Distributed Estimation through Randomized Gossip Kalman Filter

Simone Del Favero, Sandro Zampieri

Abstract— **In this paper we consider the problem of estimating a random process from noisy measurements, collected by a sensor network. We analyze a distributed two–stage algorithm. The first stage is a Kalman–like estimate update, in which each agent makes use only of its own measurements. During the second phase agents communicate with their neighbors to improve their estimate. Estimate fusion is operated by running a consensus iteration. In literature it has been considered only the case of a fixed communication strategies, i.e. described by a fixed constant consensus matrix. However, in many practical cases this is just a rough model of communications in a sensor network, that usually happen according to a randomized strategy. This strategy is more properly modeled by assuming that the consensus matrices are drawn, according to a selection probability, from an alphabet of matrices compatible with the communication graph, at each time instant. This work deals therefore with randomized communication strategies and in particular with the symmetric gossip. A mean square performance analysis is carried out and an upper–bound for the trace of the estimation error variance is derived. The proposed upper–bound has to be considered the main technical contribution of the present paper, since it is based on a highly non–trivial inequality on matrix singular values, proved in the appendix. This upper–bound is a good performance assessment index and it is assumed therefore as a cost function to be minimized. We show moreover that problem of minimizing this cost function by choosing the Kalman gain and the selection probability is convex in each of the two variables separately although it is not jointly convex. Finally simulations are presented and the results discussed.**

I. INTRODUCTION

In this paper we consider the problem of estimating a random process from noisy measurements, collected by a sensor network. More precisely, as in [1], we consider a prototypical problem of estimation in sensor networks, namely the problem of estimating the state of a scalar random process. We will analyze a distributed two–stage algorithm: the first stage is a Kalman–like estimate update, in which each agent makes use only of its own measurements, while the second phase is devoted to the estimates exchange between neighbor nodes and to the estimates fusion.

To find an optimal way to fuse local estimates is a very difficult problem, that can hardly be handled if the communication graph has cycles. Nevertheless, the problem of finding a distributed algorithm that achieves the same performance of the centralized Kalman filter has been solved in [2], [3]. To this aim the estimate fusion problem has been formulated as an consensus problem. In fact, as remarked in [1], since the local estimates mean is a sufficient statistic to compute the optimal estimate, the optimal fusion problem can be solved with consensus techniques. Unluckily, the solution proposed in [2], [3] relies on the assumption that communications are much faster than measurements, so that one can assume that consensus can be reached during two consecutive measurements. Under this assumption, the choice of the Kalman gain to be used at each node is not an issue: the centralized gain has to be used. Obviously, the assumption of fast communications does not hold in all practical cases and the proposed scheme becomes suboptimal.

In [1] and in [4] it has been studied the case of finite number of communications between two subsequent measurements. In both works, inspired by the fact that the mean of the local estimates is a sufficient statistic to compute the centralized estimate, estimate fusion was implemented as m consensus steps. In [4] and in the subsequent improvement [5], Kalman gain and consensus weights are selected at each time step in order to minimize the estimate error variance at each node in the next step. In [1] the steady state error variance was minimized in the scalar case. Both of the approaches consider a fixed communication scheme, i.e. it is assumed that all the information exchanges prescribed by the communication graph happen between 2 subsequent measurements. However, in many practical cases this is just a rough model of communications in a sensor network, valid if the time between two subsequent measurements is sufficiently large. In fact, communication in a sensor network often happens according to a randomized protocol such as broadcast [6] or symmetric gossip strategies [7]. In the first case, quite common in wireless sensor networks, one node wakes up after a random sleep time and broadcasts its estimate to its neighbors. In the symmetric gossip case one node randomly wakes up, picks up randomly one of its neighbor nodes and exchange with it its estimate. The use of randomized protocols avoids the need of cumbersome communication scheduling, reduces the need of time-synchronization and may allow to reduce power consumption. A further cause of randomness in the communication is the potential unpredictably of the environment where these protocols are implemented: packet losses and collisions are in fact rather common in a sensor network. Moreover nodes failures, arrivals and departures are common events in the large networks under study. To take these randomnesses into account, in this work the estimate fusion will be implemented as m randomized consensus steps. This means to assume that the consensus matrix, rather than being constant as in [1], is drawn at each time instant form an alphabet of matrices compatible with the graph $\mathcal{G}$. Randomized consensus algorithms have been deeply investigated in literature: the symmetric gossip algorithm for average consensus was deeply studied in the seminal work [7] and an extensive analysis of convergence properties for many common randomized consensus algorithms, in particular for symmetric gossip and broadcast, can be found in [6].

Moreover we mention [8] and [9], both dealing with dis-

Both authors are with the Dept. of Information Eng., Univ. of Padova, Via Gradenigo 9/B, 35131, Padova. Fax: +39 049 827 7614. WWW: http://www.dei.unipd.it/. E-mails: {simone.delfavero, zampi}@dei.unipd.it.

This work was supported by the CaRiPaRo foundation under the WiseWai founded project and by EU STREP project named "FeedNetBack".

tributed Kalman filtering, and [10] where a belief propagation approach to distributed Kalman filtering has been proposed.

A. Contribution and Paper organization

This paper is devoted to the analysis of a distributed Kalman filtering algorithm that makes use of randomized communication strategies, focusing in particular on symmetric gossip.

