JEGA: a joint estimation and gossip averaging algorithm for sensor network applications

Nicolas Maréchal, Jean-Benoit Pierrot, Jean-Marie Gorce

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

JEGA: a joint estimation and gossip averaging algorithm for sensor network applications

Nicolas Maréchal — Jean-Benoît Pierrot — Jean-Marie Gorce

N° 6597

25 juillet 2008

Thème COM

Rapport de recherche

ISSN 0249-6399 ISRN INRIA/RR--6597--FR+ENG

JEGA: a joint estimation and gossip averaging algorithm for sensor network applications

Nicolas Maréchal* , Jean-Benoît Pierrot† , Jean-Marie Gorce‡

Thème COM — Systèmes communicants
Projet ARES

Rapport de recherche n° 6597 — 25 juillet 2008 — 25 pages

Abstract: Distributed consensus algorithms are widely used in the area of sensor networks. Usually, they are designed to be extremely lightweight at the price of computation time. They rely on simple local interaction rules between neighbor nodes and are often used to perform the computation of spatial statistical parameters (average, variance, regression). In this paper, we consider the case of a parameter estimation from input data streams at each node. An average consensus algorithm is used to perform a spatial regularization of the parameter estimations. A two step procedure could be used: each node first estimates its own parameter, and then the network applies a spatial regularization step. It is however much more powerful to design a joint estimation/regularization process. Previous work has been done for solving this problem but under very restrictive hypotheses in terms of communication synchronicity, estimator choice and sampling rates. In this paper, we study a modified gossip averaging algorithm which fulfills the sensor networks requirements: simplicity, low memory/CPU usage and asynchronicity. By the same way, we prove that the intuitive idea of mass conservation principle for gossip averaging is stable and asympotically verified under feedback corrections even in presence of heavily corrupted and correlated measures.

Key-words: distributed algorithms, gossip algorithms, epidemic algorithms, averaging, average consensus, estimation, space-time diffusion, sensor networks

* CEA-LETI MINATEC, Grenoble, F-38054 France (email: nicolas.marechal@cea.fr)
† CEA-LETI MINATEC, Grenoble, F-38054, France (email: jean-benoit.pierrot@cea.fr)
‡ Université de Lyon, INRIA, INSA-Lyon, CITI, F-69621, France (email: jean-marie.gorce@insa-lyon.fr)

Unité de recherche INRIA Rhône-Alpes
655, avenue de l'Europe, 38334 Montbonnot Saint Ismier (France)
Téléphone : +33 4 76 61 52 00 — Télécopie +33 4 76 61 52 52

JEGA: un algorithme d'estimation et moyennage conjoints pour application aux réseaux de capteurs

Résumé : Les algorithmes distribués de consensus de moyenne (*average consensus*) sont courramment utilisés dans le domaine des réseaux de capteurs. Conçus pour être extrêmement légers au prix d'un temps de convergence accru, ils reposent sur des interactions locales entre noeuds voisins et sont principalement utilisés pour le calcul de paramètres statistiques empiriques (moyenne, variance, ...). Dans cet article, nous nous plaçons dans le cadre de l'estimation sur chaque noeud-capteur d'un paramètre à partir d'un flot d'échantillons. Un algorithme de consensus de moyenne est utilisé afin de régulariser spatiallement les paramètres estimés. Il serait possible d'utiliser une procédure en deux étapes: chaque noeud calcule tout d'abord une première estimation de son paramètre, puis le réseau uniformise ces estimations dans un deuxième temps. Il est cependant plus intéressant d'utiliser un schéma d'estimation/régularisation conjointes. Dans de précédents travaux, une solution a été proposée pour résoudre ce problème mais sous des hypothèses trop restrictives en termes de synchronie des communications, de choix d'estimateur et de cadence d'échantillonage. Dans ce rapport, nous étudions une version modifiée d'un algorithme de moyennage pair-à-pair qui répond à la problématique précédemment citée tout en respectant les spécificités des réseaux de capteurs: simplicité, faible usage des ressources (processeur, mémoire, ...) et asynchronie.

