CONSENSUS ALGORITHM DESIGN FOR DISTRIBUTED ESTIMATION

Ruggero Carli * **Alessandro Chiuso** *
Luca Schenato * **Sandro Zampieri** *

* *Department of Information Engineering, Università di Padova, Via Gradenigo 6/a, 35131 Padova, Italy*

Abstract: In this paper we study the problem of designing optimal consensus algorithms for distributed estimation of dynamical systems. In particular, we show that for fixed estimation gain, the problem of finding optimal consensus matrix is convex and can be efficiently performed numerically. We also provide some numerical examples to common scenarios, like symmetric geometric random graphs, which are representative models for wireless sensor networks.

Keywords: Average Consensus, distributed Kalman filtering

1. INTRODUCTION

Recently Wireless Sensor Networks have attracted the interest of a large research community since they can be used to monitor very large scale areas with fine resolution. However, collecting measurements from distributed wireless sensors nodes at a single location for on-line data processing may not be feasible due to long packet delay and limited bandwidth of the wireless network. As a consequence there is a growing need for in-network data processing tools and algorithms that provide high performance in terms of on-line estimation while reducing the communication load among all sensor nodes. In this work we will focus on distributed estimation of dynamical systems for which sensor nodes are not physically collocated 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 random variable 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 independent zero-mean random variables 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:

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

where $\text{mean}(y(t)) := \frac{1}{N}\sum_{i=1}^{N} y_i(t)$, and $0 < l^* < 1$ is the optimal Kalman gain that depends on the process noise covariance q and the equivalent measurement noise variance r/N. 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 multi-hop communication graph $\mathcal{G}$, and each sensor nodes 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 number only through local communication, the optimal estimate could be computed at each sensor node as follows:

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

These algorithms are known as *average consensus algorithms* and can be solved by using updates $z^+ = Qz$, where z is the vector whose entries are the quantities to be averaged [1] and Q is a doubly stochastic matrix, i.e. a matrix with properties $Q_{ij} \geq 0$, $\sum_j Q_{ij} = 1$ and $\sum_i Q_{ij} = 1$. Under some weak connectivity properties, these matrices guarantee that $\lim_{m\to\infty}[Q^m z]_i = \text{mean}(z)$, i.e. all elements of vector $Q^m z$ converge to their initial mean 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:

$$z = (1-l^*)\hat{T}_i(t) + l^* y_i(t) \quad \left\{ \begin{array}{l} \textbf{measur. \&} \\ \textbf{predict. stage} \end{array} \right.$$
$$\hat{T}_i(t+1) = [Q^m z]_i \quad \textbf{consensus stage}$$

Olfati-Saber (Olfati-Saber *et al.*, 2005a) and Spanos et al. (Spanos *et al.*, 2005a) 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. $y_i(t) = C_i T(t) + v_i$, as shown in the static scenario by Xiao et al. (Xiao *et al.*, 2005) ($A = I, w(t) = 0$) and in the dynamic scenario in (Spanos *et al.*, 2005b)(Olfati-Saber, 2005b). In this context, i.e. $m \gg 1$, it natural to optimize Q for fastest convergence rate of Q^m, which correspond to the second largest singular value of Q, for which there are already very efficient optimization tools available (Xiao *et al.*, 2004) (S.J. Xiao *et al.*, 2007). The assumption $m \gg 1$ is reasonable in applications for which communication is inexpensive as compared to sensing. 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 (Olfati-Saber *et al.*, 2005a)(Spanos *et al.*, 2005a)(Spanos *et al.*, 2005b)(Olfati-Saber, 2005b)(Xiao *et al.*, 2005). Recently, Alriksson at al. (Alriksson *et al.*, 2006) and Speranzon et al. (Fischione *et al.*, 2006), considered the case $m = 1$, i.e. sensors are allowed to communicate only once between sampling instants. Both works provide design methodologies to simultaneously design the estimator gain l and the consensus matrix Q but they rely on suboptimal heuristic optimization problems. In particular, in (Alriksson *et al.*, 2006) the authors propose a local on-step prediction strategy to alternatively optimize $l(k)$ and $Q(k)$ at any time step k, while in (Fischione *et al.*, 2006) the authors approximate it as an optimization problem based on convex relaxation that try to optimize performance while guaranteeing stability of the estimator. However, none of the two works provide optimality criteria for their strategies.

