# A majorization inequality and its application to distributed Kalman filtering ⋆

Simone Del Favero, Sandro Zampieri

*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/.*

**Abstract**

In the analysis of a recently proposed distributed estimation algorithm based on the Kalman filtering and on gossip iterations, we needed to apply a new inequality which is valid for i.i.d. matrix valued random processes. This inequality can be useful in the analysis of the convergence rate of general jump Markov linear systems.

In this paper we present this inequality. This is based on the theory of majorization and on its use in the analysis of the singular values. Finally we will show the impact of this inequality on the performance analysis of gossip based distributed Kalman filters.

*Key words:*
Majorization inequality, trace inequality, expectation of matrix-valued random variables' product, jump-Markov linear system, distributed estimation, Kalman Filter, randomized gossip.

## 1 Introduction

Assume we have a Markov random process $\mathbf{Q}(t)$, $t = 0, 1, 2, \ldots$, taking values in the set $\mathbb{R}^{N\times N}$ of $N \times N$ matrices, and assume we want to evaluate how "big" is the positive semidefinite matrix

$$P(t) := \mathbb{E}[\mathbf{Q}(t-1)\ldots\mathbf{Q}(1)\mathbf{Q}(0)\mathbf{Q}^T(0)\mathbf{Q}^T(1)\ldots\mathbf{Q}^T(t-1)] \tag{1}$$

This matrix naturally appears when analyzing the discrete time jump Markov linear system [8]

$$\mathbf{x}(t+1) = \mathbf{Q}(t)\mathbf{x}(t) \tag{2}$$

where $\mathbf{x}(t) \in \mathbb{R}^N$ for all $t = 0, 1, 2, \ldots$ is called the state of the system. Indeed, it is easy to see that

$$\mathbb{E}[||\mathbf{x}(t)||^2] = \mathbf{x}(0)^T P(t)\mathbf{x}(0)$$

where $||\cdot||$ means the Euclidean norm. Therefore the evolution of the state $\mathbf{x}(t)$ is well described by the evolution of the positive semidefinite matrix $P(t)$. One way to evaluate how big is $P(t)$ is through its 2-norm, which coincides with the largest eigenvalue of $P(t)$. An alternative way is to consider the trace of $P(t)$, which coincides with the sum of all the eigenvalues of $P(t)$. In this paper we propose an upper bound on $P(t)$ in terms of the matrix $P(1)^t$ for the particular case in which $\mathbf{Q}(t)$ are independent. This bound proved to be very useful in analyzing a particular jump Markov linear system arising in our study of recently proposed distributed estimation algorithms.

⋆ The research leading to these results has received funding from the European Community's Seventh Framework Programme under agreement n. FP7-ICT-223866-FeedNetBack

*Email addresses:* `simone.delfavero@dei.unipd.it` (Simone Del Favero), `zampi@dei.unipd.it` (Sandro Zampieri).

Recently there has been an increasing interest in the sensor networks technology and in the design of algorithms that enable to exploit the potentials of such apparatus [17,1,11]. In particular, there have been proposed distributed estimation algorithms which allow to take decisions from the huge amount of data collected by a great number of sensors which can exchange information through wireless communication. Some of the most effective distributed estimation methodologies are based of consensus or gossip algorithms (see [5,10,7,19,20]and the references therein). Examples of distributed estimation methodologies based on this technique can be found in [6,2,3,16,15,18]. Another example is described below .

Let us consider $N$ sensors, labeled with the elements of the set $\mathcal{V} = \{1 \dots N\}$. These sensors can communicate each other according to a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, meaning that we have an edge $(j,i) \in \mathcal{E}$ if and only if $j$ can transmit information to $i$.

Assume that we want to estimate, by means of such a sensor network, a discrete-time scalar random process of the form

$$\mathbf{x}(t+1) = \mathbf{x}(t) + \mathbf{w}(t) \tag{3}$$

where $\mathbf{x} \in \mathbb{R}$ and $\mathbf{w}(t)$, known as model noise, is a white noise with zero mean and variance $q$. This is the model of a random walk. Each sensor $i$ of the network can collect noisy measurements the state $\mathbf{x}(t)$, namely

$$\mathbf{y}_i(t) = \mathbf{x}(t) + \mathbf{v}_i(t) \tag{4}$$

where the measurement noise, $\mathbf{v}_i(t)$, is a white noise. Assume for simplicity that $\mathbf{v}_i(t)$ are independent, have all zero mean and the same variance $r$. Assume moreover that they are all independent of the model noise $\mathbf{w}(t)$.

The estimation algorithm consists of 2 stages. The first stage is a Kalman–like estimate update in which, at each time instant, every node collects its own measurement $\mathbf{y}_i(t)$ and updates its estimate $\hat{\mathbf{x}}_i(t)$

$$\hat{\mathbf{x}}_i^{loc}(t+1) = \ell\hat{\mathbf{x}}_i(t) + (1-\ell)\mathbf{y}_i(t).$$

where $\ell \in (0,1)$ is the Kalman gain. Since $\hat{\mathbf{x}}_i^{loc}(t)$ has been updated using only local information, it is called *local estimate.*

The second phase of the algorithm prescribes that, during two consecutive measurements, nodes improve their local estimates by exchanging information along the communication links allowed by the graph $\mathcal{G}$. More precisely, the second phase consists in the following iteration

$$\hat{\mathbf{x}}_i(t+1) = \sum_{j=1}^{N} \mathbf{Q}_{ij}(t)\hat{\mathbf{x}}_j^{loc}(t)$$

where $\mathbf{Q}_{ij}(t)$ is different for zero only if $(j,i) \in \mathcal{E}$. If we let $\mathbf{Q}(t)$ be the matrix with entries $\mathbf{Q}_{ij}(t)$ and we introduce the column vectors $\mathbf{y}(t), \mathbf{x}^{loc}(t), \mathbf{x}(t)$ having entries $\mathbf{y}_i(t), \mathbf{x}_i^{loc}(t), \mathbf{x}_i(t)$, respectively, we can write

$$\hat{\mathbf{x}}(t+1) = \mathbf{Q}(t)\left(\ell\hat{\mathbf{x}}(t) + (1-\ell)\mathbf{y}(t)\right) \tag{5}$$

To take into account the randomness introduced by the use of random protocols, [5], and by presence of unpredictable environments, we assume that $\mathbf{Q}(t)$ is randomly drown from an alphabet of stochastic matrices $Q$ compatible with the graph, namely $Q_{i,j} \neq 0$ only if $(j,i) \in \mathcal{E}$. In fact, the compatibility of the matrices with the graph ensures that the algorithm respects the communication constraints and the stochasticity is known to be sufficient to guarantee unbiasedness of the estimate (see for instance [6]).