The paper is organized as follows: after having briefly clarified the notation used in the paper (Section II), we give a detailed formulation of the estimation problem under consideration (Section III) and we introduce the estimation algorithm under analysis, namely a randomized version of the algorithm proposed in [1] (Section IV). In Section V a mean square analysis is carried out. In particular we give an upper bound on the steady state error variance of the proposed filter. This upper–bound has to be considered the main technical contribution of the present paper, since it is based on an highly non–trivial inequality on matrix singular values based on majorization theory, [11], [12], [13]. In the appendix, in fact, we show that the eigenvalues of the matrix $\mathbb{E}[Q_i \dots Q_0 Q_0 \dots Q_i]$ are submajorized by the eigenvalues of $\mathbb{E}[Q_0^2]^i$. In Section VI we discuss the optimization problem of finding a Kalman gain and randomized communication strategy, i.e. a selection probability, that minimizes a suitable cost function, namely the proposed upper–bound the trace of the steady state error variance. In particular, we show that the optimization problem is convex in each of the two variables separately although it is not jointly convex. Finally in Section VII some simulation results are presented.

II. Mathematical Preliminaries

Before proceeding, we introduce some mathematical preliminaries that will be used through the paper. We will denote with $\mathbb{1}_N$, or simply with $\mathbb{1}$, the vector $[1, \dots, 1]^T \in \mathbb{R}^N$. Recall that a stochastic matrix P is a matrix with non–negative entries such that $P\mathbb{1} = \mathbb{1}$. A matrix P is said to be doubly stochastic if both P and P^T are stochastic. Any symmetric stochastic matrix is therefore doubly stochastic. We will denote with $\mathbb{S}$ the set of symmetric matrices. Given a matrix $M \in \mathbb{R}^{N\times N}$ we denote with $\underline{\sigma}(M)$ the vector formed by the singular values of M decreasingly ordered and with $\underline{\lambda}(M)$ vector formed by the eigenvalues of M order so that $|\underline{\lambda}_0(M)| \geq \dots \geq |\underline{\lambda}_{N-1}(M)|$. Recall moreover that for any normal matrix N and therefore for any symmetric matrix, $\underline{\sigma}(N) \equiv |\underline{\lambda}(N)|$. It is moreover well known that for any stochastic P, $\lambda_1(P) = 1$. We will denote with $\otimes$ the Kronecker product. Given a matrix $M \in \mathbb{R}^{N\times N}$ we say that the graph $\mathcal{G}_M = (\mathcal{V}, \mathcal{E}_M)$, $\mathcal{V} = \{1, 2, \dots, N\}$, such that $M_{i,j} \neq 0 \;\Leftrightarrow\; (j,i) \in \mathcal{E}_M$ is *associated* with M while we say that M is *compatible with a given graph* $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ if $M_{i,j} \neq 0 \;\Rightarrow\; (j,i) \in \mathcal{E}_M$.

III. Problem Formulation

Consider a sensor network of N agents, labeled with the elements of the set $\mathcal{V} = \{1 \dots N\}$. Let us describe the communication constraints of the network with a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where the edge $(i,j) \in \mathcal{E}$ if and only if i can transmit information to j.

Our goal is to estimate, by means of such a sensor network, a discrete-time scalar random process of the form:

$$x(t+1) = x(t) + w(t)$$

where $w(t)$, known as model noise, is a white noise with variance q. Each node i of the network can collect noisy measurements the state $x(t)$:

$$y_i(t) = x(t) + v_i(t) \tag{1}$$

where the measurement noise, $v_i(t)$, is a white noise with variance r_i. It is reasonable to assume that sensors are affected by N independent measurement noises, all independent also from the model noise $w(t)$. Moreover, for simplicity, we consider a network of identical devices, with all nodes having equally reliable sensors, that is, we restrict our analysis to the case $r_i = r \;\forall i \in \mathcal{V}$.

For ease of notation collect all the N measurements in a vector $y(t) = [y_1(t), \dots, y_N(t)]^T$ and all measurement noises in a vector $v(t) = [v_1(t), \dots, v_N(t)]^T$. Then (1) can then be rewritten as:

$$y(t) = \mathbb{1}x(t) + v(t).$$

Since we assumed that all measurement noises are independent and identically distributed $\mathbb{E}[v(t)v^T(t)] = rI_N$.

IV. Proposed Algorithm

The estimation algorithm we analyze in this work is a randomized version of the one that has been analyzed in [1]. It consists of 2 stages. The first stage is a Kalman–like estimate update. At each time instant, each node collects its own measurement $y_i(t)$ and updates its estimate $\hat{x}_i(t)$:

$$\hat{x}_i^{loc}(t+1) = l\hat{x}_i(t) + (1-l)y_i(t).$$

where $l \in (0,1)$ is the Kalman gain. Since $\hat{x}_i^{loc}(t)$ has been updated using only local information, it is called *local estimate*. Again, for ease of notation, we collect all the local estimate in the vector $\hat{x}_{loc}(t) = [\hat{x}_1^{loc}(t), \dots, \hat{x}_N^{loc}(t)]$.

The second phase of the algorithm prescribes that during two consecutive measurements, nodes exchange information with their neighbors to improve their local estimates. In contrast with the previous phase, the outcome of this second phase is called *regional estimate*, $\hat{x}_i^{reg}(t)$, or simply $\hat{x}_i(t)$. Once again, we define the vector $\hat{x}(t) = [\hat{x}_1^{reg}(t), \dots, \hat{x}_N^{reg}(t)]$.

Inspired by the fact that the mean of the local estimates is a sufficient statistic to estimate the centralized estimate, we implement the estimate fusion phase with m consensus steps, as it has been done [1] and [4].

$$\hat{x}(t+1) = Q_{m-1}(t) \dots Q_0(t)\hat{x}_{loc}(t+1) = P(t)\hat{x}_{loc}$$

where $P(t) = Q_{m-1}(t) \dots Q_1(t)$, being product of stochastic matrices, is stochastic.