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, we consider a simpler scenario with a scalar state and sensors can measure the state affected by gaussian noise with the same covariance, which still captures some of the most important features of the problem. This analysis provides useful guidelines for choosing the local filter gain l and the consensus matrix Q also for more general scenarios. As a side result of our analysis we also see that the standard recipe of choosing Q optimizing the second largest eigenvalue is not necessarily the best thing to do; similarly choosing the centralized optimal gain l_c is not necessarily the optimal strategy. We also provide some numerical examples based on random geometric graphs which represents a realistic model for wireless sensor networks, and we show that our strategy outperforms the strategy proposed in (Alriksson *et al.*, 2006), thus showing that the latter does not converge to the optimal values.

2. PROBLEM FORMULATION

Consider a set V of N sensor nodes which are labeled $i = 1, 2, \ldots, N$. These sensors can communicate on 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

$$\dot{x}(t) = v(t) \tag{1}$$

where $v(t)$ is a continuous-time white noise [2] of zero mean and intensity $q \geq 0$, that is $\mathbb{E}[q(t)q(s)] = q\delta(t-s)$. The initial condition is also a random variable with expectation x_0 and variance σ.

[1] The entries of z can be real numbers, complex numbers or even matrices

[2] We 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 rigourously $x(t)$ is a Wiener process.

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

$$y_i(kT) = x(kT) + n_i(kT) \quad (2)$$

where T is the time-sampling and k the index indicating the k-th measure. Note that $y_i \in \mathbb{R}$, $\forall\, i$. We shall denote $y(kT) = [y_1(kT), \ldots, y_N(kT)]^*$ and $n(kT) = [n_1(kT), \ldots, n_N(kT)]^*$. Moreover the noise processes $n_i(kT) \in \mathbb{R}$ are such that $\mathbb{E}[n(kT)] = 0$, $E[n(kT)n(hT)] = rI\delta_{hk}$ where δ_{hk} is the Kronecker delta. Note also that (2) can be rewritten in the following vector form

$$y(kT) = x(kT)\mathbb{1} + n(kT). \quad (3)$$

where $\mathbb{1} = [1 \ldots 1]^*$. From now on we assume, without loss of generality, that $T = 1$. Suppose now that, between each pair of subsequent measurement update indices k and $k+1$, the each node exchanges m messages; we assume that these transmissions take place at the following times $k+\delta, k+2\delta, \ldots, k+(m-1)\delta, k+m\delta$, where $\delta = \frac{1}{m}$. Note that $k + m\delta = k + 1$. Moreover suppose that the i-*th* sensor possesses at each of the above indices a estimate of $x(k)$ that, by convention, we indicate by the following notation $\hat{x}_i\,(k{+}\delta|k)\,,\ \hat{x}_i\,(k{+}2\delta|k)\,, \ldots,\ \hat{x}_i\,(k{+}(m-1)\delta|k)\,,$ $\hat{x}_i\,(k{+}1|k)$. More compactly we can write

$$\hat{x}\,(k+h\delta|k) = [\hat{x}_1\,(k+h\delta|k)\,, \ldots, \hat{x}_N\,(k+h\delta|k)]^*\,.$$

We assume that these estimates are updated according to the following rule

$$\begin{cases} \hat{x}\,(k|k) \;=\; (1-l)\,\hat{x}(k|k-1) + ly(k) \\ \hat{x}\,(k+h\delta|k) \;=\; Q(k)\hat{x}\,(k+(h-1)\delta|k) \end{cases} \quad (4)$$

where $Q(k)$ is a suitable matrix compatible with the communication graph and where $0 < l < 1, \forall\, k \geq 0$. From now on we assume that $l(k) = l$ and $Q(k) = Q$, i.e. they are constant. If we impose that $\hat{x}_i$ is an unbiased estimator for each i and for each update index, we have that Q must satisfy the following condition

$$Q\mathbb{1} = \mathbb{1}. \quad (5)$$

In fact by imposing that $\mathbb{E}\,[\hat{x}\,(k+h\delta|k)] = x_0\mathbb{1}, \forall\, 0 \leq h \leq m$, it results from the update rule that

