

Distributed Kalman filtering based on consensus strategies

Ruggero Carli, Alessandro Chiuso, Luca Schenato, Sandro Zampieri

Abstract

In this paper, we consider the problem of estimating the state of a dynamical system from distributed noisy measurements. Each agent constructs a local estimate based on its own measurements and estimates from its neighbors. Estimation is performed via a two stage strategy, the first being a Kalman-like measurement update which does not require communication, and the second being an estimate fusion using a consensus matrix. In particular we study the interaction between the consensus matrix, the number of messages exchanged per sampling time, and the Kalman gain. We prove that optimizing the consensus matrix for fastest convergence and using the centralized optimal gain is not necessarily the optimal strategy if the number of exchanged messages per sampling time is small. Moreover, we showed that although the joint optimization of the consensus matrix and the Kalman gain is in general a non-convex problem, it is possible to compute them under some important scenarios. We also provide some numerical examples to clarify some of the analytical results and compare them with alternative estimation strategies.

I. INTRODUCTION

The recent technological advances in wireless communication and the decreasing in cost and size of electronic devices, are promoting the appearance of large inexpensive interconnected systems, each with computational and sensing capabilities. These complex systems of agents can be used for monitoring very large scale areas with fine resolution. However, collecting measurements from distributed wireless sensors nodes at a single location for on-line data

R. Carli, L. Schenato and S. Zampieri are with the Department of Information Engineering, Università di Padova, Via Gradenigo 6/a, 35131 Padova, Italy. A. Chiuso is with the Dipartimento di Tecnica e Gestione dei Sistemi Industriali, Università di Padova, Stradella S. Nicola, 3 - 36100 Vicenza, Italy, {carlirug|chiuso|schenato|zampi}@dei.unipd.it

This work has been supported in part by the national project *New techniques and applications of identification and adaptive control* funded by MIUR, and by European Union project SENSNET funded by Marie Curie IRG grant n.014815

processing may not be feasible due to several reasons among which long packet delay (e.g. due to multi-hop transmission) and/or limited bandwidth of the wireless network, due e.g. to energy consumption requirements.

This problem is apparent in wireless ad-hoc sensor networks where information needs to be multi-hopped from one node to another using closer neighbors. Therefore there is a growing need for in-network data processing tools and algorithms that provide high performance in terms of on-line estimation while (i) reducing the communication load among all sensor nodes, (ii) being very robust to sensor node failures or replacements and packet losses, and (iii) being suitable for distributed control applications.

The literature is very rich of contributions addressing several aspects of distributed estimation and, for obvious reasons, we shall mention only a few. Most works focus on static estimation problems: [41] derives conditions under which one can reconstruct the global sufficient statistic from local sufficient statistics; [20] investigates how much information two sensors (say S_1 and S_2) have to transmit regarding their measurements (say y_1 and y_2) in order for a fusion center to be able to evaluate certain functions of the measured data y_1 and y_2 ; this latter paper introduces the concept of *communication complexity*, also shedding some new light on well-known data fusion formulas; [14], [31], [32], [40], [39] address quantization issues and optimal estimation using quantized data; [3] studies the problem of distributed estimation from relative measurements, with applications to localization and time-synchronization; [29] aims at reconstructing, in a decentralized manner, a field (say the temperature in a certain area) from local measurements (taken for instance from temperature sensors deployed in the environment). Concerning estimation of dynamic processes we should mention the so called *Sign of Innovation* (SOI) Kalman filter [33], [17]; in this approach each sensor broadcasts to *all* sensors in the network the new information acquired by sensing the environment, so that each sensor solves a “centralized” (i.e. with all information available) estimation problem; however, in order to limit the bandwidth requirement, only the sign of the innovation process is transmitted. It is also worth recalling the paper [19], in which the authors study a decentralized problem of joint estimation and control; the overall system is decomposed into “local” subsystems of fixed structure. They restrict to (suboptimal) linear estimation and control schemes in which the estimators and controllers gains are found by solving (off-line) a constrained parameter optimization problem, aiming at optimizing an asymptotic (infinite horizon) cost.

This is very similar, in spirit, to the approach followed in this paper; we will focus on distributed estimation of dynamical systems for which sensor nodes are not physically co-located and can communicate with each other according to some underlying communication network. For example, suppose that we want to estimate the temperature in a building that changes according to a random walk, i.e. $T(t+1) = T(t) + w(t)$, where $w(t)$ is a zero mean white noise process with covariance q , and we have N sensors that can measure temperature corrupted by some noise, i.e. $y_i(t) = T(t) + n_i(t)$, where $n_i(t)$ are zero mean white noise processes with same covariance r . If all measurements were instantaneously available to a single location, it is well known from the centralized Kalman filter that the optimal steady state estimator would have the following structure:

$$\hat{T}(t+1) = (1 - \ell_0)\hat{T}(t) + \ell_0 \text{mean}(y(t))$$

where $\text{mean}(y(t)) := \frac{1}{N} \sum_{i=1}^N y_i(t)$, and $0 < \ell_0 < 1$ is the optimal Kalman gain that depends on the process noise covariance q and the equivalent measurement noise variance r/N . This expression already shows two important features of the optimal estimator. The first feature is that the optimal state estimate $\hat{T}(t+1)$ is a weighted average between the previous state estimate $\hat{T}(t)$ and the average of the sensor measurements, thus implying that *averaging reduces uncertainty*. The second is that the optimal gain needs to be tuned to optimally balance process noise and the equivalent noise of the averaged measurements. In a distributed setting, it is not possible to assume that all measurements are instantaneously available at a specific location, since communication needs to be consistent with the underlying communication graph \mathcal{G} , and each sensor node has its own temperature estimate $\hat{T}_i(t)$. However, if it was possible to provide an algorithm that computes the mean of set of numbers only through local communication, then the optimal estimate could be computed at each sensor node as follows:

$$\begin{aligned} \hat{T}_i(t+1) &= (1 - \ell_0)\text{mean}(\hat{T}(t)) + \ell_0 \text{mean}(y(t)) \\ &= \text{mean}((1 - \ell_0)\hat{T}(t) + \ell_0 y(t)) \end{aligned}$$

Algorithms able to compute the average of a set of numbers in a distributed way are known as *average consensus algorithms*. They consist in iterations like $z^+ = Qz$, where z is the vector whose entries are the quantities to be averaged and Q is a doubly stochastic matrix, i.e. a matrix with properties $Q_{ij} \geq 0$, $\sum_j Q_{ij} = 1$ and $\sum_i Q_{ij} = 1$. The consensus problem has been widely studied in terms of convergence of Markov Chains [12] [38] [28], and it has

been recently proposed as an effective approach to solve many control problems ranging from flocking [16][21] to robot rendezvous [18][8]. Many interesting results have appeared recently [30][15][25][10] just to name a few. However, a detailed discussion about the average consensus problem is beyond the scope of this paper and we address the interested reader to the survey paper [22] and references therein. Under some weak connectivity properties [6], these matrices guarantee that $\lim_{m \rightarrow \infty} [Q^m z]_i = \text{mean}(z)$, i.e. all elements of vector $Q^m z$ converge to their initial mean $\text{mean}(z)$. Therefore, provided it is possible to communicate sufficiently fast within two subsequent sensor measurements, i.e. $m \gg 1$, then intuitively we can assume that the following distributed estimation strategy yields the optimal global state estimate:

$$\begin{aligned} z &= (1 - \ell_0) \hat{T}_i(t) + \ell_0 y_i(t) && \text{measur. \& predict. stage} \\ \hat{T}_i(t+1) &= [Q^m z]_i && \text{consensus stage} \end{aligned}$$

Olfati-Saber [27] and Spanos et al. [35] were the first to propose this two-stage strategy based on computing first the mean of the sensor measurements via consensus algorithms, and then to update and predict the local estimates using the centralized Kalman optimal gains. This approach can be extended to multivariable systems where the process evolves according to $T(t+1) = AT(t) + w(t)$ and the state is only partially observable, i.e. $y_i(t) = C_i T(t) + v_i(t)$, as shown in the static scenario by Xiao et al. [44] ($A = I, w(t) = 0$) and in the dynamic scenario in [36][24]. In this context, i.e. $m \gg 1$, it is natural to optimize Q for fastest convergence rate of Q^m , which corresponds to the second largest singular value of Q , for which there are already very efficient optimization tools available [42] [43]. The assumption $m \gg 1$ is reasonable in applications for which communication is inexpensive as compared to sensing. This is the case, for example, in rendezvous control or coordination of mobile sensors where moving and sensing the position is energetically more expensive than transmitting it to their neighbors. However, there are many other important applications in which the number m of messages exchanged per sampling time per node needs to be small, as required in static battery-powered wireless sensor networks. Therefore the assumption that $[Q^m z]_i \approx \text{mean}(z)$ is not valid. In this context, for example, it is not clear whether maximizing the rate of convergence of Q is the best strategy. Moreover, also the optimal gain ℓ becomes a function of the matrix Q and the number of exchanged messages m , which is unlikely to coincide with the optimal centralized Kalman gain proposed in all the aforementioned papers [27][35][36][24][44].

Recently, Alriksson et al. [1] and Speranzon et al. [37], considered the case $m = 1$, i.e. sensors are allowed to communicate once between sampling instants. In particular, in [1], the authors consider a general MIMO scenario where the matrix $Q = Q(t)$ ($= W$ in their terminology) is selected at each time step in order to minimize the estimation error covariance of each sensor for the next time step, with the only constraint to leave the estimate unbiased corresponding to $\sum_j Q_{ij} = 1$. Also the gain $\ell = \ell_i(t)$ ($= K$ in their terminology) is different for each sensor and chosen at each time-step using the standard Kalman prediction and correction procedures in order to minimize each sensor estimation error. Simulations show that this recursive algorithm converges and provide good performance, thus providing a methodology to jointly optimize Q and ℓ . However the authors do not provide any proof of convergence nor any global optimality guarantee. In fact, this distributed optimization approach greedily minimizes the error covariance of each sensor at each time step, which might not be globally optimal. Differently, in [37] the authors do not separate the algorithm between a consensus stage and an update and correct stage, but they consider a single update equation

$$\hat{T}(t+1) = K\hat{T}(t) + Hy(t)$$

where $\hat{T} = [\hat{T}_1^* \dots \hat{T}_N^*]^*$ and $y = [y_1^* \dots y_n^*]^*$ ¹, with the additional unbiasedness constraint $\sum_j (K_{ij} + H_{ij}) = 1$, i.e. row-sum equal to unity. Using our terminology we note that we would have $K = (1 - \ell)Q$ and $H = \ell Q$, which satisfy the constraint. Then they propose to compute the design matrices K, H by formulating an optimization problem where at each time step the sum of all sensor node covariance errors is minimized. Similarly to [1], this approach seems to converge and to provide good performance, but once again without any proof of global optimality and insight about the connectivity properties of the underlying graph.

The consensus-based approach to distributed estimation is not the only approach. In fact, recently Skizas et al. [34] proposed an iterative algorithm based on local estimates and on the quality of the local estimates through special nodes called bridges, and they proved converge to the centralized optimal estimator in the ideal scenario and to maintain good performance even under quantization and non-gaussian noise measurements.

In this paper, we want to study the interaction between the consensus matrix Q , the number of messages per sampling time m , and the gain ℓ . With respect with the aforementioned works,

¹The symbol “*” denotes the conjugate transpose.

we consider a simpler scenario with a scalar state which can be measured by N identical and independent sensors, a setup which still captures some of the most important features of the problem. In fact, also in this simple setup the joint optimization of Q and ℓ is not convex, as discussed in Section V. Our goal is to provide better insights about the problem of distributed estimation using consensus matrices, rather than posing it as a black-box optimization algorithm. Therefore, we explore some important regimes, namely fast communication $m \rightarrow \infty$, “small” measurement noise ($r/q \rightarrow 0$), and “small” process noise ($q/r \rightarrow 0$).

In particular, we show that the common practise of optimizing the spectral radius $\rho(Q)$ of the consensus matrix Q is the optimal strategy only under the fast communication regime. In fact, under the “small” measurement noise ($r/q \rightarrow 0$) regime it is more appropriate to optimize the Frobenius norm $\|Q^m\|_F$. In general, the minimizer for these two criteria do not coincide. Moreover, we also provide some useful guidelines for choosing the local filter gain ℓ under some asymptotic regimes. Finally we provide some numerical examples to clarify the proposed analytical results and we compare them with alternative distributed estimation strategies available in the literature.

The paper is organized as follows. In the next section, we introduce the problem formulation and the proposed estimation approach. In Section III we study optimal design of the consensus matrix Q for fixed Kalman gain ℓ , and pairwise, in Section IV we study optimal design of the Kalman gain ℓ for fixed consensus matrix Q . In Section V, after showing that the joint optimization of Q and ℓ is not convex, we derive some asymptotic results under some specific regimes. Section VI presents some numerical examples to illustrate the analytical results obtained in the previous sections, and to compare the proposed algorithm with other approaches to the same problem present in the literature. This latter section also shows a preliminary attempt to improve estimation performance by adding memory. Finally, in Section VII we summarize the results and we suggest possible future extensions.

Mathematical Preliminaries

Before proceeding we introduce some mathematical preliminaries that will be used throughout the paper. We indicate with $\mathbb{1} = [1 \ 1 \ \dots \ 1]^*$ the vector of all ones. The superscript $*$ shall denote complex conjugate transpose. The Frobenius norm of a matrix Q is defined as $\|Q\|_F := \sqrt{\text{tr}(QQ^*)} = \sqrt{\sum_{i,j} |Q_{ij}|^2}$, where Q_{ij} is the (i, j) -th entry of the matrix Q . We say that a

matrix Q is *stochastic* if it has nonnegative entries and the rows sum to one, i.e. $Q\mathbf{1} = \mathbf{1}$. Note that the spectrum $\sigma(Q)$ of a stochastic matrix is contained in the closed unit disk.

