L}} \leq 1$, the rate of convergence $\mathfrak{r}_{\rho^\star}$ for scheme $\Sigma_{\rho^\star} = \Sigma_0$ is yielded by $\mathfrak{r}_{\rho^\star} = 1 - \lambda_1^{\mathcal{L}} \in (0, 1)$. Otherwise, if $\varsigma_{\mathcal{L}} > 1$, it is bounded as follows

$$\mathfrak{r}_{\rho^\star} \leq \max\left(-\underline{\lambda}_{n-1}^{\mathbf{F}_{\underline{\rho}^+}}, \overline{\lambda}_1^{\mathbf{F}_{\overline{\rho}^+}}\right) < 1,$$

$$\mathfrak{r}_{\rho^\star} \geq \max\left(-\min\left(0, \overline{\lambda}_{n-1}^{\mathbf{F}_{\overline{\rho}^+}}\right), \max\left(0, \underline{\lambda}_1^{\mathbf{F}_{\underline{\rho}^+}}\right)\right) \geq 0.$$

In particular, $\mathfrak{r}_{\rho^\star} = 0$ if and only if $\mathcal{G}$ is the complete graph.

Proof. The result follows immediately from Prop. 1, Prop. 2, Prop. 3, the fact that if $\lambda_{n-1}^{\mathbf{F}_0} = -1$ then $\lambda_{n-1}^{\mathbf{F}_\rho} \geq (\rho - 2d_M)/(\rho + 2d_M) > -1$ for $\rho > 0$, and the general fact that $0 < \lambda_1^{\mathcal{L}} \leq 1$ holds for a non-complete connected graph $\mathcal{G}$. Moreover, since nontrivial eigenvalues of $\mathcal{L}$ are all equal to $n(n-1)^{-1}$ if and only if $\mathcal{G}$ is the complete graph, the assignment of $\rho^+ = 2$ to the parameter ρ allows to place all the eigenvalues of $\mathbf{F}_\rho$ different from 1 in 0, achieving superlinear convergence $(\mathfrak{r}_\rho^\star = 0)$. $\square$

5. NUMERICAL RESULTS

In this section, a comparison between Σ_ρ and Σ_0 is provided accounting for different network topologies and focusing the attention on the convergence properties of the two decentralized schemes.

In each test, the number of iterations is fixed to $\bar{k} = 20$ and the two following performance indexes are considered:
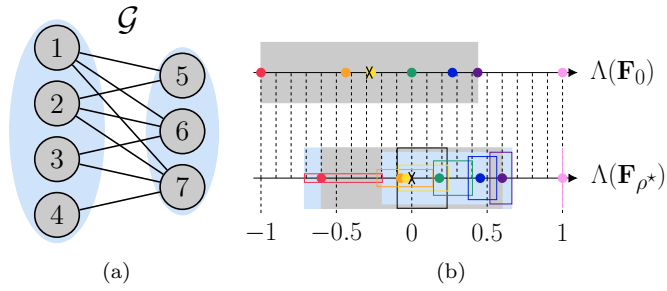
Fig. 1. (a) Nonregular bipartite graph with 7 nodes. (b) Spectral mapping due to penalty parameter tuning: the optimization of ρ allows the zeroing of the spectral centroid (black cross) and the reduction of the nonperipheral spectral radius (gray rectangle).

$$\mathfrak{r}_e = \frac{1}{\bar{k} - 2k_{neg}} \sum_{k=k_{neg}}^{\bar{k}-k_{neg}-1} \frac{\|\mathbf{x}(k+1) - \mathbf{x}^\star\|_2}{\|\mathbf{x}(k) - \mathbf{x}^\star\|_2}, \quad (17)$$

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} \left(x_i(\bar{k}) - x_i^\star \right)^2. \quad (18)$$

The former evaluates the effective convergence rate neglecting the first and last $k_{neg} = 2$ iterations and taking into account the average estimation error ratio, while the latter corresponds to the final mean square error [2].

5.1 Case study: nonregular bipartite topology

In this first test, the bipartite network in Fig. 1(a) is considered, assuming that the optimal solutions difference $x_{i+1}^\star - x_i^\star$ is equal for $i \in \{1, \ldots, n-1\}$ and that the relative measurements are corrupted by additive uniformly distributed noise.