In [1], it has been analyzed the case of a fixed communication strategy, i.e. $Q_1(t) = \dots = Q_{m-1}(t) = Q \;\forall t$ and therefore $P(t) = P = Q^m \;\forall t$. On the contrary, in this work, to take into account the randomness introduced by the use of random protocols and by unpredictable environments, we assume that for all t and i, $Q_i(t)$ is drawn from an alphabet $\{Q_\alpha,\; \alpha \in \mathcal{A}\}$ of stochastic matrices compatible with the graph $\mathcal{G}$. We will call selection probability, $p_{\mathcal{A}} = \{p_\alpha,\; \alpha \in \mathcal{A}\}$, the probability measure on the set $\{Q_\alpha,\; \alpha \in \mathcal{A}\}$, where p_α is the probability that Q_α is drawn. Therefore

$Q_i(t)$ and consequently $P(t)$ are i.i.d. random (matrix–valued) processes. It is reasonable to assume that $\forall i\ Q_i(t)$ is independent form $w(s)$ and $v(r)\ \forall t, s, r$.

Many examples of models for common random protocols can be found in [6], where a detailed analysis of randomized consensus algorithms is carried out. In particular in [6] is presented a model for both symmetric gossip and broadcast.

In this paper we will focus on the case of symmetric matrices alphabets: $Q_\alpha = Q_\alpha^T\ \forall \alpha \in \mathcal{A}$ and in particular on the symmetric gossip case.

V. Algorithm Mean Square Analysis

Let us define the local and the regional estimation error:

$$\tilde{x}_{loc}(t) = \mathbb{1}x(t) - \hat{x}_{loc}(t) \qquad \text{and} \qquad \tilde{x}(t) = \mathbb{1}x(t) - \hat{x}(t).$$

After simple computations one gets the following description of the error time evolution:

$$\begin{aligned}\tilde{x}_{loc}(t+1) &= (1-l)\tilde{x}(t) + lw(t) + \mathbb{1}v(t)\\ \tilde{x}(t+1) &= P(t)\tilde{x}_{loc}(t+1) = (1-l)P(t)\tilde{x}(t) + \\ &\qquad + lP(t)w(t) + \mathbb{1}v(t).\end{aligned}$$

First of all we will show that the mean error tends to zero when t goes to infinity. To this aim define the local and regional error mean as:

$$\mu_{loc}(t) = \mathbb{E}\left[\tilde{x}_{loc}(t)\right] \qquad \mu(t) = \mathbb{E}\left[\tilde{x}(t)\right].$$

Recalling that w and v are white zero–mean noises, one gets that

$$\mu_{loc}(t+1) = \mathbb{E}\left[(1-l)\tilde{x}(t) + lw(t) + \mathbb{1}v(t)\right] = (1-l)\mu(t)$$

$$\mu(t+1) = \mathbb{E}\left[P(t)\right]\mu_{loc}(t) = (1-l)\mathbb{E}\left[P(t)\right]\mu(t). \qquad (2)$$

Recall than that we are considering stochastic matrices alphabets, $Q_\alpha \mathbb{1} = \mathbb{1}\ \forall \alpha \in \mathcal{A}$, therefore also $\mathbb{E}\left[P(t)\right]$ is stochastic and hence the linear system described by (2) is stable for $l \in (0,1)$. This implies that $\mu(t) \to 0$ as $t \to +\infty$ and since $\mu_{loc}(t+1) = (1-l)\mu(t)$ also $\mu_{loc}(t) \to 0$.

Let us study then the variance of the (local and regional) estimation error:

$$\Sigma_{loc}(t) = \mathbb{E}\left[\tilde{x}_{loc}(t)\tilde{x}_{loc}^T(t)\right] \qquad \Sigma(t) = \mathbb{E}\left[\tilde{x}(t)\tilde{x}^T(t)\right].$$

To compute a recursive formula for the evolution of these two matrices, recall that $P(t)$, $w(s)$, $v(u)$ are independent $\forall t, s, u$. Note moreover that $\tilde{x}(t)$ is independent from $P(t)$, $w(t)$ and $v(t)$, since it depends only on $P(s)$, $w(s)$ and $v(s)$ for $s = 1, \ldots, t-1$. One can easily see then that:

$$\begin{aligned}\Sigma_{loc}(t+1) &= \mathbb{E}\left[\tilde{x}_{loc}(t+1)\tilde{x}_{loc}^T(t+1)\right]\\ &= (1-l)^2\mathbb{E}\left[P(t)\tilde{x}_{loc}(t)\tilde{x}_{loc}^T(t)P^T(t)\right] + l^2 rI + q\mathbb{1}\mathbb{1}^T.\end{aligned}$$

Since

$$\begin{aligned}&\mathbb{E}\left[P(t)\tilde{x}_{loc}(t)\tilde{x}_{loc}^T(t)P^T(t)\right] =\\ &\quad = \mathbb{E}\left[\mathbb{E}\left[P(t)\tilde{x}_{loc}(t)\tilde{x}_{loc}^T(t)P^T(t)|P(t)\right]\right]\\ &\quad = \mathbb{E}\Big[P(t)\mathbb{E}\left[\tilde{x}_{loc}(t)\tilde{x}_{loc}^T(t)|P(t)\right]P^T(t)\Big]\\ &\quad = \mathbb{E}\left[P(t)\Sigma_{loc}(t)P^T(t)\right]\end{aligned}$$

we have that

$$\begin{aligned}\Sigma_{loc}(t+1) = (1-l)^2\mathbb{E}\left[P(t)\Sigma_{loc}(t)P^T(t)\right] + \\ + l^2 rI + q\mathbb{1}\mathbb{1}^T. \qquad (3)\end{aligned}$$