For a stochastic matrix Q we define the *essential spectral radius* $\rho(Q)$ as

$$\rho(Q) = \begin{cases} 1 & \text{if } \dim \ker(Q - I) > 1 \\ \max_{\lambda \in \sigma(Q) \setminus \{1\}} |\lambda| & \text{if } \dim \ker(Q - I) = 1, \end{cases} \quad (1)$$

A stochastic matrix is said to be *primitive* if there exists an integer k such that Q^k has strictly positive entries [4]. This is equivalent to requiring that $\rho(Q) < 1$ (see [13], Theorem 8, pag. 80).

Furthermore, we shall say that Q is *doubly stochastic* if both Q and Q^* are stochastic.

II. PROBLEM FORMULATION

Consider a set V of N sensor nodes which are labeled $i = 1, 2, \dots, N$. These sensors can communicate over a network modeled as a direct graph $\mathcal{G} = (V, E)$, where the edge (i, j) is in E if and only if the node i can transmit its information to the node j . We assume that the graph \mathcal{G} is time-invariant. A physical process with state $x \in \mathbb{R}$ evolves according to the continuous-time system

$$\dot{x}(t) = v(t) \quad (2)$$

where $v(t)$ is a continuous-time white noise² of zero mean and intensity $q \geq 0$, that is $\mathbb{E}[q(t)q(s)] = q\delta_{ts}$, where δ_{hk} is the Kronecker delta. The initial condition is also assumed to be a random variable with mean x_0 and variance σ .

Each sensor take measurements of the physical process according to the equation

$$y_i(kT) = x(kT) + n_i(kT), \quad k \in \mathbb{Z}, \quad (3)$$

where T is the sampling time. Note that $y_i \in \mathbb{R}$, $\forall i$. We denote $y(kT) = [y_1(kT), \dots, y_N(kT)]^*$ and $n(kT) = [n_1(kT), \dots, n_N(kT)]^*$. Moreover the noise processes $n_i(kT) \in \mathbb{R}$ are such that $\mathbb{E}[n(kT)] = 0$, $E[n(kT)n(hT)] = rI\delta_{hk}$. Note also that (3) can be rewritten in the following vector form

$$y(kT) = x(kT)\mathbf{1} + n(kT). \quad (4)$$

²We recall that what is commonly referred to as ‘‘continuous time white noise’’ can be thought of as the ‘‘derivative’’ of a Wiener process which, unfortunately, is nowhere differentiable. More rigorously $x(t)$ is a Wiener process.

From now on we assume, without loss of generality, that $T = 1$. Suppose also that, between each pair of subsequent measurement update indexes k and $k + 1$, each node exchanges m messages; we assume that these transmissions take place at the following times $k + \delta, k + 2\delta, \dots, k + (m - 1)\delta, k + m\delta$, where $\delta = \frac{1}{m}$. Note that $k + m\delta = k + 1$.

We shall denote with $\hat{x}_i(k + h\delta|k)$ the estimate, at node i , of the state x at time $k + h\delta$ given measurements up to time k . In compact form we shall also denote with $\hat{x} := [\hat{x}_1, \dots, \hat{x}_N]^*$ the vector of estimates throughout the network; more precisely, making the dependence upon time explicit:

$$\hat{x}(k + h\delta|k) := \begin{bmatrix} \hat{x}_1(k + h\delta|k) \\ \vdots \\ \hat{x}_N(k + h\delta|k) \end{bmatrix}.$$

We assume that these estimates are updated, for $k \geq 0$, according to the following rule

$$\begin{cases} \hat{x}(k|k) &= (1 - \ell(k)) \hat{x}(k|k - 1) + \ell(k)y(k) \\ \hat{x}(k + h\delta|k) &= Q(k, h)\hat{x}(k + (h - 1)\delta|k), \quad h = 1, \dots, m \end{cases} \quad (5)$$

where $Q(k, h)$ is a suitable matrix compatible with the communication graph and where $0 < \ell(k) < 1, \forall k \geq 0$. From now on we shall use constant ‘‘gains’’ $\ell(k)$ and $Q(k, h)$, i.e. $\ell(k) = \ell$ and $Q(k, h) = Q$; furthermore, we shall also assume the recursions (5) are initialized by

$$\hat{x}(0|0) := y(0). \quad (6)$$

A natural request is also that \hat{x}_i be an unbiased estimator for all i 's and for all times, i.e. $\mathbb{E}[\hat{x}_i(k + h\delta|k)] = x_0, \forall i, \forall k \geq 0, \forall h \in [1, m]$; this leads to the condition

$$Q\mathbf{1} = \mathbf{1}. \quad (7)$$

In fact, from the update rule (5) it follows that

$$x_0\mathbf{1} = \mathbb{E}[\hat{x}(k + (h + 1)\delta|k)] = \mathbb{E}[Q\hat{x}(k + h\delta|k)] = x_0Q\mathbf{1}.$$

For $x_0 \neq 0$, the equality $x_0Q\mathbf{1} = x_0\mathbf{1}$ implies (7). Furthermore, if we restrict Q to have non-negative entries, condition (7) is equivalent to saying that Q is a stochastic matrix. Hence, from now on, we shall assume that Q is stochastic.

Now we define the new variable $\tilde{x}(k + h\delta|k) = x(k + h\delta)\mathbf{1} - \hat{x}(k + h\delta|k)$ which represents

the estimation error. In order to analyze the structure of the recursive equations that $\tilde{x}(k+h\delta|k)$ satisfies, it is convenient discretize (2) in the following way

$$x(k+(h+1)\delta) = x(k+h\delta) + w(k+h\delta)$$

where

$$w(k+h\delta) = \int_{k+h\delta}^{k+(h+1)\delta} v(\tau) d\tau.$$

Note that w is a discrete time white noise with mean zero and variance q/m , i.e. $\mathbb{E}[w(k+h\delta)] = 0$ and $\mathbb{E}[w^2(k+h\delta)] = \frac{q}{m}$. By straightforward calculations, for $h=0$, we get

$$\tilde{x}(k|k) = (1-\ell)\tilde{x}(k|k-1) - \ell n(k)$$

and, for $1 \leq h \leq m$,

$$\tilde{x}(k+h\delta|k) = Q^h \tilde{x}(k|k) + \left(\sum_{i=0}^{h-1} w(k+i\delta) \right) \mathbf{1}$$

In this paper we shall study the *second order* properties of the error, measured by the covariance matrices:

$$P(k+h\delta|k) = \mathbb{E}[\tilde{x}(k+h\delta|k) \tilde{x}(k+h\delta|k)^*],$$

defined for $0 \leq h \leq m$. One can show that $P(k+h\delta|k)$ satisfies, for $h=0$,

$$P(k|k) = (1-\ell)^2 P(k|k-1) + \ell^2 r I \quad (8)$$

and, for $h=m$,

$$P(k+1|k) = Q^m P(k|k) (Q^m)^* + q \mathbf{1} \mathbf{1}^*. \quad (9)$$

Plugging (8) into (9) we obtain the recursive equation

$$P(k+1|k) = (1-\ell)^2 Q^m P(k|k-1) (Q^*)^m + \ell^2 r Q^m (Q^*)^m + q \mathbf{1} \mathbf{1}^* \quad (10)$$

while, inserting (9) into (8) evaluated at the index $k+1$ we get:

$$P(k+1|k+1) = (1-\ell)^2 Q^m P(k|k) (Q^*)^m + (1-\ell)^2 q \mathbf{1} \mathbf{1}^* + \ell^2 r I. \quad (11)$$

From (6) it follows that the error covariance $P(0|0)$ satisfies $P(0|0) = rI$; similarly $P(1|0)$ is given by $P(1|0) = rQ^m (Q^m)^* + q \mathbf{1} \mathbf{1}^*$. Iterating the update rule (10), starting from the initial condition $P(1|0)$, we obtain

$$P(k+1|k) = (1-\ell)^{2k} Q^{km} P(1|0) (Q^*)^{km} + \ell^2 r \sum_{i=0}^{k-1} (1-\ell)^{2i} Q^{(i+1)m} (Q^*)^{(i+1)m} + q \sum_{i=0}^{k-1} (1-\ell)^{2i} \mathbf{1} \mathbf{1}^* ;$$

similarly, iterating (11) with initial condition $P(0|0)$, we get

$$P(k|k) = (1 - \ell)^{2k} Q^{km} P(0|0) (Q^*)^{km} + q \sum_{i=0}^{k-1} (1 - \ell)^{2i+2} \mathbf{1} \mathbf{1}^* + \ell^2 r \sum_{i=0}^{k-1} (1 - \ell)^{2i} Q^{im} (Q^*)^{im}$$

In this paper we shall be concerned with the asymptotic ($k \rightarrow \infty$) behavior of the error covariance. Hence we consider the limits

$$\begin{aligned} \lim_{k \rightarrow \infty} P(k+1|k) &= r \ell^2 \sum_{i=0}^{\infty} (1 - \ell)^{2i} Q^{(i+1)m} (Q^*)^{(i+1)m} + q \frac{1}{1 - (1 - \ell)^2} \mathbf{1} \mathbf{1}^* \\ \lim_{k \rightarrow \infty} P(k|k) &= q \frac{(1 - \ell)^2}{1 - (1 - \ell)^2} \mathbf{1} \mathbf{1}^* + r \ell^2 \sum_{i=0}^{\infty} (1 - \ell)^{2i} Q^{im} (Q^*)^{im} \end{aligned}$$

and define the cost functions³

$$J_1(\ell, Q; m, r, q) = \text{tr} \left\{ \lim_{k \rightarrow \infty} P(k+1|k) \right\}$$

and

$$J_2(\ell, Q; m, r, q) = \text{tr} \left\{ \lim_{k \rightarrow \infty} P(k|k) \right\}$$

These costs lead to the formulation of the following minimization problem:

Problem 1: Given a graph \mathcal{G} and a nonnegative integer m , find a real $\ell \in (0, 1)$ and a matrix $Q \in \mathcal{Q}$, where \mathcal{Q} is the set of stochastic matrices compatible with the graph \mathcal{G} , minimizing J_1 or J_2 .

In the sequel the set \mathcal{Q} may be further restricted while always being compatible with the topology of the communication network.

Remark 1: In the sequel we will consider only J_1 . The reason will be clear in the next sections where the minimization on J_1 will permit us to retrieve, for some particular cases, the results already known in the literature regarding Kalman filtering. For the sake of simplicity, we will denote this cost function simply by J in place of J_1 . Hence

$$J = r \ell^2 \text{tr} \left\{ \sum_{i=0}^{\infty} (1 - \ell)^{2i} Q^{(i+1)m} (Q^*)^{(i+1)m} \right\} + q \frac{1}{1 - (1 - \ell)^2} N \quad (12)$$

³In the remainder of the paper, when there is no risk of confusion, we might drop some arguments of the cost (e.g. denote $J_1(\ell, Q)$ rather than $J_1(\ell, Q; m, r, q)$).

Remark 2: Let us denote with λ_i , $i = 0, \dots, N-1$ the eigenvalues⁴ of Q ; since Q is stochastic, we can set, w.l.o.g., $\lambda_0 = 1$, i.e. $\sigma(Q) = \{1, \lambda_1, \lambda_2, \dots, \lambda_{N-1}\}$. Note that, if Q is a normal matrix, i.e. $QQ^* = Q^*Q$, then $\text{tr}\{Q^{im}(Q^*)^{im}\} = \sum_{h=0}^{n-1} |\lambda_h|^{2im}$; it is hence easy to see that formula (12) can be rewritten as follows:

$$J = \frac{r\ell^2 + qN}{1 - (1 - \ell)^2} + r\ell^2 \sum_{i=1}^{N-1} \frac{|\lambda_i|^{2m}}{1 - (1 - \ell)^2 |\lambda_i|^{2m}} \quad (13)$$

Also note that, if Q is normal and stochastic, then it is also doubly stochastic. Relevant subclasses of normal matrices are, for instance, Abelian Cayley matrices [2], circulant matrices and symmetric matrices.

The previous remark suggests that the following assumption is both useful and reasonable.

Assumption 1: The set \mathcal{Q} in problem 1 is the subset of normal stochastic matrices which are compatible with the graph \mathcal{G} .

Example 1: Assume that \mathcal{Q} is the set of the circulant stochastic matrices of the form

$$Q_k = \begin{bmatrix} 1-k & k & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1-k & k & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ k & 0 & 0 & 0 & \cdots & 0 & 1-k \end{bmatrix}$$

where $k \in [0, 1]$ and assume that $m = 1$. Let

$$Q^{opt}, \ell^{opt} \in \arg \min_{\ell \in (0,1); Q \in \mathcal{Q}} J(Q, \ell; m, r, q) \quad (14)$$

It is well known in the literature [11] that the eigenvalues of Q can be expressed in the following form $\lambda_h = 1 - k + ke^{j\frac{2\pi}{N}h}$, $0 \leq h \leq N-1$. Notice that $\left|1 - k + ke^{j\frac{2\pi}{N}h}\right| \geq \left|1/2 + 1/2e^{j\frac{2\pi}{N}h}\right|$, $\forall h : 0 \leq h \leq N-1$. Hence it follows that

$$Q_{\frac{1}{2}} = \arg \min_{Q \in \mathcal{Q}} J(\ell, Q; 1, r, q)$$

for all $\ell \in (0, 1)$, i.e. $Q^{opt} = Q_{\frac{1}{2}}$. In order to calculate the optimal gain we have to solve

$$\arg \min_{\ell \in (0,1)} \frac{qN}{1 - (1 - \ell)^2} + r\ell^2 \sum_{h=0}^{N-1} \frac{|\lambda_h|^2}{1 - (1 - \ell)^2 |\lambda_h|^2} \quad (15)$$

⁴Multiple eigenvalues are counted as many times as their algebraic multiplicity.

where $\lambda_h = \frac{1}{2} + \frac{1}{2}e^{j\frac{2\pi}{N}h}$. Unfortunately it is not possible to give, in general, a closed form expression for the optimizing gain ℓ^{opt} , which has to be found using numerical search techniques.