We assume that $\mathbf{Q}(t)$ is an independent and identically distributed matrix valued random process, moreover it is independent of $\mathbf{v}(s)$ and $\mathbf{w}(r)$ at all time instants, $\forall t, s, r$. Furthermore we restrict our analysis only to symmetric process: $Q(t) = Q(t)^T$ $\forall t$. The assumption of independence is not very restrictive in the sensor networks applications. On the contrary symmetry is a rather restrictive hypothesis. In fact, it is not true that all possible consensus algorithms prescribe the use of a symmetric matrix and an extension of the proposed result to the case of non symmetric stochastic matrices is subject of our present research. Nevertheless, the symmetric case is of interest itself, since various of the algorithms that have been proposed, such a symmetric gossip [5], prescribe the use of symmetric matrices only.

The algorithm under study (5) turns out to be a variation of the jump Markov linear system (2). We will show that in the performance analysis of this algorithm a bound of the matrix $P(t)$ defined in (1) will be extremely useful.

### *1.1 Paper Organization*

The paper is organized as follows: in section 2 we define the notation used in the paper while in section 3 we recall some basic concepts and some known results in majorizarion theory. In section 4 the main result is stated, we comment on the possibility to extend the obtained result and we formulate a conjecture. In section 5 the proof of the previously presented result is presented and commented. Section 6 is devoted to illustrate the application of the result to a problem of distributed estimation. Finally in section 7 we draw some conclusions.

## 2 Notation

In this paper we will use lower case letters, $a, b, c, \dots$, to denote scalars or vectors, capital letters, $A, B, C, \dots$, to denote matrices and bold letters, $\mathbf{a}, \mathbf{A}, \mathbf{b}, \mathbf{B}, \mathbf{c}, \mathbf{C}, \dots$, to denote random variables. Moreover we will denote with $\otimes$ the Kronecker product while $\odot$ will represent the Hadamard product, i.e. the entry–wise product.

Given a matrix $M \in \mathbb{R}^{N\times N}$ we will denote with $\underline{\sigma}(M)$ the vector in $\mathbb{R}^N$ formed by the singular values of $M$ decreasingly ordered and with $\underline{\lambda}(M)$ the vector in $\mathbb{C}^N$ formed by the eigenvalues of $M$ ordered so that $|\underline{\lambda}_1(M)| \geq \cdots \geq |\underline{\lambda}_N(M)|$, where each eigenvalue appears as many times as its algebraic multiplicity. Recall moreover that for any normal matrix $N$, $NN^T = N^TN$, $\underline{\sigma}(N) \equiv |\underline{\lambda}(N)|$, where with $|\underline{\lambda}(N)|$ we mean the $\mathbb{R}^n$ vector having entries $|\lambda_i(N)|$ .

Through the paper, for ease of notation, we will write $\sigma_j^i(M)$ to denote $(\sigma_j(M))^i$. Note that, using the notation we just defined, $\lambda_j(M^i) = \lambda_j^i(M)$ and, therefore, for a normal matrix $N$, $\sigma_j(N^i) = \sigma_j^i(N)$.

## 3 Preliminary Results

Let us begin by reviewing some basic concepts about majorization.

**Definition 1** *Given two vector $x, y \in \mathbb{R}^N$ whose components are ordered decreasingly, i.e. $x_1 \geq x_2 \geq \cdots \geq x_N$, we say that $x$ submajorize $y$ and we write $y \prec_w x$ if*

$$\sum_{i=1}^{k} y_i \leq \sum_{i=1}^{k} x_i \qquad \forall k = 1, \dots, N$$

*If moreover the inequality for $k = N$ holds as an equality, then we say that $x$ majorize $y$ and we write $y \prec x$*

A broad literature on majorization theory and its applications is available. An introduction to the topic can be found, for instance, in [12,13]. Majorization theory is extensively treated in [4] and in [14], a book fully devoted to this topic.

In particular, recall that the following lemma holds

**Lemma 1**
*Let $x$, $y$, and $z$ be real, non–negative decreasingly ordered vectors. Then*

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

**Proof** See [14, page 92, H.2.c] □

Recall moreover an important result on the singular values of the product of two matrices

**Lemma 2** *Given any two matrices $M_1$ and $M_2$*

$$\underline{\sigma}(M_1 M_2) \prec_w \underline{\sigma}(M_1) \odot \underline{\sigma}(M_2). \tag{6}$$

**Proof** See [13,4]. □

Furthermore there is an important result specifying the relation between eigenvalues and singular values of a matrix

**Lemma 3** *Given any matrix $M$*

$$|\underline{\lambda}(M)| \prec_w \underline{\sigma}(M). \tag{7}$$

**Proof** See [13,4]. □

Combining the previous two results we get that, given any two normal matrices $N_1$ and $N_2$,

$$|\underline{\lambda}(N_1 N_2)| \prec_w \underline{\sigma}(N_1) \odot \underline{\sigma}(N_2) = |\underline{\lambda}(N_1)| \odot |\underline{\lambda}(N_2)|. \tag{8}$$

We will use this inequality in the following.

## 4 Main Result

Let us introduce a finite indexes set $\mathbb{A}$ and consider a finite alphabet of normal matrices

$$\{Q_\alpha \ \alpha \in \mathbb{A}\}, \qquad Q_\alpha Q_\alpha^T = Q_\alpha^T Q_\alpha \quad \forall \alpha \in \mathbb{A}$$

from which we assume to randomly draw a matrix and denote with $\mathbb{P}$ the probability measure on the alphabet. We will refer to $\mathbb{P}$ as selection probability and we will denote $p_\alpha$ the probability that $Q_\alpha$ is drawn. Consider then the random process $\{\mathbf{Q}(t)\}_{t \in \mathbb{N}}$ of independent and identically distributed random variables $\mathbb{P}[\mathbf{Q}(t) = Q_\alpha] = p_\alpha \ \forall t$ that describes independent extractions form the alphabet $\{Q_\alpha \ \alpha \in \mathbb{A}\}$.