Similarly one gets:

$$\begin{aligned}\Sigma(t+1) = (1-l)^2\mathbb{E}\left[P(t)\Sigma(t)P^T(t)\right] + \\ + l^2 r\mathbb{E}\left[P(t)P^T(t)\right] + q\mathbb{1}\mathbb{1}^T. \qquad (4)\end{aligned}$$

Equations (3) and (4) represent a linear time–invariant system, as it can be more clearly recognized defining $\text{vect}(\Sigma(t)) = s(t)$ and recalling that $\text{vect}(ABC) = (C^T \otimes A)\text{vect}(B)$. Equation (4) can in fact be rewritten as:

$$\begin{aligned}\text{vect}(\Sigma(t+1)) = s(t+1) = (1-l)^2\mathbb{E}[P(t)\otimes P(t)]s(t) + \\ l^2 r\mathbb{E}[P(t)\otimes P(t)]\text{vect}(I) + q\mathbb{1}_{N^2}\end{aligned}$$

that is precisely a linear time–invariant system forced by a constant input. Note that

$$\mathbb{E}[P(t)\otimes P(t)]\mathbb{1}_{N^2} = \mathbb{E}[(P(t)\mathbb{1}_N)\otimes(P(t)\mathbb{1}_N)] = \mathbb{1}_{N^2}.$$

Since $\mathbb{E}[P(t)\otimes P(t)]$ is stochastic we have that $(1-l)^2\mathbb{E}[P(t)\otimes P(t)]$ is stable.

For ease of notation let us define the linear operator $\mathcal{L}(M) = \mathbb{E}[Q_i(t)MQ_i^T(t)]$. Note that $\text{vect}\big(\mathcal{L}(M)\big) = \mathbb{E}[Q(t)\otimes Q(t)]\text{vect}(M)$ and that

$$\begin{aligned}&\mathbb{E}[P(t)MP^T(t)] =\\ &\mathbb{E}[Q_{m-1}(t)\ldots Q_0(t)MQ_0^T(t)\ldots Q_{m-1}^T(t)] = \mathcal{L}^m(M)\end{aligned}$$

Equation (4) can then be rewritten as

$$\Sigma(t+1) = (1-l)^2\mathcal{L}^m(\Sigma(t)) + l^2 r\mathcal{L}^m(I) + q\mathbb{1}\mathbb{1}^T$$

Note moreover that $\mathcal{L}(\mathbb{1}\mathbb{1}^T) = \mathbb{E}(Q\mathbb{1}\mathbb{1}^T Q^T) = \mathbb{1}\mathbb{1}^T$.

We have therefore that, for every initial condition, the system reaches an asymptotically stable equilibrium:

$$\begin{aligned}\Sigma(\infty) &= \sum_{i=0}^{+\infty}(1-l)^{2i}\mathcal{L}^{mi}\left(l^2 r\mathcal{L}^m(I) + q\mathbb{1}\mathbb{1}^T\right)\\ &= l^2 r\sum_{i=0}^{+\infty}(1-l)^{2i}\mathcal{L}^{m(i+1)}(I) + q\sum_{i=0}^{+\infty}(1-l)^{2i}\mathbb{1}\mathbb{1}^T\\ &= l^2 r\sum_{i=0}^{+\infty}(1-l)^{2i}\mathcal{L}^{m(i+1)}(I) + \frac{q}{1-(1-l)^2}\mathbb{1}\mathbb{1}^T. \qquad (5)\end{aligned}$$

We are in particular interested in computing $\text{tr}\Sigma(\infty)$.

$$\text{tr}\Sigma(\infty) = l^2 r\sum_{i=0}^{+\infty}(1-l)^{2i}\text{tr}\mathcal{L}^{m(i+1)}(I) + \frac{qN}{1-(1-l)^2} \qquad (6)$$

Unluckily, we did not manage to find a closed form for:

$$\sum_{i=0}^{+\infty}(1-l)^{2i}\mathcal{L}^{m(i+1)}(I)$$

but we propose an upper–bound for $\text{tr}\mathcal{L}^{m(i+1)}(I)$ that allows to compute an upper–bound on $\text{tr}\Sigma(\infty)$:

Proposition 1: Given any symmetric matrix alphabet $\{Q_\alpha\ \alpha \in \mathcal{A}\}$, then, for all $i \in \mathbb{N}$,

$$\text{tr}\mathcal{L}^i(I) \le \text{tr}\left(\mathbb{E}\left[Q^2(t)\right]^i\right). \qquad (7)$$

Proof: The proposition is a trivial corollary of theorem 4, in the appendix, for $k = n$. □

This upper bound allows us to analyze the $N \times N$ matrix $\mathbb{E}[Q^2(t)]$ rather then the linear operator $\mathcal{L}$, described by the $N^2 \times N^2$ matrix $\mathbb{E}[Q(t)\otimes Q(t)]$. Note, moreover, that $\mathbb{E}[Q^2(t)]$ can be computed quite easily given a communication strategy and a graph while this is not the case for $\mathbb{E}[Q(t)\otimes Q(t)]$, as it has been remarked in [7].