However a simple expression for (15) can be obtained when the number of sensors N goes to infinity, which allows to study the behavior of the optimal ℓ for large networks. To this purpose it is convenient to define the *normalized cost*

$$\begin{aligned}\bar{J}_N &:= \frac{J}{N} = \frac{q}{1 - (1 - \ell)^2} + \frac{r\ell^2}{N} \sum_{h=0}^{N-1} \frac{|\lambda_h|^2}{1 - (1 - \ell)^2 |\lambda_h|^2} \\ &= \frac{q}{1 - (1 - \ell)^2} + \frac{r\ell^2}{N} \tilde{J}_N\end{aligned}$$

where the last equation defines \tilde{J}_N . Define also the function $f : \mathbb{C} \rightarrow \mathbb{C}$

$$f(z) = \frac{1}{2} + \frac{1}{2}z.$$

and note that $\lambda_h = f\left(e^{j\frac{2\pi}{N}h}\right)$. As $N \rightarrow \infty$, \tilde{J}_N converges to

$$\lim_{N \rightarrow \infty} \tilde{J}_N = \tilde{J}_\infty = \frac{1}{2\pi} \int_0^{2\pi} \frac{|f(e^{j\phi})|}{1 - (1 - \ell)^2 |f(e^{j\phi})|^2} d\phi = \frac{1}{2\pi j} \oint_\gamma \frac{z^{-1}(2 + z + z^{-1})}{4 - (1 - \ell)^2(2 + z + z^{-1})} dz$$

where γ is the unit circle. It is straightforward to see that the poles of $\frac{z^{-1}(2+z+z^{-1})}{4-(1-\ell)^2(2+z+z^{-1})}$ inside γ are $z_1 = \frac{2-(1-\ell)^2}{(1-\ell)^2} - \sqrt{\left(\frac{2-(1-\ell)^2}{(1-\ell)^2}\right)^2 - 1}$, and $z_2 = 0$; the integral can be computed explicitly using the residue theorem yielding:

$$\tilde{J}_\infty = \frac{1}{(1 - \ell)^2} \left(\frac{1}{\sqrt{1 - (1 - \ell)^2}} - 1 \right)$$

Hence

$$\lim_{N \rightarrow \infty} \bar{J}_N \left(Q_{\frac{1}{2}}, \ell; 1, r, q \right) = \frac{q}{1 - (1 - \ell)^2} + \frac{r\ell^2}{(1 - \ell)^2} \left(\frac{1}{\sqrt{1 - (1 - \ell)^2}} - 1 \right). \quad (16)$$

The value of ℓ^{opt} can be found by minimizing the limiting expression for \bar{J}_N in (16).

III. OPTIMAL CONSENSUS MATRIX Q FOR FIXED GAIN ℓ

In this section we assume that the estimation gain ℓ is fixed, and thus consider the optimization problem:

$$Q^{opt}(\ell; m) = \arg \min_{Q \in \mathcal{Q}} J(Q, \ell; m). \quad (17)$$

Although the analysis of this problem is quite hard in general, a detailed study can be carried out in some interesting situations. In particular in the following we will restrict to the three special cases:

- the communication graph \mathcal{G} is undirected
- the sensors can communicate arbitrarily fast within two subsequent measurements, i.e., $m \rightarrow \infty$
- the estimation gain ℓ is sufficiently large, i.e. $\ell \rightarrow 1$; this intuitively corresponds to the situation in which the variance of the measurement noise is negligible with respect to the variance of the process, i.e $\frac{r}{q} \approx 0$.

Before proceeding, we observe that

$$\min_{Q \in \mathcal{Q}} J = \frac{r\ell^2 + qN}{1 - (1 - \ell)^2} + \min_{Q \in \mathcal{Q}} r\ell^2 \sum_{j=1}^{N-1} \frac{|\lambda_j|^{2m}}{1 - (1 - \ell)^2 |\lambda_j|^{2m}}$$

and hence only the second term on the right hand side play a role in the optimization. We can therefore restrict to consider only this latter quantity which, for convenience of notation, we denote as

$$S(Q, \ell; m) = \sum_{j=1}^{N-1} \frac{|\lambda_j|^{2m}}{1 - (1 - \ell)^2 |\lambda_j|^{2m}} \quad (18)$$

A. Undirected communication graph \mathcal{G}

First of all we observe that, if the communication graph \mathcal{G} is undirected, $Q \in \mathcal{Q}$ implies that also $Q^* \in \mathcal{Q}$. For any $Q \in \mathcal{Q}$, consider its symmetric part $Q_{sym} := (Q + Q^*)/2$. Clearly, Q_{sym} is normal and it is compatible with \mathcal{G} , therefore $Q_{sym} \in \mathcal{Q}$. The following lemma provides an interesting comparison between $J(Q, \ell; m)$ and $J(Q_{sym}, \ell; m)$ showing that the former is always greater or at most equal to the latter.

Lemma 3.1: Let Q be any matrix in \mathcal{Q} and let Q_{sym} be defined as above. Then

$$J(Q_{sym}, \ell; m) \leq J(Q, \ell; m).$$

Proof: Let λ_i be any eigenvalue of Q . Then $Re\{\lambda_i\}$ is an eigenvalue of Q_{sym} , where $Re\{\lambda_i\}$ denotes the real part of λ_i . Clearly $|Re\{\lambda_i\}| \leq |\lambda_i|$ which implies that

$$\frac{|Re\{\lambda_i\}|^{2m}}{1 - (1 - \ell)^2 |Re\{\lambda_i\}|^{2m}} \leq \frac{|\lambda_i|^{2m}}{1 - (1 - \ell)^2 |\lambda_i|^{2m}}.$$

Therefore, from (18), it follows that $S(Q_{sym}, \ell; m) \leq S(Q, \ell; m)$; hence also $J(Q_{sym}, \ell; m) \leq J(Q, \ell; m)$ holds true. ■

Remark 3: It is important to note that normality plays a fundamental role in the previous lemma which cannot be generalized to stochastic matrices \mathcal{Q} . In fact, it is easy to find a non-normal \mathcal{Q} for which the symmetric part \mathcal{Q}_{sym} gives a larger cost index.

An immediate consequence of Lemma 3.1 is that, when the communication graph is undirected, the minimum of the cost function J is reached by symmetric matrices. Thus, if \mathcal{Q}_{sym} is the subset of \mathcal{Q} containing the symmetric matrices, i.e. $\mathcal{Q}_{sym} = \{Q \in \mathcal{Q} : Q = Q^*\}$, solving (17) is equivalent to solve

$$\arg \min_{Q \in \mathcal{Q}_{sym}} J(Q, \ell; m). \quad (19)$$

The following result provides a powerful characterization of (19) which has important implications when it comes to performing optimization.

Theorem 3.1: Let \mathcal{Q}_{sym} be as above. Then the cost function $J(Q, \ell; m)$ defined on \mathcal{Q}_{sym} is a convex function.

Proof: Consider the function $f : \mathcal{B} \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ defined as

$$f(x) = \sum_{i=1}^n \frac{x_i^{2m}}{1 - \alpha x_i^{2m}},$$

where $m \in \mathbb{N}$, $0 < \alpha < 1$, $x = [x_1, x_2, \dots, x_n]^T$ and where $\mathcal{B} = \{x \in \mathbb{R}^n : |x_i| \leq 1\}$. It is easy to verify that the function f is convex and symmetric, i.e. it is invariant to any permutation of the vector entries x_i . Hence, it follows from the theory of convex spectral functions that also J is a convex function [5]. ■

Theorem 3.1 states that (19) is a convex problem thus implying that the solution of (19) can be performed efficiently by suitable numerical algorithms. In fact, Xiao et al. [43] adopted this strategy to optimize similar performance costs over symmetric stochastic matrices.

B. Fast communication ($m \rightarrow \infty$)

It is well known in the literature (see [8], [26]) that the speed of convergence of consensus algorithms is governed by the essential spectral radius of the consensus matrix.

Indeed, the essential spectral radius plays also an important role in distributed estimation provided communication is ‘‘sufficiently fast’’. In fact the following theorem shows that optimizing the essential spectral radius improves the performance provided that the number of message exchanges is larger than a specified bound. However, when comparing the performance of two

consensus matrices Q_1 and Q_2 , this bound is a function of $\rho(Q_1)$ and $\rho(Q_2)$ as explained in the following theorem.

Theorem 3.2: Let Q_1 and Q_2 be two matrices such that $\rho(Q_1) > \rho(Q_2)$. Then there exists \bar{m} (depending only on $\rho(Q_1) - \rho(Q_2)$) such that

$$J(Q_1, \ell; m) > J(Q_2, \ell; m), \quad \forall m > \bar{m}.$$

Proof: Let $\rho_1 = \rho(Q_1)$, $\rho_2 = \rho(Q_2)$ and $\epsilon = \rho(Q_1) - \rho(Q_2) = \rho_1 - \rho_2$. Observe that

$$\begin{aligned} S(Q_1, \ell; m) &\geq \frac{\rho_1^{2m}}{1 - (1 - \ell)^2 \rho_1^{2m}} \\ &> \frac{(\rho_2 + \epsilon)^{2m}}{1 - (1 - \ell)^2 \rho_2^{2m}} \end{aligned}$$

and that

$$S(Q_2, \ell; m) \leq N \frac{\rho_2^{2m}}{1 - (1 - \ell)^2 \rho_2^{2m}}.$$

Hence, if $(\rho_2 + \epsilon)^{2m} > N \rho_2^{2m}$, we have that $S(Q_1, \ell; m) > S(Q_2, \ell; m)$. Straightforward calculations show that this last condition is satisfied if and only if $m > \frac{\log N}{2 \log \left(1 + \frac{\epsilon}{\rho_2}\right)}$. Note that

$\frac{\log N}{2 \log(1 + \epsilon)} \geq \frac{\log N}{2 \log \left(1 + \frac{\epsilon}{\rho_2}\right)}$. Therefore by letting $\bar{m} = \left\lceil \frac{\log N}{2 \log(1 + \epsilon)} \right\rceil$ the statement of the theorem follows. Note that \bar{m} depends only on ϵ . ■

C. Large gain ($\ell \rightarrow 1$)

In this section we shall instead consider the number of message exchanges m as fixed and assume the gain ℓ is “large”; this, as we shall also see in Section V-B, is what happens when the measurement noise is small as compared to process noise, a situation frequently encountered in practice. It is remarkable that in this case the Frobenius norm is instead the “right” way to compare consensus matrices. This is made precise by the following theorem.

Theorem 3.3: Let Q_1, Q_2 be two matrices such that $\|Q_1^m\|_F > \|Q_2^m\|_F$. Then there exists $\bar{\ell}$ (depending only on $\|Q_1^m\|_F - \|Q_2^m\|_F$) such that

$$J(Q_1, \ell; m) > J(Q_2, \ell; m), \quad \forall \ell > \bar{\ell}. \quad (20)$$

Proof: Let $f(\ell, m) = S(Q_1, \ell; m) - S(Q_2, \ell; m)$ and $\epsilon = \|Q_1^m\|_F - \|Q_2^m\|_F$. We have that $f(\ell, m)$ can be written in the following way

$$f(\ell, m) = \sum_{i=0}^{\infty} \left(\|Q_1^{(i+1)m}\|_F^2 - \|Q_2^{(i+1)m}\|_F^2 \right) (1 - \ell)^{2i} = \sum_{i=0}^{\infty} \alpha_i (1 - \ell)^{2i}$$

where $\alpha_i = \|Q_1^{(i+1)m}\|_F^2 - \|Q_2^{(i+1)m}\|_F^2$. It is straightforward to see that $\alpha_0 > \epsilon^2$ and that, for $i \geq 1$, $\alpha_i \geq -(N-1)\rho^{2m(i+1)}$, where $\rho = \rho(Q_2)$. Hence, it follows that

$$\begin{aligned} f(\ell, m) &> \epsilon^2 - (N-1) \sum_{i=1}^{\infty} \rho^{2m(i+1)} (1-\ell)^{2i} \\ &> \epsilon^2 - \frac{(1-\ell)^2 (N-1)}{1 - (1-\ell)^2} \end{aligned}$$

Clearly, if the last term is positive, then also $f(\ell, m) > 0, \forall m$. This condition is satisfied if and only if

$$\ell \geq 1 - \sqrt{\frac{\epsilon^2}{N-1 + \epsilon^2}} = \bar{\ell}.$$

Therefore, $S(Q_1, \ell; m) - S(Q_2, \ell; m) > 0, \forall \ell > \bar{\ell}$ and consequently $J(Q_1, \ell; m) > J(Q_2, \ell; m)$, thus proving the claim of the theorem. \blacksquare

Remark 4: At first sight, Theorem 3.2 and Theorem 3.3 seem in contradiction. However, this can be explained by observing that $\|Q^m\|_F^2 = 1 + \alpha\rho^{2m}(Q) + o(\rho^{2m}(Q))$, where α is the algebraic multiplicity⁵ of the second largest eigenvalue. Therefore, for large m , minimizing the Frobenius norm of Q^m or the essential spectral radius of Q is almost equivalent.