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

In order to have that $x_0 Q\mathbb{1} = x_0\mathbb{1}$, for any possible value x_0, we obtain that (5) must hold. Furthermore if we restrict to nonnegative Q, namely a matrix with nonnegative entries, condition (5) imposes that Q is a stochastic matrix. From now on, we assume that Q is stochastic. Moreover the local estimators are initialized by setting $\hat{x}(0|0) = y(0)$. Now we define the new variable $\tilde{x}\,(k+h\delta|k) = x\,(k+h\delta)\,\mathbb{1} - \hat{x}\,(k+h\delta|k)$ 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 to rewrite (1) in the following way

$$x\,(k+(h+1)\delta) = x\,(k+h\delta) + w\,(k+h\delta) \quad (6)$$

where

$$w\,(k+h\delta) = \int_{k+h\delta}^{k+h\delta} v(\tau)d\tau. \quad (7)$$

Note that $\mathbb{E}[w\,(k{+}h\delta)] = 0$ and that $\mathbb{E}\left[w^2(k{+}h\delta)\right] = \frac{q}{m}$. By straightforward calculations, we get that, for $h = 0$,

$$\tilde{x}(k|k) = (1-l)\,\tilde{x}(k|k-1) - ln(k)$$

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

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

In order to analyze the asymptotic properties of the above estimates it is convenient to introduce the following matrices

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

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

$$P(k|k) = (1-l)^2\,P(k|k-1) + l^2 rI \quad (8)$$

and, for $h = m$,

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

Plugging (8) into (9) and plugging (9) into (8) evaluated at the index $k+1$ we obtain the following recursive equations

$$\begin{aligned} P(k+1|k) =& (1-l)^2 Q^m P(k|k-1)\,(Q^*)^m + \\ &+ l^2 rQ^m\,(Q^*)^m + q\mathbb{1}\mathbb{1}^* \end{aligned}$$

and

$$\begin{aligned} P(k+1|k+1) =& (1-l)^2\,Q^m P(k|k)\,(Q^*)^m + \\ &+ (1-l)^2\,q\mathbb{1}\mathbb{1}^* + l^2 rI \end{aligned}$$

Since $\hat{x}(0|0) = y(0)$ we have that $P(0|0) = rI$ and $P(1|0) = rQ^m\,(Q^m)^* + q\mathbb{1}\mathbb{1}^*$. By rewriting the last two recursive equations as expressions depending respectively on $P(1|0)$ and $P(0|0)$ we obtain

$$\begin{aligned} P(k+1|k) =& (1-l)^{2k} Q^{km} P(1|0)\,(Q^*)^{km} + \\ &+ rl^2 \sum_{i=0}^{k-1} (1-l)^{2i} Q^{(i+1)m}\,(Q^*)^{(i+1)m} + \\ &+ q\left(\sum_{i=0}^{k-1}(1-l)^{2i}\right)\mathbb{1}\mathbb{1}^* \end{aligned}$$

and

$$P(k|k) = q\sum_{i=0}^{k-1}(1-l)^{2i+2}\mathbb{1}\mathbb{1}^* + l^2 r\sum_{i=0}^{k-1}(1-l)^{2i}Q^{im}\,(Q^*)^{im}$$

By taking the limit for $k \to \infty$, we get

$$\begin{aligned} \lim_{k\to\infty} P(k+1|k) =& rl^2\sum_{i=0}^{\infty}(1-l)^{2i}Q^{(i+1)m}(Q^*)^{(i+1)m} + \\ &+ q\frac{1}{1-(1-l)^2}\mathbb{1}\mathbb{1}^* \end{aligned}$$

and

$$\lim_{k\to\infty} P(k|k) = q\frac{(1-l)^2}{1-(1-l)^2}\mathbb{1}\mathbb{1}^* + rl^2\sum_{i=0}^{\infty}(1-l)^{2i}Q^{im}\,(Q^*)^{im}$$

Now let us define the following functionals cost [3]

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

and

$$J_2(l, Q; m, r, q) = \mathrm{tr}\left\{\lim_{k\to\infty} P(k|k)\right\}$$

We can formulate the following minimization problem.