We want to study the matrix

$$\mathbb{E}\left[\mathbf{Q}(i-1)\dots\mathbf{Q}(0)\mathbf{Q}(0)^T\dots\mathbf{Q}(i-1)^T\right].$$

We present now a result that relates the spectrum of this positive semidefinite matrix to the spectrum of $\mathbb{E}\left[\mathbf{Q}(0)\mathbf{Q}(0)^T\right]^i$. The proof is reported in Section 5.

**Theorem 4** *Given any finite normal matrix alphabet $\{Q_\alpha \ \alpha \in \mathbb{A}\}$, $\forall i \in \mathbb{N}$ we have that*

$$\underline{\sigma}\Big(\mathbb{E}\left[\mathbf{Q}(i-1)\dots\mathbf{Q}(0)\mathbf{Q}(0)^T\dots\mathbf{Q}(i-1)^T\right]\Big) \prec_w \underline{\sigma}\left(\mathbb{E}\left[\mathbf{Q}(0)\mathbf{Q}(0)^T\right]^i\right). \tag{9}$$

The above result, for $k = n$, gives the following trace inequality

**Corollary 5** *Given any finite normal matrix alphabet $\{Q_\alpha \ \alpha \in \mathbb{A}\}$, $\forall i \in \mathbb{N}$ we have that*

$$\operatorname{tr}\Big(\mathbb{E}\left[\mathbf{Q}(i-1)\dots\mathbf{Q}(0)\mathbf{Q}(0)^T\dots\mathbf{Q}(i-1)^T\right]\Big) \leq \operatorname{tr}\left(\mathbb{E}\left[\mathbf{Q}(0)\mathbf{Q}(0)^T\right]^i\right).$$

Moreover, again as a trivial corollary of the above result, we get an inequality on the largest eigenvalue $\lambda_{\max}$, already proved in [10]

**Corollary 6** *Given any finite normal matrix alphabet* $\{Q_\alpha \ \alpha \in \mathbb{A}\}$,

$$\lambda_{\max}\Big(\mathbb{E}\left[\mathbf{Q}(i-1)\ldots\mathbf{Q}(0)\mathbf{Q}(0)^T\ldots\mathbf{Q}(i-1)^T\right]\Big) \leq \lambda^i_{\max}\left(\mathbb{E}\left[\mathbf{Q}(0)\mathbf{Q}(0)^T\right]\right).$$

### *4.1 Comments on the result*

One might wonder if it is possible to relax the hypothesis of normality The answer is no, as the following counterexample shows

**Counterexample** Let $\mathbb{A} = \{1, 2\}$, let $p_{\mathbb{A}} = \{p_1 = \frac{1}{3},\ p_2 = \frac{2}{3}\}$ and let

$$\{Q_\alpha \ \alpha \in \mathbb{A}\} = \left\{ Q_1 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \quad , \quad Q_2 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}.$$

Note that $Q_1$ is not normal since

$$Q_1 Q_1^T = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \neq \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = Q_1^T Q_1$$

We get that

$$\mathbb{E}\left[\mathbf{Q}(1)\mathbf{Q}(0)\mathbf{Q}(0)^T\mathbf{Q}(1)^T\right] = \frac{1}{9}\begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix} \qquad \mathbb{E}\left[\mathbf{Q}(0)\mathbf{Q}(0)^T\right]^2 = \frac{1}{9}\begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix}$$

Therefore it does not hold that $\underline{\sigma}(\mathbb{E}\left[\mathbf{Q}(1)\mathbf{Q}(0)\mathbf{Q}(0)^T\mathbf{Q}(1)^T\right]) \prec_w \underline{\sigma}(\mathbb{E}\left[\mathbf{Q}(0)\mathbf{Q}(0)^T\right]^2)$ since

$$\mathrm{tr}\ (\mathbb{E}\left[\mathbf{Q}(1)\mathbf{Q}(0)\mathbf{Q}(0)^T\mathbf{Q}(1)^T\right]) = \frac{6}{9} > \frac{5}{9} = \mathrm{tr}\ (\mathbb{E}\left[\mathbf{Q}(0)\mathbf{Q}(0)^T\right]^2)$$

$\square$

One might wonder then if the inequality holds true for any singular value, i.e. if

$$\underline{\sigma}_i\Big(\mathbb{E}\left[\mathbf{Q}(i-1)\ldots\mathbf{Q}(0)\mathbf{Q}(0)^T\ldots\mathbf{Q}(i-1)^T\right]\Big) \leq \underline{\sigma}_i\left(\mathbb{E}\left[\mathbf{Q}(0)\mathbf{Q}(0)^T\right]^i\right) \qquad \forall i = 1\ldots N.$$

This is false as shown in the following counterexample

**Counterexample** Let $\mathbb{A} = \{1, 2\}$, let $p_{\mathbb{A}} = \{p_1 = \frac{1}{2},\ p_2 = \frac{1}{2}\}$ and let

$$\{Q_\alpha \ \alpha \in \mathbb{A}\} = \left\{ Q_1 = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \quad , \quad Q_2 = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \right\}.$$

After some simple computations one gets that

$$\mathbb{E}[Q(0)Q^T(0)]^2 = \left(p_1 Q_1 Q_1^T + p_2 Q_2 Q_2^T\right)^2 = \begin{bmatrix} 4.5 & 9 \\ 9 & 22.5 \end{bmatrix}$$

$$\underline{\sigma}(\mathbb{E}[Q(0)Q^T(0)]^2) = \begin{bmatrix} 26.2279 \\ 0.7721 \end{bmatrix}$$

while (omitting transposes, since matrices are symmetric)

$$\mathbb{E}[Q(1)Q(0)Q(0)Q(1)] = p_1^2 Q_1^2 Q_1^2 + p_1 p_2 Q_1 Q_2 Q_2 Q_1 + p_2 p_1 Q_2 Q_1 Q_1 Q_2 + p_2^2 Q_2 Q_2 = \begin{bmatrix} 5.25 & 9 \\ 9 & 21.75 \end{bmatrix}$$

$$\underline{\sigma}(\mathbb{E}[Q(1)Q(0)Q(0)Q(1)]) = \begin{bmatrix} 25.7091 \\ 1.2909 \end{bmatrix}$$