Mots-clés : algorithmes distribués, algorithmes gossip, algorithmes épidémiques, consensus de moyenne, estimation, diffusion spatio-temporelle, réseaux de capteurs

1 Introduction

Sensor networks consist of a great amount of small entities, called nodes, equipped with low cost hardware in order to balance the total network cost. They are commonly used for monitoring physical phenomena on wide areas such as hydrometry, landslides or fires, but also for tracking purposes and in military warfare ([1],[2],[3],[4],[5]). The direct drawbacks of low cost hardware are numerous: severe energy constraints (battery lifetime), poor CPU and storage abilities, low transmission rates and small communication range. Facing these limitations and objectives, wireless sensor networks (WSN) have to self-organize their exchanges and the maneer sensor nodes must achieve their mission. As in most cases, computations in a centralized fashion become untractable and/or inadaptated, robust distributed algorithms have to be designed. The biggest part of algorithm design for WSN is dedicated to improving the performances while preserving energy consumption. For example, several data fusion schemes developped in order to provide a good and compact representation of the observed phenomenon can be made on the basis of a high number of low quality measures [6] and simple local interactions between neighbor nodes (gossiping). The particular class of distributed consensus algorithms is of great interest: they provide a robust way of homogenizing parameters among network nodes [7]. More specifically, average consensus algorithms seem to be a good choice whenever the stability and the quality of the consensus point is a critical issue [8], and extend to a wide panel of data fusion tools such as estimators for statistical moments, linear regression and polynomial map fitting ([6], [9]). However, a problem occurs when data to be averaged are subject to fluctations. This is often the case when a statistical parameter is estimated from time series of noisy data samples. As the number of available samples increases, the estimation process naturally acts as a temporal regularization scheme and fluctuations are reduced: one would ask to adjust the current state of average consensus in order to account for informations with higher precision. For an additive zero-mean stationary measurement noise process, the quality of this estimation increases with the number of measures and, as a corollary, with time. Nevertheless, gossip averaging algorithms are very slow to converge in comparison with centralized algorithms: gossip-based consensus algorithms converge asymptotically, rarely in finite time. As sensor networks suffer from heavy constraints on their resources, time and energy must be saved, it thus becomes necessary to run the gossip averaging algorithm while the estimation is still in progress by using correction mechanisms. This double process should be understood as a spatio-temporal regularization scheme: each node performs individually a local regularization of extracted features from sampled data (estimation) , while a spatial regularization (averaging) is performed in order to extract a global characteristic. Previous work has been done on this topic, and an alternative version of the algorithm introduced in [9] is described in this article. As explained in this paper, the originality of our work consists mainly in the full asynchronicity that is assumed for both data exchanges and estimation processes, and in the wide range of estimators covered by the hypotheses. Moreover, our algorithm can find many application contexts: as an example, a clock synchronization scheme for wireless sensor makes use of it [10]. This paper is organized as follows: section 2 provides a short overview of gossip-based consensus algorithms , their principles and some known results. In section 3, a solution is proposed, addressing the problem of the parallel estimation and averaging while respecting the philosophy and paradigm of sensor networks. Further the convergence is proved. After these theoretical considerations, simulation results are provided in section 4 in order to give a qualitative study of its behaviour w.r.t. parameters scale. In addition to its apparant meaning, this work proves that principles of mass conservation are respected and stable through our algorithm, even when measurements are highly corrupted and correlated.

2 Distributed consensus algorithms

2.1 Principles of gossip-based consensus algorithms

Distributed consensus algorithms/protocols aim at agreeing all network nodes with a common value or a decision in a decentralized fashion. From a signal processing context, this can be understood as a spatial regularization process. When data exchanges consist of local, asynchronous and simple interactions between neighbor nodes, such algorithms refered to as gossip-based. The particular subclass of gossip-based average consensus algorithms does not limit to the computation of averages, but extends to the extraction of a large variety of aggregated and statistical quantities like sums/products, max/min values, variances, quantiles and rank statistics ([11] and [12]). More direct applications like least-squares regression of model parameters have also been adapted to this algorithms ([6],[9]). All these specificities make gossip-based consensus algorithms good candidates for sensor networks applications, where bandwidth, energy consumption and CPU/memory usage are enduring severe limitations for the sake of nodes lifetime and size. Despite their suboptimality[1], pure gossip consensus algorithm can be used as a prelude to more sophisticated algorithms by homogenizing parameters upon groups of nodes (for example, the reduction of carrier offset for reducing time drift of TDMA schemes). The performance analysis of such algorithms relies essentially on diffusion speed statistics and is then closely related to performances of flooding/multicasting processes and mixing time of Markov chains: some asymptotic bounds on convergence time are given in [12] and [11].