Problem Given a graph $\mathcal{G}$ and a nonnegative integer m, find a real l such that $0 < l < 1$, and a matrix $Q \in \mathcal{Q}$, minimizing J_1 or J_2.

Remark 2.1. In the sequel of the paper 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 the Kalman filtering. For the sake of the simplicity, we will denote this functional cost simply by J in place of J_1. Hence

$$J = rl^2 \mathrm{tr}\left\{\sum_{i=0}^{\infty}(1-l)^{2i} Q^{(i+1)m}\,(Q^*)^{(i+1)m}\right\} + q\frac{1}{1-(1-l)^2}N$$

Remark 2.2. Let us denote with $\sigma(Q) = \{1, \lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_{N-1}\}$ the spectrum of Q. Note that, if Q is a normal matrix, namely $QQ^* = Q^*Q$ then formula in the previous remark can be rewritten as

$$J = \frac{rl^2 + qN}{1-(1-l)^2} + rl^2 \sum_{j=1}^{N-1} \frac{|\lambda_j|^{2m}}{1-(1-l)^2|\lambda_j|^{2m}}$$

Also note that if Q is normal and stochastic, then it is also doubly stochastic.

From now on, we will assume that Q is a normal matrix and we will denote by $\mathcal{Q}$ the set of the normal matrices compatible with the graph $\mathcal{G}$. Relevant subclasses of normal matrices are, for instance, Abelian Cayley matrices (Carli *et al.*, 2005), circulant matrices and symmetric matrices.

3. OPTIMAL CONSENSUS MATRIX Q FOR FIXED GAIN L

In this section we assume that the estimation gain l is fixed, and thus the problem we want to solve becomes the following

$$Q(l; m) = \arg\min_{Q\in\mathcal{Q}} J(Q, l; m). \tag{10}$$

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

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

- the communication graph $\mathcal{G}$ is undirected
- the sensors can communicate arbitrarily fast within two subsequent measurements, i.e., $m \to \infty$
- the estimation gain l is sufficiently large, i.e. $l \to 1$, which 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 to treat these cases separately, we observe that

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

and hence we can restrict only to the evaluation of last term of the previous equation. Since this quantity will appear often along the section, we denote it as

$$S(Q, l; m) = \sum_{j=1}^{N-1} \frac{|\lambda_j|^{2m}}{1-(1-l)^2|\lambda_j|^{2m}} \tag{11}$$

3.1 Undirected communication graph $\mathcal{G}$

We start by noticing that the assumption that the communication graph $\mathcal{G}$ is undirected implies that, given any $Q \in \mathcal{Q}$, also Q^* belongs to $\mathcal{Q}$. Consider now the symmetric matrix $(Q + Q^*)/2$, that we denote as Q_{sym}. 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 $J(Q, l; m)$ and $J(Q_{sym}, l; m)$ 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

$$J(Q, l; m) \geq J(Q_{sym}, l; m).$$

Remark 3.2. It is important to note that the previous lemma cannot be generalized to general stochastic matrices Q. In fact, it is easy to find a non-normal Q for which the symmetrized matrix Q_{sym} defined above gives a larger cost index.

An immediate consequence of the above Lemma is that, when the communication graph is undirected, the minimum of the functional cost J is reached by symmetric matrices. Thus, if $\mathcal{Q}_{sym}$ is the subset of $\mathcal{Q}$ containing the symmetric matrices, that is $\mathcal{Q}_{sym} = \{Q \in \mathcal{Q} : Q = Q^*\}$, solving (12) is equivalent to solve

$$\arg\min_{Q\in\mathcal{Q}_{sym}} J(Q, l; m). \tag{12}$$

The following result provides a powerful characterization of (12).

Theorem 3.3. Let $\mathcal{Q}_{sym}$ be as above. Then the functional cost $J(Q,l;m)$ defined on $\mathcal{Q}_{sym}$ is a convex function.

Theorem 3.3 states that (12) is a convex problem implying thus that the solution of (12) is unique and that it can be performed efficiently by suitable numeric algorithms. In fact, Xiao et al. (Xiao *et al.*, 2007) adopted this strategy to optimize similar performance costs over symmetric stochastic matrices.