Fig. 1(b) shows how the eigenvalues of the matrix $\mathbf{F}_0$ are mapped into the spectrum of $\mathbf{F}_\rho$, when the penalty parameter ρ corresponds to the solution $\rho^\star$ of (13). Note that, according to Lem. 1 and Prop. 1, all the the eigenvalues $\lambda_i^{\mathbf{F}_{\rho^\star}}$ are located within a certain neighborhood and maintain the same order w.r.t. the elements in $\Lambda(\mathbf{F}_0)$. In addition, since it is possible to compute the spectrum of the normalized Laplacian $\Lambda(\mathcal{L})$, Prop. 3 entails that $\rho^\star$ belongs to $[\underline{\rho}^\star, \overline{\rho}^\star] \simeq [0.5626, 2.2506]$, hence, the convergence rate $\mathfrak{r}_{\rho^\star}$ results in the interval $[\underline{\mathfrak{r}}_{\rho^\star}, \overline{\mathfrak{r}}_{\rho^\star}] \simeq [0.5184, 0.7118]$.

From Fig. 2, one can note that the convergence is not ensured adopting the scheme Σ_0: the convergence rate trivially results $\mathfrak{r}_0 = 1$. In detail, neither the cost function (top panel) "approaches" zero nor the estimations (bottom panel) converge towards the optimal solutions. This issue is easily overcome when the PP-based scheme is implemented: the choice of the penalty $\rho^\star = 1.3469$ not only guarantees convergence but also speeds up the estimation dynamics ($\mathfrak{r}_{\rho^\star} = 0.5999 < \mathfrak{r}_0$). In addition, accounting for the performance indexes (17)-(18), the scheme $\Sigma_{\rho^\star}$ yields $\mathfrak{r}_e = 0.0471$ and MSE $= 2.7485 \cdot 10^{-4}$, whereas, for the scheme Σ_0 only the final mean square error can be evaluate since convergence is not attain and this results considerably higher as compared to $\Sigma_{\rho^\star}$ being MSE $= 5.8291$.

---

[2] In (17)-(18), the state components are normalized in difference by, for instance, subtracting the first component to all the others.
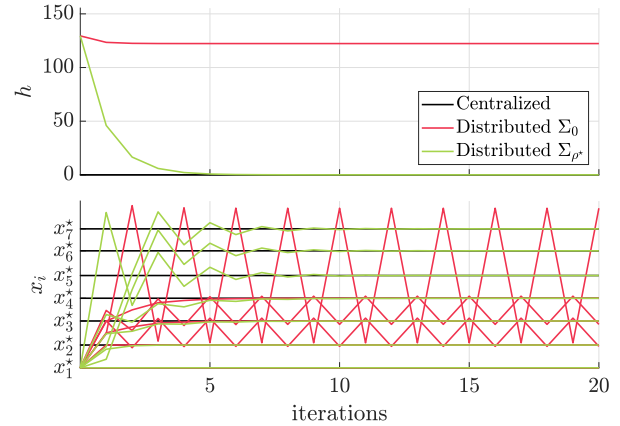

Fig. 2. (top) Cost function trend: the convergence is not attained when the scheme Σ_0 is adopted since $\mathcal{G}$ is bipartite. (bottom) Iterative schemes behavior: the state estimates oscillate around their optimal values for Σ_0, while they converge to the optimum for $\Sigma_{\rho^\star}$.

5.2 Other Topologies

In the following, the comparison between Σ_0 and Σ_ρ is carried out accounting for many different network topologies. In detail, the performances are evaluated for the networks considered in (Fabris et al., 2019), namely the complete graph K_{36}, the circulant graph $C_{36}(1,2)$, the Ramanujan graph $\mathcal{R}_{36}$ and the Cayley graph $\Gamma_{36}(3)$, all with $n = 36$ nodes. Furthermore, the following networks are also investigated: the star graph $\mathcal{S}_{36}$ with $n = 36$, i.e., a tree with 35 nodes characterized by a degree of 1 linked to a common root with 35 incidences; the small-world $SW_{9,27}$ with $n = 36$, constituted by two complete graphs K_9 and K_{27} linked together by one single edge; the complete binary trees $\mathcal{B}_4^+$ and $\mathcal{B}_6^+$ having $n = 32$ and $n = 128$ nodes, respectively, and characterized by an extra connection between the root and one of the leaves.