2.2 Gossip averaging

In practical applications, consensus algorithms are often used in order to easily homogenize some parameters. However, one should distinguish situations in which the agreed value is critical. For example, agreeing on a meeting point is not as critical as detecting the position of a sniper. Obtaining an average is in general much slower than any uniformization algorithm based on flooding/broadcasting techniques, but ensures a good qualitity of consensus.

Gossip-based average consensus algorithms (gossip averaging) have been widely studied in literature ([13],[14],[11],...). As suggested by their denomination, they aim at computing a global average of local values. This paper focus on a version based on an asynchronous peer-to-peer communications model as presented in [6]. Such a model frequently occurs in sensor networks applications where full synchronicity is neither guaranteed nor easily tractable. Under

[1] in comparison with any finite-time converging algorithm having the same objective

this assumption, an interaction consists in choosing a pair of neighbor nodes at each iteration and making a local averaging between them. A gossip averaging algorithm is then described by a linear difference equation of the form:

$$X_{k+1} = W_k X_k \tag{1}$$

where X_k is the vector of nodes' values at iteration k (X_0 contains the initial values to be averaged) and W_k is a n-by-n matrix which describes instantaneous pairwise interactions. In fact, X_k can be seen as a state vector, and its components as estimates of the global average of intial values. Following Boyd's notations, W_k could be written:

$$W_k = I - \frac{(e_i - e_j)(e_i - e_j)^\top}{2} \triangleq W_{ij} \tag{2}$$

where $e_i = [0 \ldots 0\,1\,0 \ldots 0]^\top$ is a n-dimensional vector with i^{th} entry equal to 1 (here, double index ij means that node i contacts node j). This clearly corresponds to the following rule (time into brackets, node's ID as index):

$$x_i(k+1) = x_j(k+1) = \frac{x_i(k) + x_j(k)}{2} \tag{3}$$

For convenience, the W_k are considered to be i.i.d. random matrices. Each W_{ij} is chosen with probability p_{ij}, i.e. according to the probability that node i initiates an iteration involving node j.

$$W = \mathbb{E}[W_k] = \mathbb{E}[W_0] = \frac{1}{n} \sum_{i,j} p_{ij} W_{ij} \tag{4}$$

where n stands for the number of network nodes.

In [14], the following statement is proven:

Proposition 1 ([14])**.** *If W is a doubly-stochastic ergodic matrix, then*

$$X_k \xrightarrow[k\to\infty]{} \bar{\theta}\mathbf{1}$$

where $\bar{\theta} = \frac{1}{n}\sum_{i=1}^{n} x_i(0)$, *and* $\mathbf{1} = [1, \ldots 1]^\top \in \mathbb{R}^n$.

This means that convergence to the true average value is ensured if and only if the largest eigenvalue (in modulus) of W is equal to 1, with multiplicity 1. In other words, W is the transition matrix of a Markov chain whose underlying (weighted) transition graph $\mathcal{G}$ is strongly connected, i.e. for each pair (i, j) of vertices of $\mathcal{G}$, a path from i to j and a path from j to i are existing. In [15], the authors proved that proposition 1 is equivalent to the following proposition.

Proposition 2 ([15])**.** *Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be the graph such that:*

- *the set of vertices $\mathcal{V}$ is the set of network nodes.*
- *there is an edge between two vertices only if the corresponding nodes are interacting infinitely many times.*

Then, X_k converges to $\bar{\theta}$ if $\mathcal{G}$ is connected.