IV. OPTIMAL GAIN ℓ FOR FIXED CONSENSUS MATRIX Q

In this section we assume that the consensus matrix Q is fixed. Hence the problem we want to solve is the following

$$\arg \min_{\ell \in (0,1)} J(Q, \ell; m) \tag{21}$$

The previous optimization problem is convex in ℓ . This fact can be easily checked by observing that the cost functional J can be written as sum of functions of the form:

$$g(\ell) = \frac{x\ell^2}{1 - x(1-\ell)^2}, \quad h(\ell) = \frac{x}{1 - x(1-\ell)^2}, \quad x \in [0, 1]$$

which are convex in $\ell \in (0, 1)$. Consider now a generic matrix $Q \in \mathcal{Q}$ and let

$$\ell^{opt}(Q; m) = \arg \min_{\ell \in (0,1)} J(Q, \ell; m).$$

Convexity of J allows easy computations of $\ell^{opt}(Q; m)$. In the remaining of this section we shall see that the sequence $\{\ell^{opt}(Q; m)\}_{m=0}^{\infty}$ is monotonically non-decreasing in m . Moreover, it is

⁵We assume here Q primitive, i.e. $\rho(Q) < 1$.

bounded below and above by ℓ_d^{opt} and ℓ_c^{opt} , which are the optimal gains minimizing J respectively when $Q = I$ and when $Q = \frac{1}{N}\mathbb{1}\mathbb{1}^*$, namely

$$\ell_d^{opt} = \arg \min_{\ell \in (0,1)} J(I, \ell; m), \quad \ell_c^{opt} = \arg \min_{\ell \in (0,1)} J\left(\frac{1}{N}\mathbb{1}\mathbb{1}^*, \ell; m\right)$$

Note that $Q = I$ and $Q = \frac{1}{N}\mathbb{1}\mathbb{1}^*$ represent the two extreme cases in modeling the flow of information between the sensors. Indeed, $Q = I$ corresponds to the situation in which the sensors do not communicate; in such a case there are N Kalman filters running separately (the subscript “d” in ℓ_d^{opt} means *decentralized*, i.e. no communication). In the other case, instead, we have that the underlying communication graph is complete and this means that each sensor has full knowledge about the estimates of all the other sensors (the subscript “c” in ℓ_c^{opt} means *centralized*, i.e. full communication). The following proposition characterizes precisely ℓ_d^{opt} and ℓ_c^{opt} .

Proposition 4.1: Let ℓ_d^{opt} and ℓ_c^{opt} be as above. Then

$$\ell_d^{opt} = \frac{-q + \sqrt{q^2 + 4qr}}{2r}, \quad \ell_c^{opt} = \frac{-q + \sqrt{q^2 + 4q\bar{r}}}{2\bar{r}}$$

where $\bar{r} = \frac{r}{N}$.

Proof: The proof follows from standard results known in Kalman filtering. ■

The following theorem shows also that ℓ_d^{opt} and ℓ_c^{opt} play the role of, respectively, lower and upper bounds for $\ell^{opt}(Q; m)$. Indeed, a stronger result can be obtained, which characterizes the sequence $\{\ell^{opt}(Q; m)\}_{m=0}^{\infty}$.

Theorem 4.1: Let $Q \in \mathcal{Q}$. Let $\ell^{opt}(Q; m)$ be defined as above. Then the following chain of inequalities holds true

$$\ell_d^{opt} = \ell^{opt}(Q; 0) \leq \ell^{opt}(Q; 1) \leq \dots \leq \ell^{opt}(Q; m) \leq \ell^{opt}(Q; m+1) \leq \dots \leq \ell^{opt}(Q; \infty) \leq \ell_c^{opt}.$$

The achieved (optimal) cost satisfies the reversed chain of inequalities:

$$\begin{aligned} J(Q, \ell_d^{opt}; 0) &\geq J(Q, \ell^{opt}(Q; 1); 1) \geq \dots \geq J(Q, \ell^{opt}(Q; m); m) \geq \\ &\geq J(Q, \ell^{opt}(Q; m+1); m+1) \geq \dots \geq J(Q, \ell_c^{opt}; \infty) \end{aligned}$$

Moreover $\ell^{opt}(Q; \infty) = \ell_c^{opt}$ if and only if Q is primitive [4].

Proof: Note that the equality $\ell^{opt}(Q; 0) = \ell_d^{opt}$ follows directly from the fact that $Q^0 = I$.

We prove now that $\ell^{opt}(Q; m) \leq \ell^{opt}(Q; m+1)$. We start by calculating the partial derivative

$$\frac{\partial J(Q, \ell; m)}{\partial \ell}(\ell) = 2 \left\{ \sum_{i=1}^{N-1} \frac{r\ell|\lambda_i|^{2m} - r\ell(1-\ell)|\lambda_i|^{4m}}{[1 - (1-\ell)^2|\lambda_i|^{2m}]^2} - \frac{Nq(1-\ell) - r\ell^2}{[1 - (1-\ell)^2]^2} \right\}$$

Consider now the function $p : [0, \infty) \rightarrow \mathbb{R}$ defined as

$$p(m) = \frac{x^{2m} - (1 - \ell)x^{4m}}{(1 - (1 - \ell)^2 x^{2m})^2},$$

where x is a real in $[0, 1]$. The first derivative of p is given by

$$\frac{\partial p}{\partial m}(m) = 2x^{2m} \frac{1 - (1 - \ell^2)x^{2m}}{[1 - (1 - \ell)^2 x^{2m}]^3} \log x.$$

It is easy to check that $\frac{\partial p}{\partial m}(m) < 0$. This implies that

$$\begin{aligned} \frac{\partial J(Q, \ell; m)}{\partial \ell} - \frac{\partial J(Q, \ell; m+1)}{\partial \ell} &= \\ &= 2r\ell \left\{ \sum_{i=0}^{N-1} \frac{|\lambda_i|^{2m} - (1 - \ell)|\lambda_i|^{4m}}{[1 - (1 - \ell)^2 |\lambda_i|^{2m}]^2} - \frac{|\lambda_i|^{2(m+1)} - (1 - \ell)|\lambda_i|^{4(m+1)}}{[1 - (1 - \ell)^2 |\lambda_i|^{2(m+1)}]^2} \right\} \geq 0, \end{aligned}$$

$\forall 0 < \ell < 1$ and $0 \leq x \leq 1$. Since we already know that $J(Q, \ell; m)$ is convex for all nonnegative m inside the interval $(0, 1)$ it follows that $\ell^{opt}(Q; m+1) > \ell^{opt}(Q; m)$.

In order to show that $\ell^{opt}(Q; m) \leq \ell_c^{opt}$, we remark that the matrix $\frac{1}{N}\mathbb{1}\mathbb{1}^*$ has an eigenvalue equal to 1 and $N - 1$ eigenvalues equal to 0. This implies that

$$\frac{\partial J(Q, \ell; m)}{\partial \ell} - \frac{\partial J(\frac{1}{N}\mathbb{1}\mathbb{1}^*, \ell; m+1)}{\partial \ell} = 2r\ell \sum_{i=1}^{N-1} \frac{|\lambda_i|^{2m} - (1 - \ell)|\lambda_i|^{4m}}{[1 - (1 - \ell)^2 |\lambda_i|^{2m}]^2} > 0$$

where the last inequality follows from the fact that $x - \alpha x^2 > 0$, $\forall x \in (0, 1)$ and $\forall \alpha \in (0, 1)$.

Therefore $\ell^{opt}(Q; m) \leq \ell_c^{opt}$.

We prove now $\ell^{opt}(Q; m) = \ell_c^{opt}$ if and only if Q is primitive. Standard results in Markov chains (see [4]) ensure that if Q is a primitive doubly stochastic matrix then $\lim_{k \rightarrow \infty} Q^k = \frac{1}{N}\mathbb{1}\mathbb{1}^*$.

Conversely, suppose that $Q \in \mathcal{Q}$ is not primitive, then there exists λ_i , $1 \leq i \leq N - 1$ such that $|\lambda_i| = 1$ implying that

$$\frac{\partial J(Q, \ell; m)}{\partial \ell} - \frac{\partial J(\frac{1}{N}\mathbb{1}\mathbb{1}^*, \ell; m+1)}{\partial \ell} \geq \frac{2r\ell^2}{[1 - (1 - \ell)^2]^2} > 0.$$

From this last inequality $\ell^{opt}(Q; \infty) < \ell_c^{opt}$ follows. It remains now only to verify the chain of inequalities on the cost functional. To this purpose note that the function $g : [0, 1] \rightarrow \mathbb{R}$ defined as

$$g(x) = \frac{x}{1 - \alpha x}, \quad \alpha \in (0, 1)$$

is a monotonically increasing function of x . Therefore, since $x^{m+1} \leq x^m$ for $0 \leq x \leq 1$ and using the definition of J in (13), we have that

$$J(Q, \ell, m+1) \leq J(Q, \ell, m), \quad \forall \ell \in (0, 1).$$

Therefore

$$J(Q, \ell^{opt}(Q; m+1), m+1) \leq J(Q, \ell^{opt}(Q; m+1), m) \leq J(Q, \ell^{opt}(Q; m), m).$$

This concludes the proof. ■

V. JOINT OPTIMIZATION OF Q AND ℓ : SPECIAL CASES

We have shown in the previous two sections that the cost functional J is a convex function, both in $Q \in \mathcal{Q}_{sym}$ for ℓ fixed and in $\ell \in (0, 1)$ for Q fixed. Unfortunately, as simple examples demonstrate, J is not a convex function jointly in ℓ and $Q \in \mathcal{Q}_{sym}$. Therefore, the joint minimization of J

$$(Q^{opt}(m, r, q), \ell^{opt}(m, r, q)) \in \arg \min_{\ell \in (0,1); Q \in \mathcal{Q}_{sym}} J(Q, \ell; m, r, q) \quad (22)$$

results to be quite hard in general. Nevertheless, an analytical characterization is possible when restricting to some “asymptotic cases” on the values of m , r and q . In particular we will consider the following situations:

- the sensors can communicate arbitrarily fast within two subsequent measurements, i.e., $m \rightarrow \infty$
- $\frac{r}{q} \approx 0$, i.e. the variance of the measurement noise is negligible with respect to the variance of the process
- $\frac{q}{r} \approx 0$, i.e. the variance of the process is negligible with respect to the variance of the measurement noise

First note that $Q^{opt}(m, r, q)$ and $\ell^{opt}(m, r, q)$ are indeed only functions of m and r/q . In the sequel, without risk of confusion, we shall omit arguments which are kept fixed.

A. Fast communication ($m \rightarrow \infty$)

The results of this section parallel those of section III-B. Indeed it will be shown that when $m \rightarrow \infty$, optimizing Q for “fast convergence”, i.e. minimizing the essential spectral radius is the “right” thing to do. Moreover, as expected, the optimal gain converges to the centralized gain ℓ_c^{opt} .

Theorem 5.1: Let $Q^{opt}(m)$, $\ell^{opt}(m)$ be a solution of (22). Then

$$\lim_{m \rightarrow \infty} \rho(Q^{opt}(m)) = \min_{Q \in \mathcal{Q}} \rho(Q).$$

and

$$\lim_{m \rightarrow \infty} \ell^{opt}(m) = \ell_c^{opt}.$$

Moreover, if $\arg \min_{Q \in \mathcal{Q}} \rho(Q)$ is a singleton, then also

$$\lim_{m \rightarrow \infty} Q^{opt}(m) = \arg \min_{Q \in \mathcal{Q}} \rho(Q).$$

Proof: We start by introducing the following quantity that will be useful throughout the proof: given any $Q \in \mathcal{Q}$ we define

$$J^{opt}(Q, m) \triangleq J(Q, \ell^{opt}(Q, m); m) \quad (23)$$

where $\ell^{opt}(Q, m) \in \arg \min_{\ell} J(Q, \ell; m)$. Note that $Q^{opt}(m) \in \arg \min_{Q} J^{opt}(Q, m)$.

Consider now a sequence $\{\rho(Q^{opt}(m))\}_{m=0}^{\infty}$. Let \bar{Q} be any matrix minimizing the essential spectral radius inside \mathcal{Q} . We shall now prove that

$$\lim_{m \rightarrow \infty} \rho(Q^{opt}(m)) = \rho(\bar{Q}) \quad (24)$$

For notational convenience, we denote along the proof $\rho(Q^{opt}(m))$ by ρ_m and $\rho(\bar{Q})$ by $\bar{\rho}$. Assume by contradiction that $\lim_{m \rightarrow \infty} \rho_m \neq \bar{\rho}$. This means that there exists $\epsilon > 0$ and a sequence of integers $m_1 < m_2 < m_3 < \dots$, such that $\rho_{m_i} \geq \bar{\rho} + \epsilon$, $\forall i \in \mathbb{N}$ and consequently that $\inf \{\rho_{m_i}\} \geq \bar{\rho} + \epsilon$. This implies, by Lemma 7.1 (see the Appendix), that there exist \bar{m}_ϵ , depending only on ϵ , such that $J^{opt}(Q^{opt}(m_i), m) > J^{opt}(\bar{Q}, m)$, $\forall i \in \mathbb{N}$ and $\forall m > \bar{m}_\epsilon$. Therefore, if \bar{i} is such that $m_{\bar{i}} > \bar{m}_\epsilon$, we get

$$J^{opt}(Q^{opt}(m_i), m_i) > J^{opt}(\bar{Q}, m_i), \quad \forall i \geq \bar{i},$$

contradicting the fact that $Q^{opt}(m_i) \in \arg \min_{Q} J^{opt}(Q, m_i)$ and thus proving (24).