Using the above mentioned upper–bound we get:

$$\mathrm{tr}\Sigma(\infty) = l^2 r \sum_{i=0}^{+\infty} (1-l)^{2i} \mathrm{tr}\mathcal{L}^{m(i+1)}(I) + \frac{qN}{1-(1-l)^2}$$

$$\leq l^2 r \sum_{i=0}^{+\infty} (1-l)^{2i} \mathrm{tr}\mathbb{E}\left[Q^2(t)\right]^{m(i+1)} + \frac{qN}{1-(1-l)^2}$$

$$= l^2 r \sum_{i=0}^{+\infty} (1-l)^{2i} \sum_{j=0}^{N-1} \lambda_j \left(\mathbb{E}\left[Q^2(t)\right]^{m(i+1)}\right) + \frac{qN}{1-(1-l)^2}$$

$$= \sum_{j=0}^{N-1} l^2 r \sum_{i=1}^{+\infty} (1-l)^{2i-2} \lambda_j^{mi}\left(\mathbb{E}\left[Q^2(t)\right]\right) + \frac{qN}{1-(1-l)^2}$$

$$= \sum_{j=0}^{N-1} l^2 r \frac{\lambda_j^m\left(\mathbb{E}\left[Q^2(t)\right]\right)}{1-(1-l)^2\lambda_j^m\left(\mathbb{E}\left[Q^2(t)\right]\right)} + \frac{qN}{1-(1-l)^2}.$$

VI. Optimization

One would like to study the natural optimization problem of finding $l \in (0,1)$ and the probability distribution $p_\alpha = P[Q(t) = Q_\alpha]$ such that $\frac{1}{N}\mathrm{tr}\Sigma(\infty)$ is minimized.

Since a closed form expression for $\mathrm{tr}\Sigma(\infty)$ in not available we rather consider the problem of minimizing the proposed upper–bound on $\mathrm{tr}\Sigma(\infty)$:

$$\frac{1}{N}\mathrm{tr}\Sigma(\infty) \leq J(\{p_\alpha\}_{\alpha\in\mathcal{A}}, l)$$

where

$$J(\{p_\alpha\}_{\alpha\in\mathcal{A}}, l) = \frac{1}{N}\sum_{j=0}^{N-1} l^2 r \frac{\lambda_j^m\left(\mathbb{E}\left[Q^2(t)\right]\right)}{1-(1-l)^2\lambda_j^m\left(\mathbb{E}\left[Q^2(t)\right]\right)} + \\ + \frac{q}{1-(1-l)^2}. \tag{8}$$

The above defined function J is therefore the cost functional we will study while l and $\{p_\alpha\}_{\alpha\in\mathcal{A}}$, that are the parameters we can choose, are going to be our optimization variables.

In [1] it has been shown that the problem of minimizing the cost function:

$$\tilde{J}(M, l) = \frac{1}{N}\sum_{j=0}^{N-1} l^2 r \frac{\lambda_j^m(M)}{1-(1-l)^2\lambda_j^m(M)} + \frac{q}{1-(1-l)^2} \tag{9}$$

over the set of symmetric stochastic matrices compatible with the graph $\mathcal{G}$ is convex in M for l fixed. Moreover, as it can be easily verified, the optimization of the above functional for $l \in (0,1)$ and M fixed is a convex as well. Unluckily the joint optimization problem is not convex.

Our cost functional J in (8) is the result of the composition of the cost functional $\tilde{J}$ in (9) with the map $M(p_\mathcal{A}) : \mathbb{R}^{|\mathcal{A}|} \to \mathbb{S}^N$:

$$p_\mathcal{A} \longmapsto M(p_\mathcal{A}) = \mathbb{E}_{p_\mathcal{A}}[Q(t)^2] = \sum_{\alpha\in\mathcal{A}} p_\alpha Q_\alpha^2.$$

Note that $M(p_\mathcal{A})$ is linear. Therefore if $\tilde{J}$ in (9) is convex, so it is the composed map $J = \tilde{J}(l, M(p_\mathcal{A}))$. We have therefore that the problem of minimizing (8) is convex for $l \in (0,1)$ fixed and it is convex for $p_\mathcal{A}$ fixed and $l \in (0,1)$. It can been shown that it does *not* hold true that it J is jointly convex in $p_\mathcal{A}$ and l for any arbitrary choice of matrix alphabet $\{Q_\alpha\}_{\alpha\in\mathcal{A}}$.

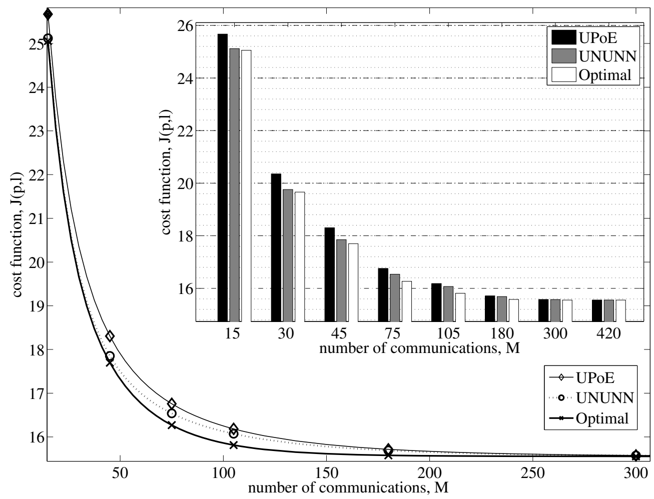

Fig. 1. Cost function for various selection probabilities

VII. Simulation Results

As an example to illustrate the results proposed in the previous sections, we consider a network of $N = 15$ agents. To simulate the behavior of a wireless sensor network, we choose as communication graph a geometric random graph. More precisely, nodes are randomly deployed in a square with side length $1m$ and we assume that two nodes can exchange information if their distance d is less than a visibility radius $r_v = 0.4m$.

The model noise variance of the process to be estimated is chosen $q = 10$ while the measurement noise variance is chosen $r = 100$.

We consider the symmetric gossip strategy: if the link (i,j) is selected, the corresponding consensus matrix Q_{ij} is:

$$Q_{ij} = I - \frac{1}{2}(e_i - e_j)(e_i - e_j)^T$$

where e_i is the vector having all entries equal to 0 except a 1 in position i.