3.2 Fast communication ($m \to \infty$)

Before stating the main result of this subsection we recall the following definition. Let Q be any matrix such that $Q\mathbb{1} = \mathbb{1}$ and assume that its spectrum $\sigma(Q)$ is contained in the closed unit disk. Define

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

It is called the essential spectral radius of Q. The following result holds.

Theorem 3.4. 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

$$J(Q_1,l;m) \geq J(Q_2,l;m), \qquad \forall\, m > \bar{m}.$$

3.3 Large gain ($l \to 1$)

We start by providing the following notational definition. Given a matrix A we denote with $\|A\|_F$ the Frobenius norm of A, namely $\|A\|_F = \sqrt{\text{tr}\ \{AA^*\}}$. Given any two matrices Q_1 and Q_2 belonging to $\mathcal{Q}$, the following result provides an interesting comparison between $J(Q_1,l;m)$ and $J(Q_2,l,m)$ when the gain l is sufficiently close to 1.

Theorem 3.5. Let Q_1, Q_2 be two matrices such that $\|Q_1^m\|_F > \|Q_2^m\|_F$. Then there exists $\bar{l}$ (depending only on $\|Q_1^m\|_F - \|Q_2^m\|_F$) such that

$$J(Q_1,l;m) - J(Q_2,l;m) > 0, \qquad \forall l > \bar{l}. \quad (13)$$

Remark 3.6. At first sight, Theorem 3.4 and Theorem 3.5 seem in contradiction. However, this can be explained by observing that $\|Q^m\|_F^2 = 1 + \rho^{2m}(Q) + o(\rho^{2m}(Q))$, therefore, for large m, minimizing the Frobenius norm of Q^m or the spectral radius of Q is almost equivalent.

4. 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

$$\arg\min_{l\in(0,1)} \ J(Q,l;m,r,q) \qquad (14)$$

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

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

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

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

Convexity of J allows easy computation of $l^{opt}(Q,m)$. In the remaining of this section we shall see that the sequence $\{l^{opt}(Q,m)\}_{m=0}^{\infty}$ is monotonically non-decreasing in m. Moreover, it is 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 $Q = \frac{1}{N}\mathbb{1}\mathbb{1}^*$, i.e.

$$l_d^{opt} = \arg\min_{l\in(0,1)} \ J(I,l;m),\ l_c^{opt} = \arg\min_{l\in(0,1)} \ J\left(\frac{1}{N}\mathbb{1}\mathbb{1}^*, l; m\right)$$

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

$$l_d^{opt} = \frac{-q+\sqrt{q^2+4qr}}{2r},\ l_c^{opt} = \frac{-q+\sqrt{q^2+4q\bar{r}}}{2\bar{r}}$$

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

The role played by l_d^{opt} and l_c^{opt} is clarified in next proposition where it is shown that they are respectively a lower bound and an upper bound for any $l^{opt}(Q,m)$. Precisely, it is stated a stronger result characterizing the sequence $\{l^{opt}(Q,m)\}_{m=0}^{\infty}$.

Theorem 4.2. Let $Q \in \mathcal{Q}$. Let $l^{opt}(Q,m)$ be defined as above. Then the following chains of inequalities hold true

$$l_d^{opt} = l^{opt}(Q,0) \leq l^{opt}(Q,1) \leq \ldots \leq l^{opt}(Q,m) \leq$$
$$\leq l^{opt}(Q,m+1) \leq \ldots \leq l^{opt}(Q,\infty) \leq l_c^{opt}$$

and

$$J\left(Q, l_d^{opt}; 0\right) \geq J\left(Q, l^{opt}(Q,1); 1\right) \geq$$
$$\geq J\left(Q, l^{opt}(Q,2); 2\right) \geq \ldots \geq J\left(Q, l_c^{opt}; \infty\right)$$

Moreover $l^{opt}(Q,\infty) = l_c^{opt}$ if and only if Q is irreducible and aperiodic.

Due to limitation of space the proof of this theorem and some other theorems in the following sections are omitted.