Another way to define $\mathcal{E}$ with respect to the set $\mathcal{E}_k$ of active links (edges) at time instant k is the following:

$$\mathcal{E} = \bigcap_{n=0}^{\infty} \bigcup_{i=n}^{\infty} \mathcal{E}_k = \limsup_n \mathcal{E}_n$$

2.3 Performance analysis of noiseless gossip averaging

The convergence speed of the standard gossip averaging algorithm is strongly related to the second largest eigenvalue (in modulus), λ_2 of W (see [15]). Given link probabilities, performances are quantified in terms of ϵ-averaging time, i.e. the minimal time that guarantees a relative error of order ϵ with probability at least $1-\epsilon$:

$$T_{ave}(\epsilon) = \sup_{X_0 \in \mathbb{R}^n} \inf \left\{ k : \Pr\left[\frac{\left\|X_k - \bar{\theta}\mathbf{1}\right\|^2}{\left\|X_0\right\|^2} \geq \epsilon \right] \leq \epsilon \right\} \tag{5}$$

Boyd and al. derived upper and lower bounds for $T_{ave}(\epsilon)$ based on the second largest eigenvalue of the corresponding matrix W:

$$\frac{0.5 \log \epsilon^{-1}}{\log \lambda_2(W)^{-1}} \leq T_{ave}(\epsilon) \leq \frac{3 \log \epsilon^{-1}}{\log \lambda_2(W)^{-1}} \tag{6}$$

The eigenvalue λ_2 is function of the link probabilities between pair of neighbor nodes. Thus, one can try to reduce λ_2 by acting on neighbor links, in two ways:

- link creation/deletion according to topology-based heuristics.
- neighbor selection uniformly and randomly chosen, or not: potential heuristics.

Some papers described solutions for the (distributed) optimization of λ_2 for a given network using semidefinite programming [16]. The choice of algorithms with such complexity is questionable for sensor network applications: their operation should remain reasonably feasible on the chosen architecture, and performance gains must be important enough to compensate for time and energy lost during optimization. However, this scheme is of great interest in a topology-stable network.

2.4 Gossip averaging of noisy measures

In some applications, the parameters to be averaged are estimated from sampled streams of data. Recently, a tremendeous work has been published to enhance performances and robustness of consensus algorithms ([17],[18]) and to define protocols for thems. However, one of the most important practical problem is the presence of additive measurement noise. When coupling noises corrupt exchanges, Boyd et al proved that the consensus point deviates from the true average value, and takes the form of a random walk over $\mathbb{R}^n$: the mean squared norm of local deviations increase linearly with time. In [19], a solution is proposed to reduce the rate of deviation, while in [9] the effort is done to provide simultaneous averaging and estimation. Nevertheless, the work done in [9] relies on exchanges synchronicity and only focus on the case of least mean square estimation. The problem is then to find a solution to desynchronizing exchanges while preserving the convergence in case of more general estimates. The main difficulty in this study remains the convergence rate of the estimation process. For example, the variance of the optimal estimator for the average of normally distributed values converges inversely proportionaly to the number of values. In [9], the proof of second-order convergence is based on the condition that the variance σ_t has a finite norm $l^2(\mathbb{N})$. In general, the variance of an estimator does not fulfill this requirement.

3 Joint estimation and gossip averaging

Similarly to the initial algorithm presented in [9], we propose to run an asynchronous spatial regularization (averaging) process working jointly with the local estimation of the parameters (temporal regularization). The main difference stands in the full asynchronoicity of our scheme, the chosen weights for interactions, and the retroaction process. Moreover, the convergence of our algorithm is prooved in the following for a wide range of local estimators under the single assumption that their spatio-temporal covariances decrease to 0 with time, without any assumption neither on:

- their convergence rate w.r.t. to the number of samples.
- data sampling rates.

3.1 Measurement process and estimation

During the measurement process, nodes collect samples related to some data of interest. The goal of the estimation phase is to extract some unkown parameters or characteristics of the original data distribution only from samples. In particular, an estimator of some parameter θ is said to be unbiased, if at any time, the mean of the estimator is θ. In this work, a proper estimation process is considered, i.e. converging to the expected parameter as the number of samples grows (the variance tends toward 0 with time): this estimation is refered to as temporal regularization. As an example, one should be interested in estimating the mean μ of some real-valued distribution $\mathcal{D}$ (of finite variance σ^2). It is well known that, given i.i.d. samples v_k from the distribution, the sample average converges to the true mean μ, i.e.:

$$
\begin{aligned}
&\forall k \in \mathbb{N}^*, v_k \sim \mathcal{D}(\mu, \sigma) \\
&\qquad \Rightarrow \left\{ \begin{array}{l} \mathbb{E}\left[\frac{1}{n}\sum_{k=1}^{n} v_k\right] = \mu \\ \mathbb{V}\left[\frac{1}{n}\sum_{k=1}^{n} v_k\right] = \frac{\sigma^2}{n} \xrightarrow[n\to\infty]{} 0 \\ \mathrm{Cov}\left(\frac{1}{m}\sum_{k=1}^{m} v_k, \frac{1}{n}\sum_{l=1}^{n} v_l\right) = \frac{\sigma^2}{\max(m,n)} \xrightarrow[\max(m,n)\to\infty]{} 0 \end{array} \right.
\end{aligned}
\tag{7}
$$

In real experiments, samples may be correlated and their distribution may change through time. However, some consistancies in the measurement and estimation processes are assumed: measurements can be taken from a time-varying distribution but the parameters to estimate must remain constant through time. For instance, interferences in wireless communications are subject to the network activity dynamics, and are usually modelled as a centered random noise. In other words, their variances (power) vary with time, but their means are constant and equal to 0.

3.2 Description of the algorithm

Let us start with some notations and conventions. For any node i, the current estimation of the parameter θ_i given samples available at node i up to time k is denoted by $\hat{\theta}_i^k$. These estimators are grouped in the n-dimensional vector Z_k:

$$Z_k = \left[\hat{\theta}_1^{(k)}, \ldots, \hat{\theta}_n^{(k)}\right]^\mathsf{T} \tag{8}$$

As local estimators are assumed unbiased, the vector $\mathbb{E}\left[Z_k\right]$ is constant through time and is equal to the vector of parameters to be estimated.

$$\bar{Z} \triangleq \mathbb{E}\left[Z_k\right] = \left[\theta_1, \ldots, \theta_n\right]^\mathsf{T} \tag{9}$$

In this paper, it is useful to consider the difference between Z_k and its mean, which is denoted by B_k:

$$B_k \triangleq Z_k - \bar{Z} = \left[b_1^{(k)}, \ldots, b_n^{(k)}\right] \tag{10}$$

We also define the covariance term C_{ij}^{kl} which measures the relation between components of Z_k through time:

$$C_{ij}^{kl} \triangleq \mathbb{E}\left[\left(\widehat{\theta_i}^{(k)} - \theta_i\right)\left(\widehat{\theta_j}^{(l)} - \theta_j\right)\right] = \mathbb{E}\left[b_i^{(k)} b_j^{(l)}\right] \tag{11}$$

In the following of this article, the proof of the convergence will rely on the only assumption that these components are asympotically uncorrelated, i.e.

$$\forall (i,j) \in [1,n]^2,\ C_{ij}^{kl} \xrightarrow[(k+l)\to\infty]{} 0 \tag{12}$$

For most of sensor network applications, this assumption is quite not restrictive as sufficiently spaced (temporally and/or spatially) samples tends to be uncorrelated too. In many cases, the Stolz-Cesaró theorem and its extensions help in finding sufficient conditions on sample convariances for ensuring assumption (12).

The proposed algorithm is based on the standard gossip averaging algorithm (1), upon which a simple feedback is added to account for estimated parameters updates. This modified algorithm takes the form of a non homogeneous system of first-order linear difference equations, and is defined by:

$$\left\{ \begin{array}{c} X_{k+1} = W_k X_k + Z_{k+1} - Z_k \\ Z_0 = X_0 = [0 \ldots 0]^\mathsf{T} \end{array} \right. \tag{13}$$

The proof of the efficiency of this algorithm relies on the convergence of every component of the state vector X_k to the spatial average of the estimated parameters $(\theta_i)_{i=1..n}$ while those of Z_k should converge to the local parameters:

$$X_k \xrightarrow[k\to\infty]{} \left(\frac{1}{n}\sum_{i=1}^{n} \theta_i\right)\mathbf{1}$$

$$Z_k \xrightarrow[k\to\infty]{} \bar{Z} = \left[\theta_1, \ldots, \theta_n\right]^\mathsf{T}$$