Consider now any sequence of integers $\tilde{m}_1 < \tilde{m}_2 < \tilde{m}_3 < \tilde{m}_4 \dots$ such that $\lim_{i \rightarrow \infty} Q^{opt}(\tilde{m}_i) = \tilde{Q}$ where \tilde{Q} is a suitable matrix inside \mathcal{Q} . It follows, by the continuity of ρ that $\lim_{i \rightarrow \infty} \rho(Q^{opt}(\tilde{m}_i)) = \rho(\tilde{Q})$. Clearly $\rho(\tilde{Q}) = \bar{\rho}$. Suppose now that $\arg \min_{Q \in \mathcal{Q}} \rho(Q)$ is a singleton and call \bar{Q} the unique element in this set. It follows that $\tilde{Q} = \bar{Q}$, thus implying, for the compactness of \mathcal{Q} , that

$$\lim_{m \rightarrow \infty} Q^{opt}(m) = \arg \min_{Q \in \mathcal{Q}} \rho(Q).$$

Consider now a sequence $\{\ell^{opt}(m)\}_{m=0}^{\infty}$ and let $\tilde{\rho}$ be such that $\bar{\rho} < \tilde{\rho} < 1$. Let us introduce a matrix \tilde{Q} such that $\sigma(\tilde{Q}) = \{1, \tilde{\rho}, \dots, \tilde{\rho}\}$, that is \tilde{Q} has $N - 1$ eigenvalues equal to $\tilde{\rho}$. It follows that

$$J(\tilde{Q}, \ell; m) = \frac{r\ell^2 + qN}{1 - (1 - \ell)^2} + (N - 1) \frac{r\ell^2 \tilde{\rho}^{2m}}{1 - (1 - \ell)^2 \tilde{\rho}^{2m}},$$

and, by recalling the expression of $\frac{\partial J(Q, \ell; m)}{\partial \ell}$, that

$$\frac{\partial J(\tilde{Q}, \ell; m)}{\partial \ell} = 2 \left\{ r\ell(N-1) \frac{\tilde{\rho}_m^{2m} - (1-\ell)\tilde{\rho}_m^{4m}}{[1 - (1-\ell)^2 \tilde{\rho}_m^{2m}]} - \frac{Nq(1-\ell) - r\ell^2}{[1 - (1-\ell)^2]^2} \right\}$$

Let now ϵ be any real such that $0 < \epsilon < \tilde{\rho} - \bar{\rho}$. We have just proved that there exists a positive integer \bar{m}_ϵ such that $\forall m > \bar{m}_\epsilon$ we have that $\rho_m < \tilde{\rho} - \epsilon$. Consider now the function $f : [0, 1] \rightarrow \mathbb{R}$ defined as

$$f(x) = \frac{x^{2m} - \alpha x^{4m}}{(1 - \alpha^2 x^{2m})^2},$$

where $m \in \mathbb{N}$ and where $\alpha \in (0, 1)$. By straightforward calculations we have that

$$\frac{\partial f(x)}{\partial \ell} = \frac{2mx^{2m}(1 + \alpha^2 x^{2m} - 2\alpha x^{2m})}{x(1 - \alpha^2 x^{2m})^3}.$$

It is possible to show that $\frac{\partial f(x)}{\partial \ell} > 0$, $\forall x \in (0, 1]$. Since, by the definition of essential spectral radius, we have that all the eigenvalues of $Q^{opt}(m)$ different from 1, are, in absolute value, smaller than ρ_m and therefore of $\tilde{\rho}$, it follows that

$$\frac{\partial J(Q(m), \ell; m)}{\partial \ell} \leq \frac{\partial J(\tilde{Q}, \ell; m)}{\partial \ell}, \quad \forall \ell \in (0, 1). \quad (25)$$

Let now $\tilde{\ell}^{opt}(m) = \arg \min_{\ell \in (0, 1)} J(\tilde{Q}, \ell; m)$. From Theorem 4.1 we have that $\lim_{m \rightarrow \infty} \tilde{\ell}^{opt}(m) = \ell_c^{opt}$. On the other hand, by the convexity of J on ℓ and by (25) it follows that $\tilde{\ell}^{opt}(m) \leq \ell^{opt}(m)$, $\forall m > \bar{m}_\epsilon$. Therefore $\lim_{m \rightarrow \infty} \ell^{opt}(m) = \ell_c^{opt}$. ■

B. Small measurement noise ($r/q \rightarrow 0$)

In this subsection we treat the case in which the variance of the measurement noise is negligible with respect of the variance of the process, that is $r/q \rightarrow 0$. This parallels the case analyzed in Section III-C. Also here, as in Section III-C, it is the Frobenius norm of Q which plays a crucial role.

Indeed, while in Section III-C the gain was *assumed* to be large, here ℓ is shown to converge to 1 as r/q goes to zero.

Theorem 5.2: Let $Q^{opt}(r/q)$, $\ell^{opt}(r/q)$ be a solution of (22) and let

$$\bar{Q} \in \arg \min_{Q \in \mathcal{Q}} \|Q^m\|_F.$$

Then

$$\lim_{r/q \rightarrow 0} \|(Q^{opt}(r/q))^m\|_F = \|\bar{Q}^m\|_F.$$

Moreover

$$\ell^{opt}(r/q) = 1 - \frac{\|\bar{Q}\|_F^2 r}{N} + o(r/q).$$

In addition if $\arg \min_{Q \in \mathcal{Q}} \|Q^m\|_F$ is a singleton also

$$\lim_{r/q \rightarrow 0} Q^{opt}(r/q) = \bar{Q}.$$

holds.

Proof: We start by introducing the following quantity that will be useful throughout the proof: given any $Q \in \mathcal{Q}$ we define

$$J^{opt}(Q; r/q) \triangleq J(Q, \ell^{opt}(Q, r/q); r, q) \quad (26)$$

where $\ell^{opt}(Q, r/q) \in \arg \min_{\ell} J(Q, \ell; m, r, q)$. Note that $Q^{opt}(r/q) \in \arg \min_{Q} J^{opt}(Q, r/q)$.

We shall now prove that

$$\lim_{r/q \rightarrow 0} \|(Q^{opt}(r/q))^m\|_F = \|\bar{Q}^m\|_F. \quad (27)$$

Assume by contradiction that $\lim_{r/q \rightarrow 0} \|(Q^{opt}(r/q))^m\|_F \neq \|\bar{Q}^m\|_F$. This means that there exists $\epsilon > 0$ and a sequence $\frac{r_1}{q_1} > \frac{r_2}{q_2} > \frac{r_3}{q_3} > \dots$, such that $\|(Q^{opt}(r_i/q_i))^m\|_F \geq \|\bar{Q}^m\|_F + \epsilon$, $\forall i \in \mathbb{N}$ and consequently that $\inf \{\|(Q^{opt}(r_i/q_i))^m\|_F\} \geq \|\bar{Q}^m\|_F + \epsilon$.

This implies, by Lemma 7.2 (see the Appendix), that there exists δ_ϵ , depending only on ϵ , such that

$$J^{opt}(Q^{opt}(r_i/q_i), r/q) > J^{opt}(\bar{Q}, r/q),$$

for each pair r, q such that $\frac{r}{q} < \delta_\epsilon$ and $\forall i \in \mathbb{N}$. Therefore, if \bar{i} is such that $\frac{r_{\bar{i}}}{q_{\bar{i}}} < \delta_\epsilon$, we get

$$J^{opt}(Q^{opt}(r_{\bar{i}}/q_{\bar{i}}), r_{\bar{i}}/q_{\bar{i}}) > J^{opt}(\bar{Q}, r_{\bar{i}}/q_{\bar{i}}), \quad \forall i \geq \bar{i},$$

contradicting the fact that $Q^{opt}(r_i/q_i) \in \arg \min_{Q} J^{opt}(Q, r_i/q_i)$ and thus proving (27).

Consider now any sequence $\frac{r_{i_1}}{q_{i_1}} > \frac{r_{i_2}}{q_{i_2}} > \frac{r_{i_3}}{q_{i_3}} > \dots$, such that $\lim_{h \rightarrow \infty} Q^{opt}(r_{i_h}/q_{i_h}) = \tilde{Q}$ where \tilde{Q} is a suitable matrix inside \mathcal{Q} . It follows that $\lim_{h \rightarrow \infty} \|(Q^{opt}(r_{i_h}/q_{i_h}))^m\|_F = \|(\tilde{Q})^m\|_F$. Clearly $\|(\tilde{Q})^m\|_F = \|\bar{Q}^m\|_F$. Suppose now that $\arg \min_{Q \in \mathcal{Q}} \|Q^m\|_F$ is a singleton and call \bar{Q} the unique element in this set. It follows that $\tilde{Q} = \bar{Q}$, thus implying, for the compactness of \mathcal{Q} , that

$$\lim_{r/q \rightarrow 0} Q^{opt}(r/q) = \bar{Q} = \arg \min_{Q \in \mathcal{Q}} \|Q^m\|_F$$

Consider now $\ell^{opt}(r/q)$. Preliminarily we observe that $\lim_{r/q \rightarrow 0} \ell_d^{opt} = 1$. Since by Theorem 4.1 we have that $\ell^{opt}(Q, r/q) \geq \ell_d^{opt}$, it follows that also $\lim_{r/q \rightarrow 0} \ell^{opt}(Q, r/q) = 1$ and $\lim_{r/q \rightarrow 0} \ell^{opt}(r/q) = 1$. Using the Implicit Function Theorem it follows that $\ell^{opt}(Q, r/q)$ is differentiable around $r/q = 0$, i.e.

$$\ell^{opt}(Q, r/q) = 1 + \alpha(Q) \frac{r}{q} + o\left(\frac{r}{q}\right)$$

where the coefficient $\alpha(Q)$ is given by

$$\alpha(Q) = - \left[\left(\frac{\partial^2 J}{\partial \ell^2} \right)^{-1} \frac{\partial^2 J}{\partial \ell \partial r/q} \right]_{|\ell=1, Q, r/q=0} = - \frac{\|Q^m\|_F^2}{N}$$

Since $\ell^{opt}(r/q) = \ell^{opt}(Q^{opt}(r/q), r/q)$ then

$$\ell^{opt}(r/q) = 1 + \alpha(Q^{opt}(r/q)) r/q + o(r/q).$$

Using the fact that

$$\lim_{r/q \rightarrow 0} \|Q^{opt}(r/q)\|_F = \|\bar{Q}\|_F$$

then

$$\ell^{opt}(r/q) = 1 + \alpha(\bar{Q}) r/q + o(r/q),$$

concluding the proof. ■

Remark 5: It is interesting to observe that the communication graph \mathcal{G} determines the coefficient of the first order expansion of ℓ in r/q . Note that, in the extreme cases $Q^{opt} = I$ (no communication, decentralized estimation) and $Q^{opt} = \frac{\mathbf{1}\mathbf{1}^*}{N}$ (communication graph fully connected, centralized estimation) one recovers, respectively, $\ell_d^{opt} = 1 - \frac{r}{q} + o(\frac{r}{q})$ and $\ell_c^{opt} = 1 - \frac{1}{N} \frac{r}{q} + o(\frac{r}{q})$, which can also be easily obtained from the standard expressions of ℓ_d^{opt} and ℓ_c^{opt} .

C. High measurement noise ($q/r \rightarrow 0$)

Similarly to the previous section, we now consider the other limiting case for $q/r \approx 0$. The following result holds.

Theorem 5.3: Let $Q^{opt}(q/r)$, $\ell^{opt}(q/r)$ be defined as above and denote with $p(Q)$ the number of eigenvalues of Q on the unit circle. Then

$$\lim_{q/r \rightarrow 0} p(Q^{opt}(q/r)) = \min_{Q \in \mathcal{Q}} p(Q) =: p^{opt}.$$

Moreover

$$\ell^{opt}(q/r) = \sqrt{\frac{N}{p^{opt}}} \sqrt{\frac{q}{r}} + o\left(\sqrt{q/r}\right).$$

Proof: We start by observing that $\lim_{q/r \rightarrow 0} \ell_d^{opt} = 0$ and $\lim_{q/r \rightarrow 0} \ell_c^{opt} = 0$. More precisely it is possible to see $\ell_d^{opt} = \sqrt{\frac{q}{r}} + o\left(\frac{q}{r}\right)$ and $\ell_c^{opt} = \sqrt{N} \sqrt{\frac{q}{r}} + o\left(\frac{q}{r}\right)$. Define

$$\ell^{opt}(Q, q/r) \triangleq \arg \min_{\ell \in (0,1)} J(Q, \ell; r, q).$$

From the previous observation and from Theorem 4.1 it follows that

$$\lim_{q/r \rightarrow 0} \ell^{opt}(Q, q/r) = 0. \quad (28)$$

Assume now by contradiction that $\lim_{q/r \rightarrow 0} p(Q^{opt}(q/r)) \neq p^{opt}$. This means that there exists a sequence $\frac{q_1}{r_1} > \frac{q_2}{r_2} > \frac{q_3}{r_3} > \dots$, such that $p(Q^{opt}(q/r)) \geq p^{opt} + 1$, $\forall i \in \mathbb{N}$ and consequently that $\inf \{p(Q^{opt}(q_i/r_i))\} \geq p^{opt} + 1$. Let us rewrite J in the following way

$$\begin{aligned} J(Q, \ell; r, q) &= \frac{qN + r\ell^2 p(Q)}{1 - (1 - \ell)^2} + r\ell^2 \sum_{i: |\lambda_i| \neq 1} \frac{|\lambda_i|^{2m}}{1 - (1 - \ell)^2 |\lambda_i|^{2m}} \\ &= \frac{qN + r\ell^2 p(Q)}{1 - (1 - \ell)^2} + \frac{r\ell^2}{1 - (1 - \ell)^2} f(\ell) \end{aligned}$$

where $f(\ell) = (1 - (1 - \ell)^2) \sum_{i: |\lambda_i| \neq 1} \frac{|\lambda_i|^{2m}}{1 - (1 - \ell)^2 |\lambda_i|^{2m}}$. Note that $\lim_{\ell \rightarrow 0} f(\ell) = 0$. Hence

$$J(Q, \ell; r, q) = \frac{qN + r\ell^2 p(Q)}{1 - (1 - \ell)^2} + o\left(\frac{r\ell^2}{1 - (1 - \ell)^2}\right) \quad (29)$$

Let now \bar{Q} be such that $p(\bar{Q}) = p^{opt}$. From (29) it follows that

$$J(Q, \ell; r, q) - J(\bar{Q}, \ell; r, q) = (p(Q) - p^{opt}) \frac{r\ell^2}{1 - (1 - \ell)^2} + o\left(\frac{r\ell^2}{1 - (1 - \ell)^2}\right)$$

Equation (28) implies that also $\lim_{i \rightarrow \infty} \ell^{opt}(Q, q_i/r_i) = 0$ so that for i large enough

$$J(Q^{opt}(q_i/r_i), \ell^{opt}(q_i/r_i); r_i, q_i) > J(\bar{Q}, \ell^{opt}(q_i/r_i); r_i, q_i)$$

contradicting the fact that $Q^{opt}(q_i/r_i) \in \arg \min_{Q \in \mathcal{Q}} J(Q, \ell^{opt}(q_i/r_i); r_i, q_i)$.