5. JOINT OPTIMIZATION OF Q AND L: SPECIAL CASES

We have shown in the previous two sections that the functional cost J is a convex function, both if we fix the gain and we assume J defined on the set of the symmetric matrices and if we fix the consensus matrix and we assume that the gain is the independent variable. We ask now whether J is a convex function jointly in l and $Q \in \mathcal{Q}_{sym}$. One can see numerically that this is not true for any value of q and r. Therefore, the joint minimization of J

$$Q^{opt}(m,r,q)\; l^{opt}(m,r,q) \in \underset{l\,\in\,(0,1);\,Q\,\in\,\mathcal{Q}}{\arg\min}\; J(Q,l;m,r,q) \tag{15}$$

results to be quite hard in general. Nevertheless, restricting to some asymptotic case on the values of m, r and q, it is possible to provide an analytical characterization of the above problem. In particular we will consider the following situations:

- the sensors can communicate arbitrarily fast within two subsequent measurements, i.e., $m \to \infty$
- $\frac{r}{q} \approx 0$, 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 $Q^{opt}(m,r,q)$ $l^{opt}(m,r,q)$ are indeed only functions of m and r/q. In the sequel, without risk of confusion, we shall omit arguments which are kept fixed.

5.1 Fast communication ($m \to \infty$)

Let $Q^{opt}(m), l^{opt}(m)$ be a solution of (15). In this subsection we provide a characterization of $Q^{opt}(m)$ and $l^{opt}(m)$ when $m \to \infty$. Then the following result holds.

Theorem 5.1. Let $Q^{opt}(m)$, $l^{opt}(m)$ be as defined above. Then

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

and

$$\lim_{m\to\infty} l^{opt}(m) = l_c^{opt}.$$

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

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

5.2 Small measurement noise ($r/q \to 0$)

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 $r/q \to 0$. Let $Q^{opt}(r/q)$, $l^{opt}(r/q)$ be a solution of (15), then the following result holds.

Theorem 5.2. Let $Q^{opt}(r/q)$, $l^{opt}(r/q)$ be defined above and let

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

Then

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

Moreover

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

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

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

holds.

Note that $l_c^{opt} = 1 - \frac{1}{N}\frac{r}{q} + o(\frac{r}{q})$ and $l_d^{opt} = 1 - \frac{r}{q} + o(\frac{r}{q})$, showing that the communication graph $\mathcal{G}$ determines the coefficient of the first order expansion in r/q.

5.3 High measurement noise $q/r \to 0$

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

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

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

Moreover

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

Note that $l_c^{opt}(q/r) = \sqrt{N}\sqrt{q/r} + o(\sqrt{q/r})$ and $l_d^{opt}(q/r) = \sqrt{q/r} + o(\sqrt{q/r})$, therefore, the optimal gain depends on the communication structure of the underlying communication graph. In fact, if sensors cannot communicate, then necessarily $Q^{opt} = I$, therefore $l^{opt}(q/r) = l_d^{opt}(q/r)$, while if the communication graph is fully connected, then $Q^{opt} = \frac{1}{N}\mathbb{1}\mathbb{1}^*$, therefore $l^{opt}(q/r) = l_c^{opt}(q/r)$.

The previous theorem states also that for $q << r$ then it is necessary to choose a matrix Q consistent with the communication graph that minimizes the number of unitary eigenvalues.

6. AN ILLUSTRATIVE EXAMPLE

In this section we provide a numerical comparison between the approach presented in this paper and the method proposed in (Alriksson *et al.*, 2006). The authors in (Alriksson *et al.*, 2006), analyze a general MIMO scenario where the gain $l = l(t)$ (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 (Alriksson *et al.*, 2006) with our approach we assume the averaging matrix W in (Alriksson *et al.*, 2006) corresponds to performing m consensus iterations using the matrix Q, i.e. $W = Q^m$. The matrix gain l is chosen to minimize the estimation error covariance of the local estimators (i.e. in a decentralized fashion). In (Alriksson *et al.*, 2006) l is different for each sensor and the consensus matrix Q is chosen so that the estimation error covariance of the local estimators is minimized after consensus (weighted averaging in (Alriksson *et al.*, 2006)). In our simulation, we used 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 $r = 1$. We consider both the minimization of J_1 and J_2, with J_1, J_2 defined as in the Section II. We use the following notational conventions. $Q_1^{opt}(m)$, $l_1^{opt}(m)$ and $Q_2^{opt}(m)$, $l_2^{opt}(m)$ 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:

$$Q_1^{opt}(m),\, l_1^{opt}(m) \in \underset{l\in(0,1),\, Q\in\mathcal{Q}}{\arg\min} \; J_1(Q,l;m,r,q)$$
$$Q_2^{opt}(m),\, l_2^{opt}(m) \in \underset{l\in(0,1),\, Q\in\mathcal{Q}}{\arg\min} \; J_2(Q,l;m,r,q)$$