In figure 1 it is reported the comparison of three different selection strategies. More precisely it is depicted the value of the cost function (8) for $l = 0.5$ and for different values of m, the number of communications between two subsequent measurements. The three different selection strategies we compare are:

- UPoE, Uniform Probability over the Edges: All links are selected with the same probability, $p_{ij}^u = \frac{1}{|\mathcal{E}|}$.
- UNUNN Uniform Node Uniform Neighbor Node: At each time instant one node randomly wakes up, with uniform probably among all nodes. This node picks up randomly one node with uniform probability among all its neighbor nodes. Therefore the selection probability of the link (i,j) is: $p_{ij} = \frac{1}{N}\frac{1}{\mathrm{degree}(i)}$.
- Optimal: p^{opt} is the probability that minimize the cost function (8).

Figure 1 shows that, at least in this example, optimization does not seem to give a significant improvement in the estimation performance and the easily implementable strategy UNUNN gives performance which is quite close to the optimal selection strategy. Moreover, it can be noted that, over a certain value of m, $\bar{m} \cong 300$, there is no more performance difference between the three strategies. This is due to the fact that when m is large there are enough communications to reach consensus between two subsequent measurements independently of the selection strategy. No further estimation improvement can therefore be obtained. Analogously, if only very few communications are allowed between 2 measurements, the impact of the optimization is very little.

It has however to be noted that, for m in the range $N < m < 10N$, optimization may play an important role. In fact, the same estimation performance is reach by the optimized strategy with a significantly smaller amount of communications with respect to UNUNN. For instance, figure 1 shows that the performance achieved by the optimal strategy for $m = 150$ is reached by UNUNN only for $m = 300$.

VIII. Conclusions

In this work we analyzed a randomized version of the distributed Kalman filter proposed in [1]. We carried out a mean square-analysis, proving that the error variance remains bounded and providing an upper–bound for this quantity. To prove the upper–bound a stronger result has been derived. It relates the eigenvalues of the matrix $\mathbb{E}[Q_i \dots Q_0 Q_0 \dots Q_i]$ and the eigenvalues of $\mathbb{E}[Q^2]^i$: the first are, in fact, submajorized by the latter.

Moreover, we studied the problem of minimizing the proposed upper–bound and we showed that this problem is convex in the selection probability and in the Kalman gain separately, but not jointly convex.

Simulations are finally presented. They seem to show that optimization does not give a significant improvement in the estimation performance. Whether or not this is a general fact for the proposed algorithm is an issue that deserves to be explored in detail and that will be object of further investigations.

Appendix

Derivation of an upper bound for $\mathrm{tr}(\Sigma(t))$

In this section we will present a result that relates the eigenvalues of the matrix $\mathbb{E}[Q_i \dots Q_0 Q_0 \dots Q_i]$ and the eigenvalues of $\mathbb{E}[Q^2]^i$. We will show, in fact, that the first are submajorized by the latter:

$$\underline{\lambda}\Big(\mathbb{E}\left[Q_{i-1} \dots Q_0 Q_0 \dots Q_{i-1}\right]\Big) \prec_w \underline{\lambda}\left(\mathbb{E}\left[Q_0^2\right]^i\right) \quad \forall i \in \mathbb{N}.$$

That, for decreasingly ordered vector, is equivalent to:

$$\sum_{j=1}^{k} \lambda_j\Big(\mathbb{E}\left[Q_{i-1} \dots Q_0 Q_0 \dots Q_{i-1}\right]\Big) \leq \sum_{j=1}^{k} \lambda_j\left(\mathbb{E}\left[Q_0^2\right]^i\right).$$

For $k = 1, \dots, N$ and $\forall i \in \mathbb{N}$. (10)

The relations of majorization and submajorization are treated in detail in [11], [12], [13], to which we refer the interested reader.

Note that (10) is much stronger than the result proposed in proposition 1.

To prove (10) let us recall the following lemma:

Lemma 2: Let x, y, and z be real, non–negative decreasingly ordered vectors, i.e. $x_1 \geq x_2 \geq \dots \geq x_N \geq 0$

$$y \prec_w x \quad \Rightarrow \quad y \odot z \prec_w x \odot z,$$

where $\odot$ represents the entry–wise (Hadamard) product.

Proof: [14, pag. 92, H.2.c] □

It is worth moreover to recall an important result on the singular values of the product of two matrices:

Lemma 3: Given any two matrices A and B

$$\underline{\sigma}(AB) \prec_w \underline{\sigma}(A) \odot \underline{\sigma}(B). \tag{11}$$

Proof: See [12], [13]. □

We are now ready to prove the main technical contribution of this work:

Theorem 4: Given any symmetric matrix alphabet $\{Q_\alpha \ \alpha \in \mathcal{A}\}$, it holds the following:

$$\underline{\sigma}\Big(\mathbb{E}\left[Q_{i-1} \dots Q_0 Q_0 \dots Q_{i-1}\right]\Big) \prec_w \underline{\sigma}\left(\mathbb{E}\left[Q_0^2\right]^i\right) \ \forall i \in \mathbb{N}.$$

that is

$$\sum_{j=1}^{k} \sigma_j\Big(\mathbb{E}\left[Q_{i-1} \dots Q_0 Q_0 \dots Q_{i-1}\right]\Big) \leq \sum_{j=1}^{k} \sigma_j\left(\mathbb{E}\left[Q_0^2\right]^i\right) \tag{12}$$

$\forall k = 1, \dots, N$ and $\forall i \in \mathbb{N}$.

Proof

We will prove the theorem by induction.