Consider now $\ell^{opt}(q/r)$. Let us first compute the partial derivative

$$\frac{\partial J(Q^{opt}(q/r), \ell; r, q)}{\partial \ell} = \frac{r \left(2p(Q^{opt}(q/r))\ell^2 - 2\gamma^2 N + 2\gamma^2 N\ell + 2\ell^2 f(\ell) + 2\ell^3 \dot{f}(\ell) - \ell^4 \ddot{f}(\ell) \right)}{(1 - (1 - \ell)^2)}$$

where $\gamma = \sqrt{\frac{q}{r}}$. Let now $F(\gamma, \ell) = 2p(Q^{opt}(q/r))\ell^2 - 2\gamma^2 N + 2\gamma^2 N\ell + 2\ell^2 f(\ell) + 2\ell^3 \dot{f}(\ell) - \ell^4 \ddot{f}(\ell)$ and consider the equation $F(\gamma, \ell) = 0$. We adopt an argument similar to the proof of the Implicit Function Theorem. By applying the Taylor's Theorem around $(0, 0)$ we get

$$F(\gamma, \ell) = -4N\gamma^2 + 4p(Q^{opt}(q/r))\ell^2 + o(\|\gamma^2\| + \|\ell^2\|).$$

and equating this expression to 0 we obtain that

$$\begin{aligned} \ell^{opt}(Q^{opt}(q/r), q/r) &= \gamma \sqrt{\frac{N}{p(Q^{opt}(q/r))}} + o(\gamma) \\ &= \sqrt{\frac{N}{p(Q^{opt}(q/r))}} \sqrt{\frac{q}{r}} + o\left(\sqrt{\frac{q}{r}}\right) \end{aligned}$$

Therefore, using an argument similar to the one adopted in the proof of Theorem 5.2, it follows that

$$\ell^{opt}(r/q) = \sqrt{\frac{N}{p^{opt}}} \sqrt{\frac{q}{r}} + o\left(\sqrt{\frac{q}{r}}\right),$$

which concludes the proof. ■

Remark 6: Also in this case the constant in the first order expansion of the optimal gain as a function of q/r depends on the communication graph; the extreme cases are, respectively, $\ell_c^{opt}(q/r) = \sqrt{N} \sqrt{q/r} + o(\sqrt{q/r})$ and $\ell_d^{opt}(q/r) = \sqrt{q/r} + o(\sqrt{q/r})$. In fact, if sensors cannot communicate, then necessarily $Q^{opt} = I$, therefore $\ell^{opt}(q/r) = \ell_d^{opt}(q/r)$, while if the communication graph is fully connected, then $Q^{opt} = \frac{1}{N} \mathbf{1}\mathbf{1}^*$, therefore $\ell^{opt}(q/r) = \ell_c^{opt}(q/r)$.

The previous theorem shows also that for $q \ll r$ the optimizing Q , while being consistent with the communication graph, has to minimize the number of unitary eigenvalues.

VI. ILLUSTRATIVE EXAMPLES

In this section we provide some examples illustrating the approach proposed in this paper. In particular in Example 2 we solve analytically the minimization problem formulated in Section III. In Example 3 and in Example 4 we provide a numerical comparison between the approach presented here and the method proposed in [1]. In Example 5, inspired by [23], we propose a new scheme for updating the estimates which seems to be very promising in order to improve the performance. In particular we allow the sensors to keep in memory also the previously received estimates, and not only the current one, and to use them to build an updated estimate. More precisely the estimate $\hat{x}(t + (h + 1)\delta)$ will be a suitable weighted combination of the

current estimate $\hat{x}(t + h\delta)$ and of the previous estimate $\hat{x}(t + (h - 1)\delta)$. It has been shown by Muthukrihnan et al. [23] in the context of load balancing, and by Cao et al.[7] in the context of gossip algorithms, that the use of memory permits to speed up convergence and to improve performance. We show here, by means of simulative results, that for the set of matrices considered in Example 2 and 3 the presence of one level of memory permits to reach better performance in terms of the cost function J .

Example 2: Assume that \mathcal{Q} is the set of the circulant stochastic matrices of the form

$$Q_k = \begin{bmatrix} 1-2k & k & 0 & 0 & \cdots & 0 & k \\ k & 1-2k & k & 0 & \cdots & 0 & 0 \\ 0 & k & 1-2k & k & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ k & 0 & 0 & 0 & \cdots & k & 1-2k \end{bmatrix} \quad (30)$$

where k is a real in $[0, \frac{1}{2}]$ and assume that $m = 1$. We want to solve the minimization problem considered in Section III, i.e.

$$Q^{opt} = \arg \min_{Q \in \mathcal{Q}} J = \arg \min_{Q \in \mathcal{Q}} \sum_{i=1}^{N-1} \frac{|\lambda_i|^{2m}}{1 - (1 - \ell)^2 |\lambda_i|^{2m}}. \quad (31)$$

Unfortunately, even if \mathcal{Q} exhibits a particular structure, it is not possible to give, in general, a closed form expression for the optimizing matrix Q^{opt} , which has to be found using numerical search techniques.

However, similarly to Example 1, a more detailed study of (31) can be provided when the number of sensors N goes to infinity. To this purpose it is convenient to consider the normalized cost:

$$\bar{J}_N := \frac{1}{N} \sum_{i=0}^{N-1} \frac{|\lambda_i|^2}{1 - (1 - \ell)^2 |\lambda_i|^2}$$

and the function $f : \mathbb{C} \rightarrow \mathbb{C}$

$$f(z) = 1 - 2k + kz + kz^{-1}.$$

The eigenvalues of Q_k can be expressed as $\lambda_h = 1 - 2k + ke^{j\frac{2\pi}{N}h} + ke^{-j\frac{2\pi}{N}h}$ (see e.g. [11]). Note that $\lambda_h = f\left(e^{j\frac{2\pi}{N}h}\right)$. As the number of agents grows, the normalized cost J_N converges to:

$$\lim_{N \rightarrow \infty} J_N = \bar{J}_\infty = \frac{1}{2\pi} \int_0^{2\pi} \frac{|f(e^{j\phi})|^2}{1 - (1 - \ell)^2 |f(e^{j\phi})|^2} d\phi = \frac{1}{2\pi j} \oint_\gamma \frac{z^{-1} f^2(z)}{1 - (1 - \ell)^2 f^2(z)} dz$$

where γ is the unit circle. It is straightforward to see that the poles of $\frac{z^{-1}f^2(z)}{1-(1-\ell)^2f^2(z)}$ inside γ are $z_1 = \frac{\ell+2(1-\ell)k-\sqrt{\ell^2+4\ell(1-\ell)k}}{2(1-\ell)k}$, $z_2 = \frac{\ell-2-2(1-\ell)k+\sqrt{(2-\ell)^2+4(1-\ell)(2-\ell)k}}{2(1-\ell)k}$ and $z_3 = 0$; the integral can be computed explicitly using the residue theorem yielding:

$$\bar{J}_\infty = -\frac{1}{(1-\ell)^2} + \frac{1}{2(1-\ell)^2} \left(\frac{1}{\sqrt{\ell^2+4\ell(1-\ell)k}} - \frac{1}{\sqrt{(2-\ell)^2-4(1-\ell)(2-\ell)k}} \right)$$

By equating the first derivative of \bar{J}_∞ to zero, we get that Q^{opt} tends to the matrix $Q_{k^{opt}}$ for $N \rightarrow \infty$, where $k^{opt} = \frac{2-2\ell+\sqrt[3]{(2-\ell)\ell^2-\sqrt[3]{\ell(2-\ell)^2}}}{4(1-\ell)}$.

Example 3: In [1] the authors analyze a general MIMO scenario where the gain $\ell = \ell(t)$ (K in their terminology) and the consensus matrix $Q = Q(t)$ are time varying matrices which are chosen recursively at each time step. In order to compare the results in [1] with our approach we assume that the averaging matrix W in [1] corresponds to performing m consensus iterations using the matrix Q , i.e. $W = Q^m$. In [1] the gain ℓ is chosen to minimize the estimation error covariance of the local estimators (i.e. in a decentralized fashion) and it is different for each sensor. Moreover the consensus matrix Q is chosen so that the estimation error covariance of the local estimators is minimized after consensus (weighted averaging in [1])⁶.

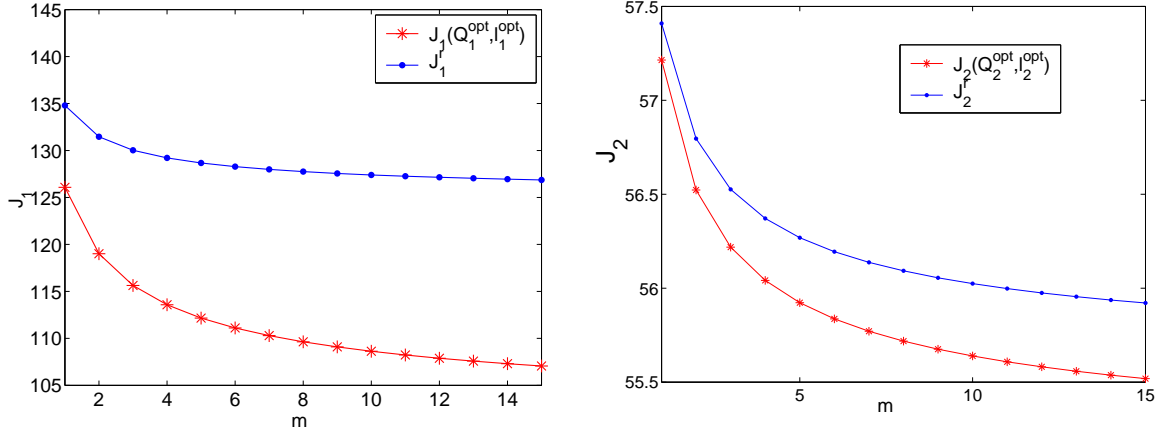


Fig. 1. Comparison between $J_1(Q_1^{opt}(m), \ell_1^{opt}(m), m)$, and $J_1^I(m)$ (left). Comparison between $J_2(Q_2^{opt}(m), \ell_2^{opt}(m), m)$, and $J_2^I(m)$ (right).

In the simulation reported in this example, we assume that $N = 100$, $q = 1$ and $r = 1$. Moreover we assume that Q is the same set of circulant matrices defined in the previous example.

⁶Technically, the approach proposed in [1] was applied only for the case $m = 1$.

We consider the minimization of both J_1 and J_2 , with J_1, J_2 defined as in the Section II. We use the following notational conventions: $Q_1^{opt}(m), \ell_1^{opt}(m)$ and $Q_2^{opt}(m), \ell_2^{opt}(m)$ are the optimal consensus matrices and the optimal gains respectively for J_1 and J_2 obtained by solving numerically the problem formulated at the end of Section II, given by:

$$\begin{aligned} (Q_1^{opt}(m), \ell_1^{opt}(m)) &\in \arg \min_{\ell \in (0,1), Q \in \mathcal{Q}} J_1(Q, \ell; m, r, q) \\ (Q_2^{opt}(m), \ell_2^{opt}(m)) &\in \arg \min_{\ell \in (0,1), Q \in \mathcal{Q}} J_2(Q, \ell; m, r, q) \end{aligned}$$

As mentioned earlier, in [1] the optimal gain and the optimal consensus matrix are found *recursively* at each time step t . The fact that the consensus matrix is of the form (30) implies that the gain is the same for all the sensors: we denote it by $\ell^r(t, m)$. Moreover $Q^r(t, m)$ represents the optimal consensus matrix (the superscript "r" means *recursively*). The asymptotic values (in t) of $\ell^r(t, m)$ and $Q^r(t, m)$ are denoted respectively by $\ell^r(m)$ and $Q^r(m)$, i.e. $\lim_{t \rightarrow \infty} \ell^r(t, m) = \ell^r(m)$ and $\lim_{t \rightarrow \infty} Q^r(t, m) = Q^r(m)$. Similarly, we indicate by $J_1^r(m)$ and $J_2^r(m)$ the asymptotic cost values⁷ to which J_1 and J_2 converge using the method proposed in [1]. Finally, since the set \mathcal{Q} is parameterized by k , we shall identify a matrix Q_k with the corresponding value of the parameter k . Hence, we use $k_1^{opt}(m), k_2^{opt}(m)$ and $k^r(m)$ in lieu of $Q_1^{opt}(m), Q_2^{opt}(m)$ and $Q^r(m)$. Simulations for m ranging in the interval $[1, 15]$ are shown in Figures 1 and 2.