As said before, in (Alriksson *et al.*, 2006) the optimal matrix gains and the optimal consensus matrix are found *recursively* at each time step t. We denote by $J_1^r(m)$ and $J_2^r(m)$ (the superscript "r" means *recursively*) the asymptotic cost values [4] to which tends J_1 and J_2 by the method proposed in (Alriksson *et al.*, 2006). We run simulations for m ranging in the interval $[1, 10]$. A few remarks are now in order.

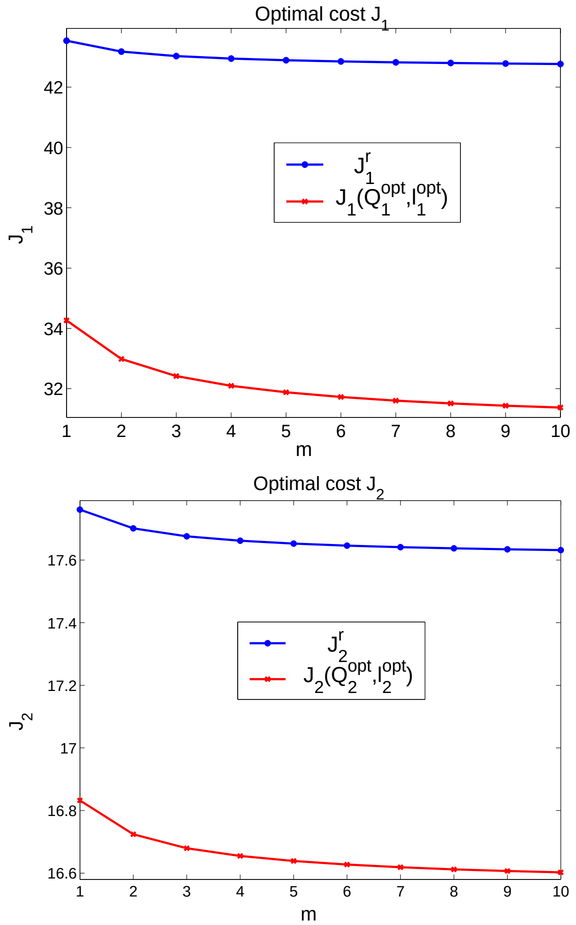

Fig. 1. Comparison between $J_1(Q_1^{opt}(m), l_1^{opt}(m), m)$, and $J_1^r(m)$ ***(top)***. Comparison between $J_2(Q_2^{opt}(m), l_2^{opt}(m), m)$ and $J_2^r(m)$ ***(bottom)***.

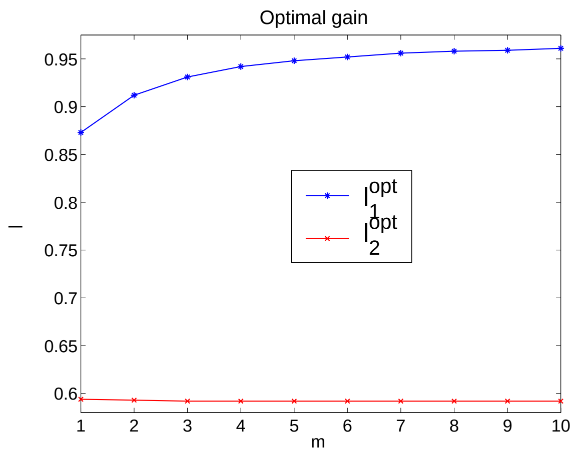

Fig. 2. Optimal gain l.

In top panel of Figure 1 we report the value of J_1 corresponding to the "optimal" parameter pair $l_1^{opt}(m), Q_1^{opt}(m)$ and of $J_1^r(m)$. Clearly [5]

[4] There is no proof of convergence in (Alriksson *et al.*, 2006); however this is observed experimentally.