Hence:

$$\underline{\sigma}_1(\mathbb{E}[Q(1)Q(0)Q(0)Q(1)]) = 25.7091 \leq 26.2279 = \underline{\sigma}_1(\mathbb{E}[Q(0)Q(0)]^2)$$

and

$$\text{tr }(\mathbb{E}[Q(1)Q(0)Q(0)Q(1)]) = 27 \leq 27 = \text{tr }(\mathbb{E}[Q(0)Q(0)]^2)$$

as prescribed by inequality (9), but

$$\underline{\sigma}_2(\mathbb{E}[Q(1)Q(0)Q(0)Q(1)]) = 1.2909 \not\leq 0.7721 = \underline{\sigma}_2(\mathbb{E}[Q(0)Q^T(0)]^2)$$

□

Many numerical experiments instead supported the conjecture that the statement of Theorem 4 is still holds true in the case of stochastic, possibly non-normal, matrices. The proposed conjecture have relevant applications in distributed estimation over wireless sensors network since in that case the matrices of the alphabet are known to be always stochastic.

## 5 Proof of Theorem 4

The proof is based on the following lemma

**Lemma 7** *Given any normal matrix alphabet $\{Q_\alpha \; \alpha \in \mathbb{A}\}$, and any symmetric positive semi-definite matrix $P \geq 0$, we have that*

$$\underline{\sigma}\Big(\mathbb{E}\left[\mathbf{Q}(0)P\mathbf{Q}(0)^T\right]\Big) \prec_w \underline{\sigma}(P) \odot \underline{\sigma}\left(\mathbb{E}\left[\mathbf{Q}(0)\mathbf{Q}(0)^T\right]\right). \tag{10}$$

**Proof** Note that the thesis (10) is equivalent to prove that, $\forall k = 1, \dots, N$

$$\sum_{j=1}^{k} \sigma_j\left(\sum_{\alpha\in\mathbb{A}} p_\alpha Q_\alpha P Q_\alpha^T\right) \leq \sum_{j=1}^{k} \sigma_j(P)\, \sigma_j\left(\sum_{\alpha\in\mathbb{A}} p_\alpha Q_\alpha Q_\alpha^T\right). \tag{11}$$

To this aim recall ([13,4]) that for any matrix $M$

$$\sum_{j=1}^{k} \sigma_j(M) = \max_{U^T U = I_k} \text{tr }\left(U^T M U\right), \tag{12}$$

therefore

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

Note, moreover, that

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

where

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

Therefore

$$\begin{aligned}
\max_{U^T U = I_k} \mathrm{tr}\, U^T \left( \sum_{\alpha \in \mathbb{A}} p_\alpha Q_\alpha P Q_\alpha^T \right) U \leq \max_{U^T U = I_k} \mathrm{tr}\, U^T \left( \sum_{\alpha \in \mathbb{A}} p_\alpha Q_\alpha \bar{P} Q_\alpha^T \right) U + \\
+ \sigma_k(P) \mathrm{tr}\, U^T \left( \sum_{\alpha \in \mathbb{A}} p_\alpha Q_\alpha Q_\alpha^T \right) U. \qquad (13)
\end{aligned}$$

One has, again for (12),

$$\sigma_k(P) \mathrm{tr}\, \left( U^T \left( \sum_{\alpha \in \mathbb{A}} p_\alpha Q_\alpha Q_\alpha^T \right) U \right) \leq \sigma_k(P) \sum_{j=1}^{k} \sigma_j \left( \sum_{\alpha \in \mathbb{A}} p_\alpha Q_\alpha Q_\alpha^T \right). \qquad (14)$$

The other term of the sum in (13) can be upper–bounded by noting that

$$\begin{aligned}
\mathrm{tr}\, \left( U^T \left( \sum_{\alpha \in \mathbb{A}} p_\alpha Q_\alpha \bar{P} Q_\alpha^T \right) U \right) &= \sum_{\alpha \in \mathbb{A}} p_\alpha \mathrm{tr}\, \left( U^T Q_\alpha \bar{P}^{\frac{1}{2}} \bar{P}^{\frac{1}{2}} Q_\alpha^T U \right) \\
&= \sum_{\alpha \in \mathbb{A}} p_\alpha \mathrm{tr}\, \left( \bar{P}^{\frac{1}{2}} Q_\alpha^T U U^T Q_\alpha \bar{P}^{\frac{1}{2}} \right). \qquad (15)
\end{aligned}$$

Noting that

$$U U^T \leq I$$

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 (15), one gets

$$\begin{aligned}
\sum_{\alpha \in \mathbb{A}} p_\alpha \mathrm{tr}\, \left( \bar{P}^{\frac{1}{2}} Q_\alpha^T U U^T Q_\alpha \bar{P}^{\frac{1}{2}} \right) &\leq \sum_{\alpha \in \mathbb{A}} p_\alpha \mathrm{tr}\, \left( \bar{P}^{\frac{1}{2}} Q_\alpha^T Q_\alpha \bar{P}^{\frac{1}{2}} \right) \\
&= \mathrm{tr}\, \left( \bar{P}^{\frac{1}{2}} \left( \sum_{\alpha \in \mathbb{A}} p_\alpha Q_\alpha^T Q_\alpha \right) \bar{P}^{\frac{1}{2}} \right) \\
&= \sum_{j=1}^{N} \lambda_j \left( \bar{P}^{\frac{1}{2}} \left( \sum_{\alpha \in \mathbb{A}} p_\alpha Q_\alpha^T Q_\alpha \right) \bar{P}^{\frac{1}{2}} \right)
\end{aligned} \qquad (16)$$