It is trivially true for $i = 1$. Suppose that (12) holds true for i and let us prove that this implies it holds true for $i + 1$. Call $P_i = \mathbb{E}[Q_{i-1} \dots Q_0 Q_0 \dots Q_{i-1}]$. We want to prove

$$\sum_{j=1}^{k} \sigma_j\left(\mathbb{E}\left[Q_i P_i Q_i\right]\right) \leq \sum_{j=1}^{k} \sigma_j^{i+1}\left(\mathbb{E}\left[Q_0^2\right]\right) \ \forall k = 1, \dots, N,$$

that is, $\forall k = 1, \dots, N$

$$\sum_{j=1}^{k} \sigma_j\left(\sum_{\alpha \in \mathcal{A}} p_\alpha Q_\alpha P_i Q_\alpha\right) \leq \sum_{j=1}^{k} \sigma_j^{i+1}\left(\sum_{\alpha \in \mathcal{A}} p_\alpha Q_\alpha^2\right).$$

Let us start by proving that, $\forall k = 1, \dots, N$:

$$\sum_{j=1}^{k} \sigma_j\left(\sum_{\alpha \in \mathcal{A}} p_\alpha Q_\alpha P_i Q_\alpha\right) \leq \sum_{j=1}^{k} \sigma_j\left(P_i\right) \sigma_j\left(\sum_{\alpha \in \mathcal{A}} p_\alpha Q_\alpha^2\right). \tag{13}$$

To this aim recall ([11], [13]) that for any matrix M

$$\sum_{j=1}^{k} \sigma_j\left(M\right) = \max_{U^T U = I_k} \mathrm{tr}\left(U^T M U\right), \tag{14}$$

therefore

$$\sum_{j=1}^{k} \sigma_j\left(\sum_{\alpha \in \mathcal{A}} p_\alpha Q_\alpha P_i Q_\alpha\right) = \\ = \max_{U^T U = I_k} \mathrm{tr}\left(U^T\left(\sum_{\alpha \in \mathcal{A}} p_\alpha Q_\alpha P_i Q_\alpha\right) U\right).$$

Note, moreover, that:

$$\begin{aligned} P_i &= V^T \mathrm{diag}\{[\sigma_1(P_i) \dots \sigma_{k-1}(P_i), \sigma_k(P_i) \dots \sigma_N(P_i)]\} V \\ &\leq V^T \mathrm{diag}\{[\sigma_1 - \sigma_k, \dots, \sigma_{k-1} - \sigma_k, 0, \dots, 0]\} V + \\ &\qquad + \sigma_k(P_i) I = \bar{P} + \sigma_k(P_i) I, \end{aligned}$$

where

$$\bar{P} = V^T \mathrm{diag}\{[\sigma_1 - \sigma_k, \dots, \sigma_{k-1} - \sigma_k, 0, \dots, 0]\} V.$$

Therefore

$$\max_{U^TU=I_k} \mathrm{tr} U^T \left(\sum_{\alpha\in\mathcal{A}} p_\alpha Q_\alpha P_i Q_\alpha\right) U$$
$$\leq \max_{U^TU=I_k} \mathrm{tr} U^T \left(\sum_{\alpha\in\mathcal{A}} p_\alpha Q_\alpha \bar{P} Q_\alpha\right) U)+$$
$$+ \sigma_k(P_i) \mathrm{tr} U^T \left(\sum_{\alpha\in\mathcal{A}} p_\alpha Q_\alpha^2\right) U. \tag{15}$$

One has, again for (14):

$$\sigma_k(P_i)\mathrm{tr}\left(U^T\left(\sum_{\alpha\in\mathcal{A}} p_\alpha Q_\alpha^2\right)U\right)$$
$$\leq \sigma_k(P_i)\sum_{j=1}^{k}\sigma_j\left(\sum_{\alpha\in\mathcal{A}} p_\alpha Q_\alpha^2\right). \tag{16}$$

The other term of the sum in (15) can be upper–bounded by noting that:

$$\mathrm{tr}\Big(U^T\Big(\sum_{\alpha\in\mathcal{A}} p_\alpha Q_\alpha \bar{P} Q_\alpha\Big)U\Big)$$
$$= \sum_{\alpha\in\mathcal{A}} p_\alpha \mathrm{tr}\left(U^T Q_\alpha \bar{P}^{\frac{1}{2}} \bar{P}^{\frac{1}{2}} Q_\alpha U\right)$$
$$= \sum_{\alpha\in\mathcal{A}} p_\alpha \mathrm{tr}\left(\bar{P}^{\frac{1}{2}} Q_\alpha U U^T Q_\alpha \bar{P}^{\frac{1}{2}}\right). \tag{17}$$

Noting that $UU^T \leq I$ in the sense of the positive semidefinite matrices and recalling that, given two positive semidefinite matrices A and B, such that $A \leq B$, it holds that $\mathrm{tr} X^T A X \leq \mathrm{tr} X^T B X$. Form (17), one gets:

$$\sum_{\alpha\in\mathcal{A}} p_\alpha \mathrm{tr}\left(\bar{P}^{\frac{1}{2}} Q_\alpha U U^T Q_\alpha \bar{P}^{\frac{1}{2}}\right) \leq \sum_{\alpha\in\mathcal{A}} p_\alpha \mathrm{tr}\left(\bar{P}^{\frac{1}{2}} Q_\alpha^2 \bar{P}^{\frac{1}{2}}\right)$$
$$= \mathrm{tr}\left(\bar{P}^{\frac{1}{2}}\left(\sum_{\alpha\in\mathcal{A}} p_\alpha Q_\alpha^2\right)\bar{P}^{\frac{1}{2}}\right)$$
$$= \sum_{j=1}^{N}\lambda_j\left(\bar{P}^{\frac{1}{2}}\left(\sum_{\alpha\in\mathcal{A}} p_\alpha Q_\alpha^2\right)\bar{P}^{\frac{1}{2}}\right)$$
$$\leq \sum_{j=1}^{N}\lambda_j\left(\bar{P}\right)\lambda_j\left(\sum_{\alpha\in\mathcal{A}} p_\alpha Q_\alpha^2\right)$$
$$= \sum_{j=1}^{k-1}(\sigma_j(P_i)-\sigma_k(P_i))\sigma_j\left(\sum_{\alpha\in\mathcal{A}} p_\alpha Q_\alpha^2\right). \tag{18}$$