[5] $l_1^{opt}(m), Q_1^{opt}(m)$ are found minimizing J_1.

Section III	undirected graph	$m \to \infty$	$l \to 1$
Fixed l	Q^{opt} symmetric Section III – A	$Q^{opt} \in \arg\min_{Q\in\mathcal{Q}} \rho(Q)$ Section III – B	$Q^{opt} \in \arg\min_{Q\in\mathcal{Q}} \lVert Q^m \rVert_F$ Section III – C
Section IV	$m = 0$	$0 < m < \infty$	$m \to \infty$ $+Q$ aperiodic irreducible
Fixed Q	$l^{opt}(Q,0) = l_d^{opt}$ Theorem 4.1	$l_d^{opt} < l^{opt}(Q,m) \leq l^{opt}(Q,m+1) < l_c^{opt}$ Theorem 4.1	$l^{opt}(Q,\infty) = l_c^{opt}$ Theorem 4.1
Section V	$m \to \infty$	$r/q \to 0$	$r/q \to \infty$
Optimal l and Q	$Q^{opt} \in \arg\min_{Q\in\mathcal{Q}} \rho(Q)$, $l^{opt} \to l_c^{opt}$ Section V – A	$Q^{opt} \in \arg\min_{Q\in\mathcal{Q}} \lVert Q^m \rVert_F$, $l^{opt} \simeq 1 - \frac{\lVert \bar{Q}^m \rVert_F^2}{N} \frac{r}{q}$ Section V – B	$Q^{opt} \in \arg\min_{Q\in\mathcal{Q}} p(Q)$ $l_c^{opt} \simeq \sqrt{\frac{N}{p^{opt}}}\sqrt{\frac{q}{r}}$ Section V – C

Fig. 3. Summarizing table of results

$l_1^{opt}(m), Q_1^{opt}(m)$ yields a value of J_1 which is better than J_1^r. A similar consideration holds for J_2 (bottom panel of Figure 1. In this case $l_2^{opt}(m), Q_2^{opt}(m)$ gives the best performance, again $J_2^r(m)$ yields the worst value. It is remarkable that, even though the optimal matrix gain and the optimal consensus matrix in (Alriksson *et al.*, 2006) are chosen minimizing step by step an estimation error (and hence a cost which resembles J_2) its asymptotic value does not provide the minimum of J_2. In figure 2 we depict the behavior of $l_1^{opt}(m)$ and $l_2^{opt}(m)$. Notice that $l_1^{opt}(m)$ grows with m whereas $l_2^{opt}(m)$ remains almost constant.

7. 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 3.

Although these results have been obtained for a rather simple scenario 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 fact, as discussed in Section V the joint optimization of Q and l is not convex even in our simple setup.

Finally, we compared our approach with the recursive optimization proposed by Alriksson et al. (Alriksson *et al.*, 2006), showing also that their strategy fails to minimize the steady state cost (see Figures 1).

REFERENCES

L. Xiao and S. Boyd, Fast linear iterations for distributed. averaging In *Systems and Control Letters*, vol. 53, n.1, pp. 65–78, 2004.

P. Alriksson and A. Rantzer, Distributed Kalman Filtering using Weighted Averaging. 2006.

R. Olfati-Saber and J.S. 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, 2005 a

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*, 2005 b

D.P. Spanos and R.M. Murray, Distributed sensor fusion using dynamic consensus. In *Proceedings of the 16th IFAC World Congress*, 2005 a

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 b

A. Speranzon and C. Fischione and K.H. Johansson, Distributed and Collaborative Estimation over Wireless Sensor Networks, In *Proceedings of the IEEE Conference on Decision and Control, San Diego*, 2006

L. Xiao and 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

L. Xiao and S. Boyd and S.J. Kim, Distributed Average Consensus with Least-Mean-Square Deviation In *Journal of Parallel and Distributed Computing* , vol.67, n.1, pp.33–46, 2007

R. Carli and F. Fagnani and A. Speranzon and S. Zampieri , Communication constraints in the average consensus problem. Note:*Submitted to Automatica*, 2005