Using (8) and recalling that, for any semi–positive definite matrix $A \geq 0$, $\underline{\lambda}(A) \in \mathbb{R}^N$ and $\underline{\lambda}(A) \geq 0$ one gets

$$\sum_{j=1}^{N} \lambda_j \left( \bar{P}^{\frac{1}{2}} \left( \sum_{\alpha \in \mathbb{A}} p_\alpha Q_\alpha^T Q_\alpha \right) \bar{P}^{\frac{1}{2}} \right) \leq \sum_{j=1}^{N} \lambda_j \left( \bar{P}^{\frac{1}{2}} \right) \lambda_j \left( \sum_{\alpha \in \mathbb{A}} p_\alpha Q_\alpha^T Q_\alpha \right) \lambda_j \left( \bar{P}^{\frac{1}{2}} \right)$$

Therefore

$$\sum_{\alpha\in\mathbb{A}} p_\alpha \mathrm{tr}\,\left(\bar{P}^{\frac{1}{2}} Q_\alpha^T U U^T Q_\alpha \bar{P}^{\frac{1}{2}}\right) \leq \sum_{j=1}^{N} \lambda_j\left(\bar{P}\right)\lambda_j\left(\sum_{\alpha\in\mathbb{A}} p_\alpha Q_\alpha^T Q_\alpha\right)$$
$$= \sum_{j=1}^{k-1}(\sigma_j(P)-\sigma_k(P))\sigma_j\left(\sum_{\alpha\in\mathbb{A}} p_\alpha Q_\alpha^T Q_\alpha\right). \tag{17}$$

Now the assumptions of normality of the matrix alphabet,

$$Q_\alpha^T Q_\alpha = Q_\alpha Q_\alpha^T,$$

comes into play, allowing to combine (13), (14) and (17) to get

$$\begin{aligned}
\sum_{j=1}^{k}\sigma_j\left(\sum_{\alpha\in\mathbb{A}} p_\alpha Q_\alpha P Q_\alpha^T\right) &= \max_{U^TU=I_k} \mathrm{tr}\,\left(U^T\left(\sum_{\alpha\in\mathbb{A}} p_\alpha Q_\alpha P Q_\alpha^T\right)U\right)\\
&\leq \sum_{j=1}^{k-1}(\sigma_j(P)-\sigma_k(P))\sigma_j\left(\sum_{\alpha\in\mathbb{A}} p_\alpha Q_\alpha Q_\alpha^T\right) + \sigma_k(P)\sum_{j=1}^{k}\sigma_j\left(\sum_{\alpha\in\mathbb{A}} p_\alpha Q_\alpha Q_\alpha^T\right)\\
&= \sum_{j=1}^{k-1}\sigma_j(P)\sigma_j\left(\sum_{\alpha\in\mathbb{A}} p_\alpha Q_\alpha Q_\alpha^T\right) - \sum_{j=1}^{k-1}\sigma_k(P)\sigma_j\left(\sum_{\alpha\in\mathbb{A}} p_\alpha Q_\alpha Q_\alpha^T\right) +\\
&\qquad + \sigma_k(P)\sum_{j=1}^{k}\sigma_j\left(\sum_{\alpha\in\mathbb{A}} p_\alpha Q_\alpha Q_\alpha^T\right)\\
&= \sum_{j=1}^{k-1}\sigma_j(P)\sigma_j\left(\sum_{\alpha\in\mathbb{A}} p_\alpha Q_\alpha Q_\alpha^T\right) + \sigma_k(P)\sigma_k\left(\sum_{\alpha\in\mathbb{A}} p_\alpha Q_\alpha Q_\alpha^T\right)\\
&= \sum_{j=1}^{k}\sigma_j(P)\sigma_j\left(\sum_{\alpha\in\mathbb{A}} p_\alpha Q_\alpha Q_\alpha^T\right).
\end{aligned} \tag{18}$$

which concludes the proof of the lemma. □

**Proof of the theorem** We will prove the theorem by induction.
It is trivially true for $i = 1$.
Suppose that the thesis (9) holds true for $i$ and let us prove that this implies it holds true for $i+1$. Let us define

$$P(i) = \mathbb{E}\left[\mathbf{Q}(i-1)\ldots\mathbf{Q}(0)\mathbf{Q}(0)^T\ldots\mathbf{Q}(i-1)^T\right]$$

Note that $P(i)$ is symmetric and positive semidefinite $P(i) \geq 0\ \forall i$
Recalling that $\mathbf{Q}(0)\ldots\mathbf{Q}(i)$ are independent and identically distributed, one gets

$$\begin{aligned}
P(i+1) &= \mathbb{E}\left[\mathbf{Q}(i)\mathbf{Q}(i-1)\ldots\mathbf{Q}(0)\mathbf{Q}(0)^T\ldots\mathbf{Q}(i-1)^T\mathbf{Q}(i)^T\right]\\
&= \mathbb{E}\left[\mathbb{E}\left[\mathbf{Q}(i)\mathbf{Q}(i-1)\ldots\mathbf{Q}(0)\mathbf{Q}(0)^T\ldots\mathbf{Q}(i-1)^T\mathbf{Q}(i)^T|\mathbf{Q}(i)\right]\right]\\
&= \mathbb{E}\left[\mathbf{Q}(i)\mathbb{E}\left[\mathbf{Q}(i-1)\ldots\mathbf{Q}(0)\mathbf{Q}(0)^T\ldots\mathbf{Q}(i-1)^T|\mathbf{Q}(i)\right]\mathbf{Q}(i)^T\right]\\
&= \mathbb{E}\left[\mathbf{Q}(i)\mathbb{E}\left[\mathbf{Q}(i-1)\ldots\mathbf{Q}(0)\mathbf{Q}(0)^T\ldots\mathbf{Q}(i-1)^T\right]\mathbf{Q}(i)^T\right]\\
&= \mathbb{E}\left[\mathbf{Q}(i)P(i)\mathbf{Q}(i)^T\right] = \mathbb{E}\left[\mathbf{Q}(0)P(i)\mathbf{Q}(0)^T\right],
\end{aligned}$$

where in the last equality follows from the fact that the process $Q(t)$ is i.i.d..We want therefore to prove that

$$\sum_{j=1}^{k}\sigma_j\left(\mathbb{E}\left[\mathbf{Q}(0)P(i)\mathbf{Q}(0)^T\right]\right) \leq \sum_{j=1}^{k}\sigma_j^{i+1}\left(\mathbb{E}\left[\mathbf{Q}(0)\mathbf{Q}(0)^T\right]\right)\ \forall k=1,\ldots,N$$