It is worth to recall that some of these basic topologies set the foundations for the development of key applications as camera networks, smart grids and social networks (Chen et al., 2019). As a result, Tab. 1 summarizes the parameters and all the main properties related to the convergence behaviors, jointly with some other graph-related quantities, as the *average degree* $d_{av}(\mathcal{G}) = n^{-1} \sum_{i=1}^{n} d_i$, the *density* $\text{dens}(\mathcal{G}) = 2|\mathcal{E}|/(n(n-1))$ and the *diameter* $\phi(\mathcal{G}) = \max\{\min(|l_{ij}| - 1)\}$, where $|l_{ij}|$ is the number of vertices belonging to the path that links nodes i and j.

From the data reported in this table, some considerations are in order.

- K_{36}: the estimation performance is good for both Σ_0 and $\Sigma_{\rho^\star}$, although the latter method allows for a faster convergence;
- $C_{36}(1,2)$ and $\mathcal{R}_{36}$: for these specific (regular) graphs the performances of Σ_0 and $\Sigma_{\rho^\star}$ are the same, for both indexes $\mathfrak{r}_e$ and MSE (as a further remark, note that for some circulant graphs, as the ring graphs, the performance results better by employing the PP-based approach);

Table 1. Summary of the validation test for the proposed iterative schemes in a heterogeneous topology framework.

Topology	K_{36}	$C_{36}(1,2)$	$\mathcal{R}_{36}$	$\Gamma_{36}(3)$	$SW_{9,27}$	$\mathcal{S}_{36}$	$\mathcal{B}_4^+$	$\mathcal{B}_6^+$
Regular	yes	yes	yes	yes	no	no	no	no
Bipartite	no	no	no	yes	no	yes	no	no
dens	1	0.1143	0.0857	0.0857	0.6159	0.0556	0.0667	0.0159
ϕ	1	9	6	6	3	2	8	12
d_{av}	35	4	3	3	21.5556	1.9444	2	2
d_m	35	4	3	3	8	1	1	1
d_M	35	4	3	3	27	35	3	3
$\lambda_1^{\mathcal{L}}$	1.0286	0.0377	0.1548	0.1181	0.0133	1	0.0261	0.0050
$\lambda_{n-1}^{\mathcal{L}}$	1.0286	1.5567	1.9407	2	1.1456	2	1.9888	1.9980
$\varsigma_{\mathcal{L}}$	1.0286	0.7972	0.9720	1.0590	0.5795	1.5	1.0074	1.0015
$\underline{\rho}^\star$	2	0	0	0.3542	0	1.0000	0.0148	0.0030
$\rho^\star$	2	0	0	0.3542	0	1.8972	0.0283	0.0059
$\overline{\rho}^\star$	2	0	0	0.3542	0	35.0000	0.0445	0.0090
$\underline{\mathfrak{r}}_{\rho^\star}$	0	0.9623	0.9131	0.8885	0.9867	0.0264	0.9740	0.9950
$\mathfrak{r}_{\rho^\star}$	0	0.9623	0.9131	0.8885	0.9867	0.4868	0.9741	0.9950
$\overline{\mathfrak{r}}_{\rho^\star}$	0	0.9623	0.9131	0.8885	0.9867	0.9472	0.9794	0.9961
$\mathfrak{r}_0$	0.0286	0.9623	0.9131	1	0.9867	1	0.9888	0.9980
$\mathfrak{r}_e(\Sigma_0)$	$5.2521 \cdot 10^{-5}$	0.2648	0.1940	–	0.7344	–	0.3149	0.4343
$\mathfrak{r}_e(\Sigma_{\rho^\star})$	$7.6734 \cdot 10^{-16}$	0.2648	0.1940	0.0399	0.7344	0.0289	0.2863	0.4336
MSE(Σ_0)	$1.6732 \cdot 10^{-14}$	3.9255	0.8262	0.2566	14.331	18.5849	2.2622	12.1184
MSE$(\Sigma_{\rho^\star})$	$1.6711 \cdot 10^{-14}$	3.9255	0.8262	0.0296	14.331	$1.1920 \cdot 10^{-5}$	1.9135	12.0192