From left panel of Figure 2 we can see that the iterative (local) optimization proposed in [1] converges to values of the parameters $\ell^r(m)$ which are different from the optimal values obtained minimizing the asymptotic cost proposed in this paper. Indeed, the recursive approach seems to give a worse performance than the approach proposed in this paper both in terms of the performance costs J_1 and J_2 , as shown in Figure 1. Moreover, for large m , the gain $\ell_1^{opt}(m)$ converges to the optimal centralized gain ℓ_c^{opt} , as shown in Section IV.

Note also from right panel of Figure 2 that the optimal consensus parameter k of the matrix Q_k for small m is close to the value that minimizes the Frobenius norm $\|Q\|_F$, while for large m it converges to the value that minimizes the spectral radius $\rho(Q)$, as shown in Section III.

Example 4: In this simulation we consider a strongly connected random geometric graph generated by choosing N points at random in the unit square, and then placing an edge between each pair of points at distance less than 0.3. We assume that $N = 30$, $q = 1$ and $r = 1$. It is

⁷There is no proof of convergence in [1]; however this is observed experimentally.

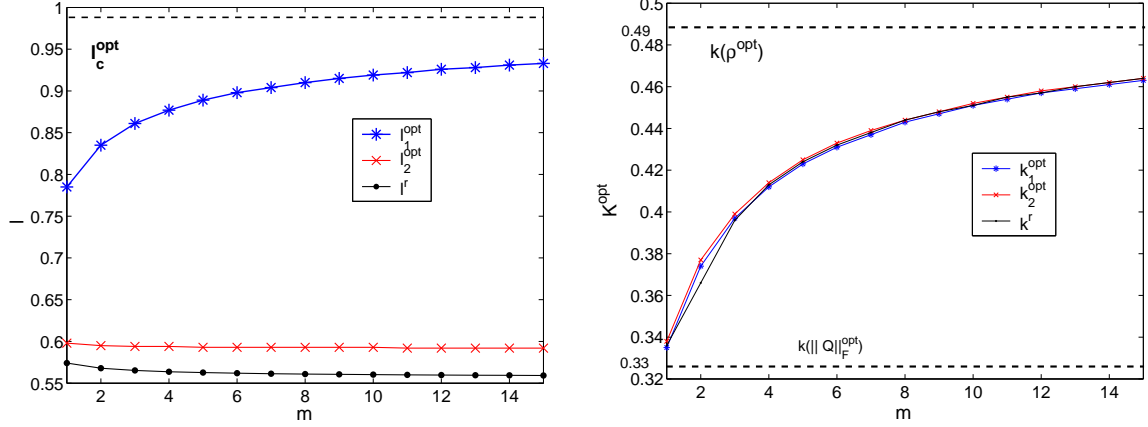


Fig. 2. Optimal gains (*left*) and consensus matrices (*right*) as a function of number of exchanged messages m . Optimal consensus matrices Q are parameterized by k . In the right panel we also report the values $k(\rho^{\text{opt}})$, which minimizes the spectral radius $\rho(Q)$ and $k(\|Q\|_F^{\text{opt}}$ which minimizes the Frobenius norm $\|Q\|_F$.

worth noticing that in this case differently from the previous one, the gains, calculated by the method proposed in [1] recursively at each instant time t , are in general different for each sensor. Hence we report here only the comparison between the cost functions. We run simulations for m ranging in the interval $[1, 10]$. The results obtained are depicted in Figure 3. The notation is the same used previously. Note that $\ell_1^{\text{opt}}(m), Q_1^{\text{opt}}(m)$ and $\ell_2^{\text{opt}}(m), Q_2^{\text{opt}}(m)$ yield again values

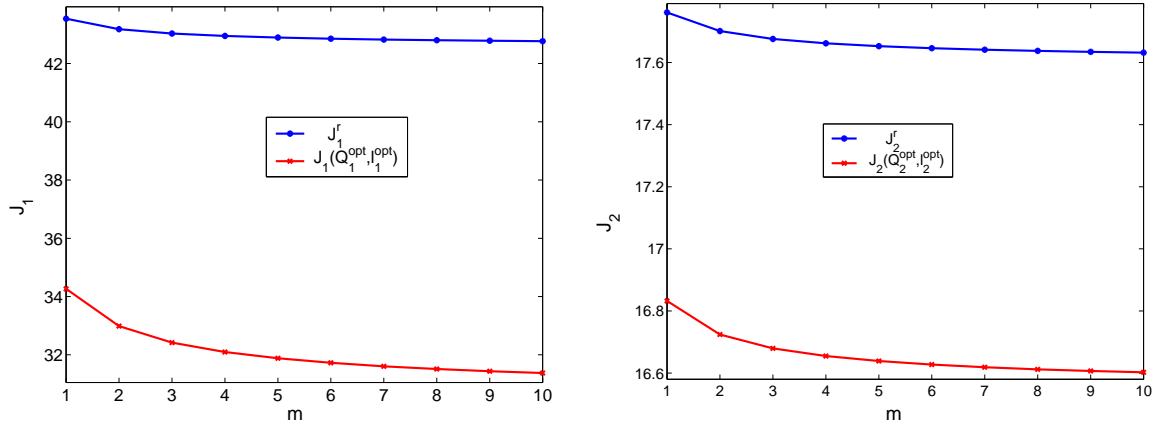


Fig. 3. Comparison between $J_1(Q_1^{\text{opt}}(m), \ell_1^{\text{opt}}(m), m)$, $J_1(Q_2^{\text{opt}}(m), \ell_2^{\text{opt}}(m), m)$ and $J_1^r(m)$ (*left*). Comparison between $J_2(Q_2^{\text{opt}}(m), \ell_2^{\text{opt}}(m), m)$, $J_2(Q_1^{\text{opt}}(m), \ell_1^{\text{opt}}(m), m)$ and $J_2^r(m)$ (*right*).

of J_1 and J_2 which are better respectively than J_1^r and J_2^r .

Remark 7: It is worth remarking that the optimization strategy proposed in [1] gives worse performance in terms of the asymptotic values for both costs J_1 and J_2 in both the previous examples. This is somewhat to be expected since the former approach recursively optimize the cost at the next time step, not the steady state cost. This recursive strategy gives the optimal steady state performance only for the centralized scenario, i.e. when the communication graph is fully connected, as well known from any standard textbook on optimal estimation and Kalman filtering [9]. When the graph is not fully connected, this strategy is not guaranteed to converge to the optimal value, as indeed shown in the previous numerical examples.

Example 5: Inspired by the work of Muthukrishnan et al. [23], in this example we propose the following scheme of updating the estimate $\hat{x}(k + h\delta|k)$ which adds a memory buffer,

$$\hat{x}(k + (h + 1)\delta|k) = Q\hat{x}(k + h\delta|k) \quad \text{if } h = 0,$$

and

$$\hat{x}(k + (h + 1)\delta|k) = \nu Q\hat{x}(k + h\delta|k) + (1 - \nu)\hat{x}(k + (h - 1)\delta|k) \quad \text{if } h = 1, \dots, m-1 \quad (32)$$

where ν is a weighting parameter. It has been proved in [23], that, when we are dealing with a consensus algorithm, if $\nu \notin [0, 2]$ then (32) does not yield the convergence. Let $\tilde{x}(k + h\delta|k) = x(k + h\delta)\mathbf{1} - \hat{x}(k + h\delta|k)$ denote again the estimation error and let us introduce, for $1 \leq h \leq m$, the following quantities

$$z(k + h\delta|k) = \begin{bmatrix} \tilde{x}(k + h\delta|k) \\ \tilde{x}(k + (h - 1)\delta|k) \end{bmatrix}$$

and

$$P(k + h\delta|k) = \mathbb{E}[z(k + h\delta|k)z^*(k + h\delta|k)] = \begin{bmatrix} P_{11}(k + h\delta|k) & P_{12}(k + h\delta|k) \\ P_{21}(k + h\delta|k) & P_{22}(k + h\delta|k) \end{bmatrix}.$$

where

$$P_{11}(k + h\delta|k) = \mathbb{E}[\tilde{x}(k + h\delta|k)\tilde{x}^*(k + h\delta|k)]$$

$$P_{12}(k + h\delta|k) = \mathbb{E}[\tilde{x}(k + h\delta|k)\tilde{x}^*(k + (h - 1)\delta|k)]$$

$$P_{21}(k + h\delta|k) = \mathbb{E}[\tilde{x}(k + (h - 1)\delta|k)\tilde{x}^*(k + h\delta|k)]$$

$$P_{22}(k + h\delta|k) = \mathbb{E}[\tilde{x}(k + (h - 1)\delta|k)\tilde{x}^*(k + (h - 1)\delta|k)]$$

Section III	undirected graph	$m \rightarrow \infty$	$\ell \rightarrow 1$
Fixed ℓ	$Q^{opt} \in$ symmetric Section III – A	$Q^{opt} \in \arg \min_{Q \in \mathcal{Q}} \rho(Q)$ Section III – B	$Q^{opt} \in \arg \min_{Q \in \mathcal{Q}} \ Q^m\ _F$ Section III – C
Section IV	$m = 0$	$0 < m < \infty$	$m \rightarrow +\infty$ Q primitive
Fixed Q	$\ell^{opt}(Q, 0) = \ell_d^{opt}$ Theorem 4.1	$\ell_d^{opt} < \ell^{opt}(Q, m) \leq \ell^{opt}(Q, m+1) < \ell_c^{opt}$ Theorem 4.1	$\ell^{opt}(Q, \infty) = \ell_c^{opt}$ Theorem 4.1
Section V	$m \rightarrow \infty$	$r/q \rightarrow 0$	$r/q \rightarrow \infty$
Optimal ℓ and Q	$Q^{opt} \in \arg \min_{Q \in \mathcal{Q}} \rho(Q)$ $\ell^{opt} \rightarrow \ell_c^{opt}$ Section V – A	$Q^{opt} \in \arg \min_{Q \in \mathcal{Q}} \ Q^m\ _F,$ $\ell^{opt} = 1 - \frac{\ \bar{Q}^m\ _F^2}{N} \frac{r}{q} + o\left(\frac{r}{q}\right)$ Section V – B	$Q^{opt} \in \arg \min_{Q \in \mathcal{Q}} p(Q),$ $\ell^{opt} = \sqrt{\frac{N}{p^{opt}}} \sqrt{\frac{q}{r}} + o\left(\sqrt{\frac{q}{r}}\right)$ Section V – C

TABLE I

SUMMARY OF THE RESULTS

Note that $P_{12}(k + h\delta|k) = P_{21}^*(k + h\delta|k)$. It is possible to see that the above quantities satisfy the following recursive equations

$$P_{11}(k + (h+1)\delta) = \nu^2 Q P_{11}(k + h\delta) Q^* + \nu(1-\nu) [Q P_{12}(k + h\delta) + P_{21}(k + h\delta) Q^*] + 2\nu(1-\nu) \frac{q}{m} \mathbf{1}\mathbf{1}^* + \frac{q}{m} \mathbf{1}\mathbf{1}^* + (1-\nu)^2 P_{22}(k + h\delta) + (1-\nu)^2 \frac{q}{m} \mathbf{1}\mathbf{1}^*$$

$$P_{12}(k + (h+1)\delta) = \nu Q P_{11}(k + h\delta) + (1-\nu) P_{21}(k + h\delta) + (1-\nu) \frac{q}{m} \mathbf{1}\mathbf{1}^*$$

$$P_{22}(k + (h+1)\delta|k) = P_{11}(k + h\delta)$$

In our simulation we assume that $N = 50$, $m = 5$, $q = 1$, $r = 1$ and that $Q = Q_{\frac{1}{3}}$ where $Q_{\frac{1}{3}}$ is as the matrix introduced in Example 2. Moreover $\ell = \ell^{opt}$ where

$$\ell^{opt} = \arg \min_{\ell \in (0,1)} J\left(Q_{\frac{1}{3}}, \ell; 5, 1, 1\right) \approx 0.879.$$

Clearly, if $\nu = 1$, then $\lim_{k \rightarrow +\infty} \text{tr}(P_{11}(k+1|k)) = J\left(Q_{\frac{1}{3}}, \ell^{opt}; 5, 1, 1\right)$, which is the cost function introduced in Eqn. (13) for which no memory is added. In Figure 4 we depict the behavior of $\lim_{k \rightarrow +\infty} \text{tr}(P_{11}(k+1|k))$ for ν ranging in $[0, 2]$. It is remarkable to note that the minimum is reached when $\nu \approx 1.44$, meaning that the presence of memory can improve the performance.

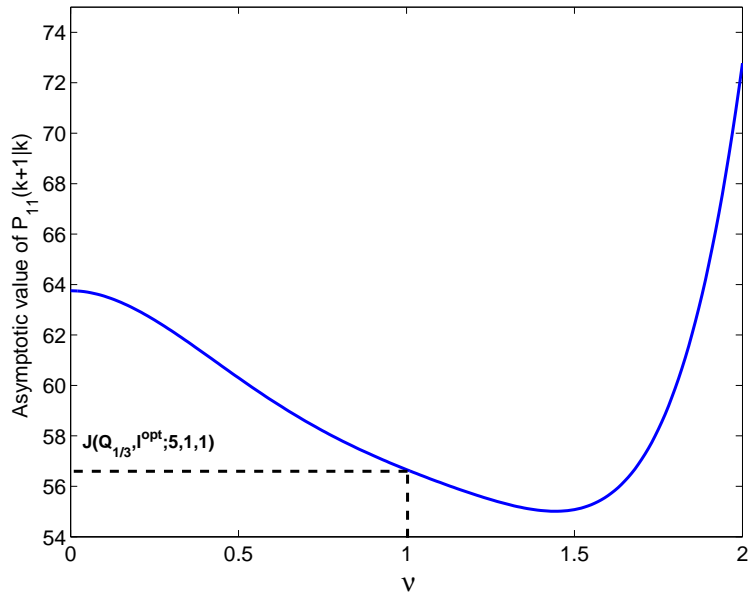


Fig. 4. Steady state error $\text{tr}(P_{11}(k+1|k))$ of distributed optimal estimation with memory as a function of weighting parameter ν . Figure also indicates optimal cost in the absence of memory which corresponds to $\nu = 1$.