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

$$\sum_{j=1}^{k}\sigma_j\left(\sum_{\alpha\in\mathbb{A}}p_\alpha Q_\alpha P(i)Q_\alpha^T\right)\le\sum_{j=1}^{k}\sigma_j^{i+1}\left(\sum_{\alpha\in\mathbb{A}}p_\alpha Q_\alpha Q_\alpha^T\right).$$

From lemma 7 we know that

$$\sum_{j=1}^{k}\sigma_j\Big(\mathbb{E}\left[\mathbf{Q}(0)P(i)\mathbf{Q}(0)^T\right]\Big)\le\sum_{j=1}^{k}\sigma_j\left(P(i)\right)\sigma_j\left(\mathbb{E}\left[\mathbf{Q}(0)\mathbf{Q}(0)^T\right]\right) \tag{19}$$

Moreover, by inductive hypothesis, we have that

$$\sum_{j=1}^{k}\sigma_j\left(P(i)\right)\le\sum_{j=1}^{k}\sigma_j^{i}\left(\mathbb{E}\left[\mathbf{Q}(0)\mathbf{Q}(0)^T\right]\right).$$

Therefore, using lemma 1, we get

$$\begin{aligned}\sum_{j=1}^{k}\sigma_j\left(\mathbb{E}\left[\mathbf{Q}(0)P(i)\mathbf{Q}(0)^T\right]\right)&\le\sum_{j=1}^{k}\sigma_j^{i}\left(\mathbb{E}\left[\mathbf{Q}(0)\mathbf{Q}(0)^T\right]\right)\sigma_j\left(\mathbb{E}\left[\mathbf{Q}(0)\mathbf{Q}(0)^T\right]\right)=\\&=\sum_{j=1}^{k}\sigma_j^{i+1}\left(\mathbb{E}\left[\mathbf{Q}(0)\mathbf{Q}(0)^T\right]\right),\end{aligned}$$

that completes the proof. □

**Remark 1** *Note that the assumption of normality allows to combine (13), (14) and (17) to get (18) which amounts to the thesis.*

**Remark 2** *The proposed theorem holds true also in the case of infinite alphabet sets, both countable and uncountable. In fact, the assumption of finiteness of the alphabet does not play a central role in the proof and can be dropped, taking care to guarantee the boundedness of the expectations that appear in the proof. To this aim, as one can easily see, it is sufficient to add the further assumption of boundedness of* $\mathbb{E}\left[\mathbf{Q}(0)\mathbf{Q}(0)^T\right]$.

**Remark 3** *Note moreover that in the proposed proof it has never been used the fact that* $\mathbb{P}(\alpha)$ *is such that* $\sum_{\alpha\in\mathbb{A}}p_\alpha=1$. *The presented result holds therefore for any* $p(\alpha)$ *of non-negative weight function,* $p(\alpha)\ge 0\ \forall\alpha\in\mathbb{A}$ *even if* $\sum_{\alpha\in\mathbb{A}}p_\alpha\ne 1$ ($\int_{\mathbb{A}}p(\alpha)d\alpha\ne 1$).

## 6 Application of the result to a distributed estimation problem

In the following section we describe a relevant application of the proposed result in the field of distributed estimation for sensor networks. A more detailed analysis of the specific application can be found in [9].

As we mentioned in the introduction, we consider a sensor network of $N$ agents that are allowed to communicate according to the communication graph $\mathcal{G}=(\mathcal{V},\mathcal{E})$.

We want to estimate, by means of such a sensor network, a discrete-time scalar random walk described by equation (3) from noisy measurements (4). For ease of notation collect all the $N$ measurements in a vector $\mathbf{y}(t)=[\mathbf{y}_1(t),\ldots,\mathbf{y}_N(t)]^T$ and all measurement noises in a vector $\mathbf{v}(t)=[\mathbf{v}_1(t),\ldots,\mathbf{v}_N(t)]^T$. Equation (4) becomes then

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

where we denote by $\mathbb{1}$ the vector $[1,\ldots,1]^T\in\mathbb{R}^N$. Since we assumed that all measurement noises are independent and identically distributed, we have that $\mathbb{E}[v(t)v^T(t)]=rI_N$.

The estimation algorithm we consider is presented in detail in [9] and it is a randomized version of the one that has been proposed and analyzed in [6]. It consists on the recursion

$$\hat{\mathbf{x}}(t+1) = \mathbf{Q}(t)\left(\ell\hat{\mathbf{x}}(t) + (1-\ell)\mathbf{y}(t)\right)$$

where $\ell \in (0,1)$ is the Kalman gain and $\mathbf{Q}(t)$ is drown at each time instant from an alphabet of stochastic matrices compatible with the graph, namely $Q_{i,j} \neq 0$ only if $(j,i) \in \mathcal{E}$.

### *6.1 Mean Square Analysis*

Let us consider the estimation error of the algorithm

$$\tilde{\mathbf{x}}(t) = \mathbb{1}\mathbf{x}(t) - \hat{\mathbf{x}}(t).$$

After simple computations one gets the following description of its time evolution

$$\tilde{\mathbf{x}}(t+1) = (1-\ell)\mathbf{Q}(t)\tilde{\mathbf{x}}(t) + \ell\mathbf{Q}(t)\mathbf{w}(t) + \mathbb{1}\mathbf{v}(t).$$

Let us study the variance of the estimation error

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

Recalling that $\mathbf{Q}(t)$, $\mathbf{w}(s)$, $\mathbf{v}(u)$ can be assumed to be independent $\forall t,s,u$ and noting that $\tilde{\mathbf{x}}(t)$ is independent from $\mathbf{Q}(t)$, $\mathbf{w}(t)$ and $\mathbf{v}(t)$, since it depends only on $\mathbf{Q}(s)$, $\mathbf{w}(s)$ and $\mathbf{v}(s)$ for $s = 1,\ldots,t-1$, one can easily see then that

$$\begin{aligned}\Sigma(t+1) &= \mathbb{E}\left[\tilde{\mathbf{x}}(t+1)\tilde{\mathbf{x}}^T(t+1)\right]\\ &= (1-\ell)^2\mathbb{E}\left[\mathbf{Q}(t)\tilde{\mathbf{x}}(t)\tilde{\mathbf{x}}^T(t)\mathbf{Q}^T(t)\right] + \ell^2 r\mathbb{E}\left[\mathbf{Q}(t)\mathbf{Q}^T(t)\right] + q\mathbb{1}\mathbb{1}^T.\end{aligned}$$