- $\Gamma_{36}(3)$ and $\mathcal{S}_{36}$: for the bipartite nature of these graphs, only the PP-based estimation can be applied successfully ($\Gamma_{36}(3)$ is regular and $\mathcal{S}_{36}$ is not);
- $SW_{9,27}$: the two update rules Σ_0 and $\Sigma_{\rho^\star}$ perform equally, for this nonregular graph;
- $\mathcal{B}_4^+$ and $\mathcal{B}_6^+$: estimation performances are similar between the two methods, nonetheless, the PP-based procedure obtains better results both in estimation speed and estimation accuracy.

By recalling Remark 1, colored cells in Tab. 1 highlight how bipartite and complete graphs improve information spreading due to their connectivity patterns, (with $\varsigma_{\mathcal{L}} > 1$, using $\Sigma_{\rho^\star}$) while small-world networks tend to isolate information in node clusters (with $\varsigma_{\mathcal{L}} < 1$, using Σ_0).

6. CONCLUSIONS

This work focuses on system state estimation problem consisting in the computation of the optimal state estimation for all the components of a networked multi-agent system characterized by a set of noisy relative measurements. To solve this problem, a distributed iterative PP-based scheme is provided in the form of linear state-space model. The proposed solution is proven to converge in terms of relative differences to the corresponding centralized solution. In particular, its convergence features rests upon the spectral properties of stochastic matrices strictly related to the network topology and the normalized Laplacian matrix of a graph. Lower and upper bounds for the convergence rate of the provided PP scheme are reported along with the indications to compute an optimal value for the penalty parameter. Numerical simulations run in different scenarios validate the proposed estimation scheme and support the theoretical considerations on its performance.

Future directions for this work are represented by a performance comparison between the PP-based scheme, the spectral-based scheme Σ_η derived in (Fabris et al., 2019) and other schemes, as well as their generalization.

REFERENCES

Aster, R.C., Borchers, B., and Thurber, C.H. (2018). *Parameter estimation and inverse problems*. Elsevier.

Bai, J., Li, J., Dai, P., and Li, J. (2019). General parameterized proximal point algorithm with applications in statistical learning. *Int. J. of Computer Mathematics*, 96(1), 199–215.

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

Carron, A., Todescato, M., Carli, R., and Schenato, L. (2014). An asynchronous consensus-based algorithm for estimation from noisy relative measurements. *IEEE Trans. on Control of Network Systems*, 1(3), 283–295.

Chen, F., Ren, W., et al. (2019). On the control of multi-agent systems: A survey. *Foundations and Trends in Systems and Control*, 6(4), 339–499.

Cintuglu, M.H. and Ishchenko, D. (2019). Secure distributed state estimation for networked microgrids. *IEEE Internet of Things J.*

Drazin, M.P. and Haynsworth, E.V. (1962). Criteria for the reality of matrix eigenvalues. *Mathematische Zeitschrift*, 78(1), 449–452.

Fabris, M., Michieletto, G., and Cenedese, A. (2019). On the distributed estimation from relative measurements: a graph-based convergence analysis. In *2019 18th European Control Conf. (ECC)*, 1550–1555. IEEE.

Olfati-Saber, R., Fax, J.A., and Murray, R.M. (2007). Consensus and cooperation in networked multi-agent systems. *Proc. of the IEEE*, 95(1), 215–233.

Parikh, N., Boyd, S., et al. (2014). Proximal algorithms. *Foundations and Trends in Optimization*, 1(3), 127–239.

Ravazzi, C., Chan, N.P., and Frasca, P. (2018). Distributed estimation from relative measurements of heterogeneous and uncertain quality. *IEEE TSIPN*, 5(2), 203–217.