VII. CONCLUSIONS

In this paper we have studied a prototypical problem of distributed estimation for Sensor Networks; the state of a scalar linear system is estimated via a two stage procedure which consists in (i) a standard (and decentralized) Kalman-like update and (ii) information propagation using consensus strategies. To this purpose two design parameters, i.e. the Kalman gain ℓ and the consensus matrix Q have to be designed. This choice is made by optimizing the steady state prediction (or estimation) error. We have discussed, under specific circumstances, the behavior of the “optimal” parameters. This is summarized in table of Figure VI.

Although these results have been obtained for simple scenarios where the state is scalar and all sensors are equal, they provide useful guidelines for choosing the local filter gain ℓ and the consensus matrix Q also for more general scenarios. In particular, we showed that the common practise of finding algorithms that minimize the spectral radius for the consensus matrix Q is not necessarily the optimal strategy in the context of optimal estimation of time-varying signals. In fact, we showed that depending on some specific regimes, it is more convenient to optimize the Frobenius norm $\|Q^m\|_F$ or the sum of the unitary eigenvalues $p(Q)$. Moreover, as discussed

in Section V, we showed that the joint optimization of Q and ℓ is not convex even in our simple setup. We also compared our approach with the recursive optimization proposed by Alriksson et al. [1], showing also that their strategy based on minimization of the estimation error at the next time step, fails to minimize the steady state cost (see Figures 1 and 3).

Many research avenues still deserve to be explored. The most promising one is the use of memory to improve performance. As shown in Example 5 in the previous section, memory can considerably improve estimation performance. However, the impact of memory length on performance and optimization algorithms for the weighting parameters are still open problems. Another important aspect is the extension of the results in this paper to the multivariable case where the gains for the sensors can be different. Finally, it is fundamental to find provably optimal strategies to simultaneously optimize the consensus matrix Q and the update gains ℓ , since the steady state cost function is non-convex in these parameters.

APPENDIX

Lemma 7.1: Let Q and \bar{Q} such that $\rho(Q) \geq \rho(\bar{Q}) + \epsilon$ where $\epsilon > 0$. Then the cost function defined in (23) satisfies

$$J^{opt}(Q, m) > J^{opt}(\bar{Q}, m), \quad \forall m > \bar{m}_\epsilon,$$

where \bar{m}_ϵ depends only on ϵ .

Proof: From Theorem 3.2 it follows that there exists \bar{m}_ϵ depending only on ϵ such that

$$J^{opt}(Q, m) > J(\bar{Q}, \ell^{opt}(Q, m); m), \quad \forall m > \bar{m}_\epsilon.$$

Clearly

$$J(\bar{Q}, \ell^{opt}(Q, m); m) \geq J(\bar{Q}, \ell^{opt}(\bar{Q}, m); m)$$

where, by definition $J(\bar{Q}, \ell^{opt}(\bar{Q}, m); m) \triangleq J^{opt}(\bar{Q}, m)$. ■

Lemma 7.2: Let $\epsilon > 0$. Let Q_1 and Q_2 be such that $\|Q_1^m\|_F \geq \|Q_2^m\|_F + \epsilon$. Then there exists δ_ϵ depending only on ϵ such that if $\frac{r}{q} < \delta_\epsilon$, the cost function defined in (26) satisfies

$$J^{opt}(Q_1, r/q) > J^{opt}(Q_2, r/q).$$

Proof: From Theorem 3.3 it follows that there exists ℓ_ϵ depending only on ϵ such that

$$J(Q_1, \ell; r, q) > J(Q_2, \ell; r, q), \quad \forall \ell > \bar{\ell}_\epsilon.$$

Hence if $\ell_{Q_1}^{opt} > \bar{\ell}_\epsilon$ we have that

$$J^{opt}(Q_1, r/q) > J(Q_2, \ell^{opt}(Q_1, r/q); r, q).$$

Clearly

$$J(Q_2, \ell^{opt}(Q_1, r/q); r, q) \geq J(Q_2, \ell^{opt}(Q_2, r/q); r, q),$$

where by definition $J(Q_2, \ell^{opt}(Q_2, r/q); r, q) \triangleq J^{opt}(Q_2, r/q)$.

Now we recall that if $\frac{r}{q} \rightarrow 0$ then $\ell_d^{opt} \rightarrow 1$. This implies that $\forall \eta > 0$ there exists $\delta_\eta > 0$ such that if $\frac{r}{q} < \delta_\eta$ then $\ell_d^{opt} > 1 - \eta$. Let now $\eta = 1 - \bar{\ell}_\epsilon$ and let us denote δ_ϵ the corresponding δ_η . Therefore, if $\frac{r}{q} < \delta_\epsilon$ then $\ell_d^{opt} > \bar{\ell}_\epsilon$. Since $\ell^{opt}(Q_1, r/q) \geq \ell_d^{opt}$ we have that also $\ell^{opt}(Q_1, r/q) > \bar{\ell}_\epsilon$. This concludes the proof. ■

REFERENCES

- [1] P. Aliksson and A. Rantzer. Distributed Kalman filtering using weighted averaging. In *Proceedings of the 17th International Symposium on Mathematical Theory of Networks and Systems*, Kyoto, Japan, 2006.
- [2] L. Babai. Spectra of Cayley graphs. *Journal of Combinatorial Theory, Series B*, 27:180–189, 1979.
- [3] P. Barooah and J.P.Hespanha. Distributed estimation from relative measurements in sensor networks. In *Proc. of the 2nd Int. Conf. on Intelligent Sensing and Information Processing*, Dec. 2005.
- [4] E. Behrends. *Introduction to Markov Chains (with Special Emphasis on Rapid Mixing)*. Vieweg Verlag, 1999.
- [5] J.-M. Borwein and A.-S. Lewis. *Convex Analysis and Nonlinear Optimazation*. CMS Books in Mathematics. Canadian Mathematical Society, 2000.
- [6] M. Cao, A. Morse, and B. Anderson. Reaching a consensus in a dynamically changing environment a graphical approach. *submitted to SIAM Journal on Control and Optimization*. [Online] Available at <http://www.eng.yale.edu/controls/pending/flockp1.pdf>.
- [7] M. Cao, D. A. Spielman, and E. M. Yeh. Accelerated gossip algorithms for distributed computation. In *44th Annual Allerton Conference on Communication, Control, and Computation.*, 2006.
- [8] R. Carli, F. Fagnani, A. Speranzon, and S. Zampieri. Communication constraints in the average consensus problem. *Automatica*, to appear.
- [9] G. Chen, G. Chen, and S. Hsu. *Linear Stochastic Control Systems*. CRC Press, 1995.
- [10] L. G. D. Bauso and R. Pesenti. Distributed consensus in networks of dynamic agents. In *Proceedings of IEEE Conference on Decision and Control (CDC'05)*, pages 7054–7059, 2005.
- [11] P. J. Davis. *Circulant matrices*. A Wiley-Interscience Publication, Pure and Applied Mathematics. John Wiley & Sons, New York-Chichester-Brisbane, 1979.
- [12] J. L. Doob. *Stochastic Processes*. John Wiley & Sons, Inc., New York, 1953.
- [13] F. R. Gantmacher. *The theory of matrices*. New York : Chelsea publ., 1959.
- [14] J. Gubner. Distributed estimation and quantization. *IEEE Transactions on Information Theory*, 39(4):1456–1459, July 1993.

- [15] Y. Hatano and M. Mesbahi. Agreement over random networks. *IEEE Transactions on Automatic Control*, 50(11):1867–1872, Nov. 2005.
- [16] A. Jadbabaie, J. Lin, and A. S. Morse. Coordination of groups of mobile autonomous agents using nearest neighbor rules. *IEEE Transactions on Automatic Control*, 48(6):988–1001, June 2003.
- [17] J. Xiao, A. Ribeiro, Z. Luo, and G. Giannakis. Distributed compression-estimation using wireless sensor networks. *IEEE Signal Processing Magazine*, pages 27–46, July 2006.
- [18] J. Lin, A. Morse, and B. Anderson. The multi-agent rendezvous problem. In *Proceedings of the 42nd IEEE Conference on Decision and Control (CDC'03)*, volume 2, pages 1508–1513, December 2003.
- [19] D. Looze, P. Houp, N. S. Jr., and M. Athans. On decentralized estimation and control with application to freeway ramp metering. *IEEE Trans. on Aut. Contr.*, 23(2):268–275, 1978.
- [20] Z. Luo and J. Tsitsiklis. Data fusion with minimal communication. *IEEE Trans. on Information Theory*, 40(5):1551–1563, Sept. 1994.
- [21] L. Moreau. Stability of multiagent systems with time-dependent communication links. *IEEE Transactions on Automatic Control*, 50(2):169–182, Feb. 2005.
- [22] R. Murray, J. Fax, R. O. Saber, and D. Spanos. Consensus and cooperation in multi-agent networked systems. *to appear in Proceedings of IEEE*, February 2007.
- [23] S. Muthukrishnan, B. Ghosh, and S. M. H. First and second order diffusive methods for rapid, coarse, distributed load balancing. *Theory of Computing Systems*, 31:331–354, 1998.
- [24] R. Olfati-Saber. Distributed Kalman filter with embedded consensus filters. In *Proceedings of the 44th IEEE Conference on Decision and Control, and European Control Conference*, December 2005.
- [25] R. Olfati-Saber. Ultrafast consensus in small-world networks. In *Proceedings of the 2005 American Control Conference (ACC'05)*, volume 4, pages 2371–2378, 2005.
- [26] R. Olfati-Saber and R. M. Murray. Consensus problems in networks of agents with switching topology and time-delays. *IEEE Trans. Automat. Control*, 49(9):1520–1533, 2004.
- [27] R. Olfati-Saber and J. Shamma. Consensus filters for sensor networks and distributed sensor fusion. In *Proceedings of the 44th IEEE Conference on Decision and Control, and European Control Conference*, December 2005.
- [28] A. Olshevsky and J. Tsitsiklis. Convergence rates in distributed consensus and averaging. In *Proc. of IEEE Conf. on Dec. and Control*, pages 3387–3392, San Diego, CA, USA, December 2006.
- [29] J. Predd, S. Kulkarni, and H. Poor. Regression in sensor networks: Training distributively with alternating projections. Technical report, July 2005.
- [30] W. Ren and R. Beard. Consensus seeking in multiagent systems under dynamically changing interaction topologies. *IEEE Transactions on Automatic Control*, 50(5):655–661, May. 2005.
- [31] A. Ribeiro and G. Giannakis. Bandwidth-constrained distributed estimation for wireless sensor networks - Part I: Gaussian case. *IEEE Transactions on Signal Processing*, 54(3):1131–1143, 2006.
- [32] A. Ribeiro and G. Giannakis. Bandwidth-constrained distributed estimation for wireless sensor networks - Part II: Unknown probabilistic density function. *IEEE Transactions on Signal Processing*, 54(7):2784–2796, 2006.
- [33] A. Ribeiro, G. Giannakis, and S. Roumeliotis. SOI-KF: Distributed Kalman filtering with low-cost communications using the sign of innovations. *IEEE Transactions on Signal Processing*, 54(12):4782–4795, 2006.
- [34] I. Schizas, A. Ribeiro, and G. Giannakis. Consensus in ad hoc wsns with noisy links - Part I: Distributed estimation of deterministic signals. *to appear in IEEE Transactions on Signal Processing*, 2007.

- [35] D. P. Spanos and R. M. Murray. Distributed sensor fusion using dynamic consensus. In *Proceedings of the 16th IFAC World Congress*, July 2005.
- [36] D. P. Spanos, R. Olfati-Saber, and R. M. Murray. Distributed Kalman filtering in sensor networks with quantifiable performance. In *Proceedings of the Information Processing for Sensor Networks (IPSN'05)*, 2005.
- [37] A. Speranzon, C. Fischione, and K. Johansson. Distributed and collaborative estimation over wireless sensor networks. In *Proceedings of the IEEE Conference on Decision and Control (CDC'06)*, pages 1025–1030, December 2006.
- [38] J. N. Tsitsiklis, D. P. Bertsekas, and M. Athans. Distributed asynchronous deterministic and stochastic gradient optimization algorithms. *IEEE Transactions on Automatic Control*, 31(9):803–812, Sep. 1986.
- [39] I. Vajda. On convergence of information contained in quantized observations. *IEEE Trans. on Information Theory*, 48(8):2163–2172, 2002.
- [40] P. Venkitasubramaniam, G. Mergen, L. Tong, and A. Swami. Quantization for distributed estimation in large scale sensor networks. In *Proc. of ICISIP*, pages 121– 127, 2005.
- [41] R. Viswanathan. A note on distributed estimation and sufficiency. *IEEE Trans. on Information Theory*, 39(5):1765–1767, 1993.
- [42] L. Xiao and S. Boyd. Fast linear iterations for distributed averaging. *Systems and Control Letters*, 53(1):65–78, September 2004.
- [43] L. Xiao, S. Boyd, and S.-J. Kim. Distributed average consensus with least-mean-square deviation. *Journal of Parallel and Distributed Computing*, 67(1):33–46, September 2007.
- [44] L. Xiao, S. Boyd, and S. Lall. A scheme for robust distributed sensor fusion based on average consensus. In *Proceedings of the Information Processing for Sensor Networks (IPSN'05)*, 2005.