Since

$$\begin{aligned}&\mathbb{E}\left[\mathbf{Q}(t)\tilde{\mathbf{x}}(t)\tilde{\mathbf{x}}^T(t)\mathbf{Q}^T(t)\right] =\\ &\quad= \mathbb{E}\left[\mathbb{E}\left[\mathbf{Q}(t)\tilde{\mathbf{x}}(t)\tilde{\mathbf{x}}^T(t)\mathbf{Q}^T(t)|\mathbf{Q}(t)\right]\right]\\ &\quad= \mathbb{E}\Big[\mathbf{Q}(t)\mathbb{E}\left[\tilde{x}(t)\tilde{x}^T(t)|\mathbf{Q}(t)\right]\mathbf{Q}^T(t)\Big]\\ &\quad= \mathbb{E}\left[\mathbf{Q}(t)\Sigma(t)\mathbf{Q}^T(t)\right],\end{aligned}$$

we have that

$$\Sigma(t+1) = (1-\ell)^2\mathbb{E}\left[\mathbf{Q}(t)\Sigma(t)\mathbf{Q}^T(t)\right] + \ell^2 r\mathbb{E}\left[\mathbf{Q}(t)\mathbf{Q}^T(t)\right] + q\mathbb{1}\mathbb{1}^T. \tag{20}$$

For ease of notation let us define the linear operator

$$\mathcal{L}(M) = \mathbb{E}[\mathbf{Q}(t)M\mathbf{Q}^T(t)].$$

Equation (20) represents a linear time invariant system that can be rewritten as

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

Note that, using this notation, the matrix $P(t)$ defined in (1) is given by

$$P(t) = \mathcal{L}^t(I).$$

Note moreover that $\mathcal{L}(\mathbb{1}\mathbb{1}^T) = \mathbb{E}(\mathbf{Q}\mathbb{1}\mathbb{1}^T\mathbf{Q}^T) = \mathbb{1}\mathbb{1}^T$. To see more clearly that equation (20) represents a linear time invariant system, define $s(t)$ the vectorization of $\Sigma(t)$, $s(t) = \mathrm{vect}(\Sigma(t))$. Recalling then that $\mathrm{vect}(ABC) = (C^T \otimes A)\mathrm{vect}(B)$ and note that $\mathrm{vect}\big(\mathcal{L}(M)\big) = \mathbb{E}[\mathbf{Q}(t)\otimes\mathbf{Q}(t)]\mathrm{vect}(M)$. Equation (20) can then be rewritten as

$$\mathrm{vect}(\Sigma(t+1)) = s(t+1) = (1-\ell)^2\mathbb{E}[\mathbf{Q}(t)\otimes\mathbf{Q}(t)]s(t) + \ell^2 r\mathbb{E}[\mathbf{Q}(t)\otimes\mathbf{Q}(t)]\mathrm{vect}(I) + q\mathbb{1}_{N^2},$$

where $\mathbb{1}_{N^2}$ is the vector of $\mathbb{R}^{N^2}$ having all entries 1, that is precisely a linear time–invariant system forced by a constant input. Note that

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

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

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

$$\begin{aligned}
\Sigma(\infty) &= \sum_{i=0}^{+\infty}(1-\ell)^{2i}\mathcal{L}^i\left(\ell^2 r\mathcal{L}(I) + q\mathbb{1}\mathbb{1}^T\right) \\
&= \ell^2 r\sum_{i=0}^{+\infty}(1-\ell)^{2i}\mathcal{L}^{i+1}(I) + q\sum_{i=0}^{+\infty}(1-\ell)^{2i}\mathbb{1}\mathbb{1}^T \\
&= \ell^2 r\sum_{i=0}^{+\infty}(1-\ell)^{2i}\mathcal{L}^{i+1}(I) + \frac{q}{1-(1-\ell)^2}\mathbb{1}\mathbb{1}^T.
\end{aligned} \tag{21}$$

One is in particular interested in computing the trace of $\Sigma(\infty)$.

$$\text{tr}\ (\Sigma(\infty)) = \ell^2 r\sum_{i=0}^{+\infty}(1-\ell)^{2i}\text{tr}\ \mathcal{L}^{i+1}(I) + \frac{qN}{1-(1-\ell)^2} \tag{22}$$

Unluckily we did not manage to find a closed form expression for

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

but it can be upper–bounded using the trace inequality presented in corollary 5. We have in fact that

$$\text{tr}\ \mathcal{L}^i(I) \leq \text{tr}\ \left(\mathbb{E}\left[\mathbf{Q}(t)\mathbf{Q}^T(t)\right]^i\right) = \text{tr}\ \left(\mathbb{E}\left[\mathbf{Q}^2(t)\right]^i\right), \tag{23}$$

where we used the fact that we are restring our analysis to the case of symmetric matrix alphabet.

Equation (23) allows us to analyze the $N \times N$ matrix $\mathbb{E}[\mathbf{Q}^2(t)]$ rather then the linear operator $\mathcal{L}$, described by the $N^2 \times N^2$ matrix $\mathbb{E}[\mathbf{Q}(t) \otimes \mathbf{Q}(t)]$.