Combining (15), (16) and (18) one gets:

$$\sum_{j=1}^{k}\sigma_j\left(\sum_{\alpha\in\mathcal{A}} p_\alpha Q_\alpha P_i Q_\alpha\right)$$
$$= \max_{U^TU=I_k} \mathrm{tr}\left(U^T\left(\sum_{\alpha\in\mathcal{A}} p_\alpha Q_\alpha P_i Q_\alpha\right)U\right)$$
$$\leq \sum_{j=1}^{k-1}(\sigma_j(P_i)-\sigma_k(P_i))\sigma_j\left(\sum_{\alpha\in\mathcal{A}} p_\alpha Q_\alpha^2\right)+$$
$$\sigma_k(P_i)\sum_{j=1}^{k}\sigma_j\left(\sum_{\alpha\in\mathcal{A}} p_\alpha Q_\alpha^2\right)$$
$$= \sum_{j=1}^{k-1}\sigma_j(P_i)\sigma_j\left(\sum_{\alpha\in\mathcal{A}} p_\alpha Q_\alpha^2\right) - \sum_{j=1}^{k-1}\sigma_k(P_i)\sigma_j\left(\sum_{\alpha\in\mathcal{A}} p_\alpha Q_\alpha^2\right)$$
$$+ \sigma_k(P_i)\sum_{j=1}^{k}\sigma_j\left(\sum_{\alpha\in\mathcal{A}} p_\alpha Q_\alpha^2\right)$$
$$= \sum_{j=1}^{k-1}\sigma_j(P_i)\sigma_j\left(\sum_{\alpha\in\mathcal{A}} p_\alpha Q_\alpha^2\right) + \sigma_k(P_i)\sigma_k\left(\sum_{\alpha\in\mathcal{A}} p_\alpha Q_\alpha^2\right)$$
$$= \sum_{j=1}^{k}\sigma_j(P_i)\sigma_j\left(\sum_{\alpha\in\mathcal{A}} p_\alpha Q_\alpha^2\right).$$

By inductive hypothesis we have that

$$\sum_{j=1}^{k}\sigma_j\left(P_i\right) \leq \sum_{j=1}^{k}\sigma_j^i\left(\sum_{\alpha\in\mathcal{A}} p_\alpha Q_\alpha^2\right).$$

Therefore, according to lemma 2, we get

$$\sum_{j=1}^{k}\sigma_j\left(\sum_{\alpha\in\mathcal{A}} p_\alpha Q_\alpha P_i Q_\alpha\right)$$
$$\leq \sum_{j=1}^{k}\sigma_j^i\left(\sum_{\alpha\in\mathcal{A}} p_\alpha Q_\alpha^2\right)\sigma_j\left(\sum_{\alpha\in\mathcal{A}} p_\alpha Q_\alpha^2\right) =$$
$$= \sum_{j=1}^{k}\sigma_j^{i+1}\left(\sum_{\alpha\in\mathcal{A}} p_\alpha Q_\alpha^2\right).$$

□

References

[1] R. Carli, A. Chiuso, L. Schenato, and S. Zampieri, "Distributed kalman filtering based on consensus strategies," *IEEE Journal in Selected Areas of Communications*, to be published.

[2] R. Olfati-Saber, "Ultrafast consensus in small-world networks," in *Proceedings of the 2005 American Control Conference (ACC'05)*, vol. 4, 2005, pp. 2371– 2378.

[3] ——, "Distributed kalman filter with embedded consensus filters," in *Proceedings of the 44th IEEE Conference on Decision and Control, and European Control Conference*, December 2005.

[4] P. Alriksson 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, July 2006.

[5] ——, "Model based information fusion in sensor networks," in *Proceedings of the 17th IFAC World Congress*, Seoul, Korea, July 2008.

[6] F. Fagnani and S. Zampieri, "Randomized consensus algorithms over large scale networks," *IEEE Journal on Selected Areas of Communications*, to be published.

[7] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE Transactions on Information Theory/ACM Transactions on Networking*, vol. 52, no. 6, pp. 2508–2530, June 2006.

[8] 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)*, December 2006, pp. 1025–1030.

[9] I. D. Schizas, G. B. Giannakis, S. I. Roumeliotis, and A. Ribeiro, "Anytime optimal distributed kalman filtering and smoothing," *Statistical Signal Processing, 2007. SSP '07. IEEE/SP 14th Workshop on*, pp. 368–372, Aug. 2007.

[10] D. Bickson, O. Shental, and D. Dolev, "Distributed kalman filter via gaussian belief propagation," *Communication, Control, and Computing, 2008 46th Annual Allerton Conference on*, pp. 628–635, Sept. 2008.

[11] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, Cambridge, UK, 1990.

[12] ——, *Topics in Matrix Analysis*. Cambridge University Press, Cambridge, UK, 1991.

[13] R. Bathia, *Matrix Analysis*. Springer, 1997.

[14] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications*, ser. Mathematics in science and engineering. Academic press, New York, 1979, vol. 143.