Using (23) we get the following upper–bound on tr $\Sigma(\infty)$

$$\begin{aligned}
\text{tr}\ \Sigma(\infty) &= \ell^2 r\sum_{i=0}^{+\infty}(1-\ell)^{2i}\text{tr}\ \mathcal{L}^{i+1}(I) + \frac{qN}{1-(1-\ell)^2} \\
&\leq \ell^2 r\sum_{i=0}^{+\infty}(1-\ell)^{2i}\text{tr}\ \mathbb{E}\left[\mathbf{Q}^2(t)\right]^{i+1} + \frac{qN}{1-(1-\ell)^2} \\
&= \ell^2 r\sum_{i=0}^{+\infty}(1-\ell)^{2i}\sum_{j=1}^{N}\lambda_j\left(\mathbb{E}\left[\mathbf{Q}^2(t)\right]^{i+1}\right) + \frac{qN}{1-(1-\ell)^2} \\
&= \sum_{j=1}^{N}\ell^2 r\sum_{i=1}^{+\infty}(1-\ell)^{2i-2}\lambda_j^i\left(\mathbb{E}\left[\mathbf{Q}^2(t)\right]\right) + \frac{qN}{1-(1-\ell)^2} \\
&= \sum_{j=1}^{N}\ell^2 r\frac{\lambda_j\left(\mathbb{E}\left[\mathbf{Q}^2(t)\right]\right)}{1-(1-\ell)^2\lambda_j\left(\mathbb{E}\left[\mathbf{Q}^2(t)\right]\right)} + \frac{qN}{1-(1-\ell)^2}.
\end{aligned}$$

On an application point of view, one would like to study the natural optimization problem of finding the probability distribution $p_\alpha = \mathbb{P}[\mathbf{Q}(t) = Q_\alpha]$ such that $\frac{1}{N}\mathrm{tr}\ \Sigma(\infty)$ is minimized. Numerical experiments show that this optimization problem is not convex. One can rather consider then as a cost function the proposed upper–bound on $\mathrm{tr}\ \Sigma(\infty)$

$$\frac{1}{N}\mathrm{tr}\ \Sigma(\infty) \leq J(\{p_\alpha\}_{\alpha\in\mathbb{A}}) = \frac{1}{N}\sum_{j=1}^{N} \ell^2 r \frac{\lambda_j\left(\mathbb{E}\left[\mathbf{Q}^2(t)\right]\right)}{1-(1-\ell)^2\lambda_j\left(\mathbb{E}\left[\mathbf{Q}^2(t)\right]\right)} + \frac{q}{1-(1-\ell)^2}$$

Therefore it is relevant to study the optimization problem of minimizing the cost functional $J(\{p_\alpha\}_{\alpha\in\mathbb{A}})$ in the variables $\{p_\alpha\}_{\alpha\in\mathbb{A}}$. As it has be shown in [9] such a problem is convex. This fact has a relevant impact since it shows that this optimization problem is intrinsically easy and can be solved using any standard convex optimization technique. Even on a numerical point of view, one can make use of one of the many convex optimization software available.

## 7 Conclusions

In this paper we presented a majorization inequality on the singular values of the expectation of matrix–valued random variables' product. It is shown that, if the alphabet is made of normal matrices, the singular values of the matrix $\mathbb{E}[\mathbf{Q}(i)\ldots\mathbf{Q}(0)\mathbf{Q}^T(0)\ldots\mathbf{Q}^T(i)]$ are submajorized by the singular values of $\mathbb{E}[\mathbf{Q}(0)\mathbf{Q}(0)^T]^i$. As a straightforward corollary of this result, we proposed a novel trace inequality.

Furthermore, we illustrated a relevant application of this majorization inequality to the field of distributed estimation over sensor networks. More precisely, we analyzed a randomized version of the distributed Kalman filter proposed in [6]. A mean–square analysis was carried out and an upper–bound on the trace of the error variance was proposed using the previously presented trace inequality. It was shown moreover that the problem of minimizing this upper–bound is convex, while the original problem is not.

More generally, the proposed result applies to the analysis of jump–Markov linear systems where the transition matrix is an i.i.d. random process. Another relevant example of application in this framework is the case of a linear plant controlled by a linear time–varying feedback, in which at each time instant a feedback matrix $K(t)$ is drawn from a fixed set of matrices $K_1, \ldots, K_N$ and in which the designer only chooses the probability that the various feedback matrices are applied.

Many numerical experiments support the conjecture that the statement of Theorem 4 holds true also in the case of stochastic, possibly non-normal, matrices. We expect an eventual proof of such a conjecture to leverage on arguments different from the ones we used here. In fact, normality is a key assumption in our proof and we do not see how to take advantage of stochasticity in our arguments.

## References


[1] Sensor networks and applications. *Proceedings of the IEEE, Special issue*, 91, August 2003.

[2] Peter Alriksson and Anders 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.

[3] Peter Alriksson and Anders Rantzer. Model based information fusion in sensor networks. In *Proceedings of the 17th IFAC World Congress*, Seoul, Korea, July 2008.

[4] Rajendra Bathia. *Matrix Analysis*. Springer, 1997.

[5] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah. Randomized gossip algorithms. *IEEE Transactions on Information Theory/ACM Transactions on Networking*, 52(6):2508–2530, June 2006.

[6] 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.

[7] R. Carli, F. Fagnani, A. Speranzon, and S. Zampieri. Communication constraints in the average consensus problem. *Automatica*, 44(3):671–684, March 2008.

[8] O.L.V. Costa, M.D. Fragoso, and R.P. Marques. *Discrete-Time Markov Jump Linear Systems (Probability and its Applications)*. Springer, 2005.

[9] S. Del Favero and S. Zampieri. Distributed estimation through randomized gossip Kalman filter. In *Proc. 48th IEEE Conference on Decision and Control and 28th Chinese Control Conference. Shanghai*, pages 7049 – 7054, Dec 2009.

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

[11] J.P. Hespanha, P. Naghshtabrizi, and Y. Xu. A survey of recent results in networked control systems. *Proceedings of the IEEE, special issue*, 95(1):138–162, January 2007.

[12] Roger A. Horn and Charles R. Johnson. *Matrix Analysis.* Cambridge University Press, Cambridge, UK, 1990.

[13] Roger A. Horn and Charles R. Johnson. *Topics in Matrix Analysis.* Cambridge University Press, Cambridge, UK, 1991.

[14] Albert W Marshall and Ingram Olkin. *Inequalities: Theory of Majorization and Its Applications*, volume 143 of *Mathematics in science and engineering.* Academic press, New York, 1979.

[15] 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.

[16] 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.

[17] Special Issue Proceedings of the IEEE, editor. *Sensor networks and applications*, volume 91, August 2003.

[18] A. Speranzon, C. Fischione, and K.H. 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.

[19] L. Xiao and S. Boyd. Fast linear iterations for distributed averaging. *Systems and Control Letters*, 53(1):65–78, September 2004.

[20] 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.