In system (13), the (random) matrix W_k fulfills the same conditions as in the classic gossip avaraging algorithms: it can be constant through time or taken at random, but its mean must be a doubly stochastic[2] ergodic[3] matrix. The behaviour of our system is very easy to describe qualitatively. The paracontracting matrices (see [20]) W_k homogenize the values of X_k while the components of Z_k are stabilizing. If each $z_i(k)$ is ergodic, $Z_{k+1} - Z_k$ tends to 0 as X_k tends to a vector colinear to $[1 \dots 1]$. The term $Z_{k+1} - Z_k$ implies a permanent correction of the total weights of X_k. One can then conclude that after an infinite time, the sum of the (identical) components of X_k is equal to the sum of those of $\bar{Z}$. In the next two parts, a formal proof of this analysis is given.

3.3 First-order moment convergence

The first result asserts that the mean limit of X_k is naturally the average of estimated parameters, i.e. $\bar{\theta} = \frac{1}{n}\sum_{i=1}^{n}\theta_i$:

Proposition 3.

$$\lim_{k\to\infty} \mathbb{E}\left[X_k\right] = \bar{\theta}\mathbf{1} \tag{14}$$

Proof. All along the proof, we take benefit of the recursive form efficiency of update equation (13):

$$\begin{aligned}
\mathbb{E}\left[X_{k+1}\right] &= \mathbb{E}\left[W_k X_k + Z_{k+1} - Z_k\right] &(15)\\
&= \mathbb{E}\left[W_k\right]\mathbb{E}\left[X_k\right] + \mathbb{E}\left[Z_{k+1}\right] - \mathbb{E}\left[Z_k\right] &(16)\\
&= W\mathbb{E}\left[X_k\right] + \bar{Z} - \bar{Z} &(17)
\end{aligned}$$

This relation is true except for $k = 0$: $\mathbb{E}\left[X_1\right] = \mathbb{E}\left[Z_1\right] = \bar{Z}$. It follows:

$$\begin{aligned}
\mathbb{E}\left[X_{k+1}\right] &= W^k\mathbb{E}\left[Z_1\right] &(18)\\
&= W^k\bar{Z} &(19)
\end{aligned}$$

The ergodicity of W states that W^k converges to $\frac{\mathbf{1}\mathbf{1}^{\intercal}}{n}$ as k grows. Together with relation (19), it yelds to the expected result:

$$\mathbb{E}\left[X_k\right] \xrightarrow[k\to\infty]{} \frac{\mathbf{1}\mathbf{1}^{\intercal}}{n}\bar{Z} = \bar{\theta}\mathbf{1} \tag{20}$$

□

3.4 Second-order moment convergence

Now that our algorithm is proved to converge on average towards the constant vector $\bar{\theta}\mathbf{1}$, the quality of this convergence must be analyzed. In other words, can the mean distance between consensus and true average (statistically) be made arbitrarily small ? A positive answer is demonstrated below. Technical issues stand in the random nature of matrices appearing in the algorithm, and also in the simple fact that covariances are not assumed to have any finite ℓ^p norm (in any direction). More interestingly, the second order convergence is usefull to state the convergence in probability as explained in the next section.

[2] its rows and columns sum to 1

[3] it is the transition matrix of an ergodic (aperiodic and irreducible) Markov chain

Proposition 4.

$$\lim_{k\to\infty} \mathbb{E}\left[\left\|X_k - \bar{\theta}\mathbf{1}\right\|^2\right] = 0 \tag{21}$$

Proof. The proof of proposition 4 is more fastidious than technical. As the next equation shows, the deviation is directly related to the second-order moment of X_k:

$$\mathbb{E}\left[\left\|X_k - \bar{\theta}\mathbf{1}\right\|^2\right] = \mathbb{E}\left[\left(X_k - \bar{\theta}\mathbf{1}\right)^\mathsf{T}\left(X_k - \bar{\theta}\mathbf{1}\right)\right] \tag{22}$$

The right-hand side of equation (22) can be developped as:

$$\mathbb{E}\left[X_k^\mathsf{T} X_k\right] - 2\bar{\theta}\mathbb{E}\left[\mathbf{1}^\mathsf{T} X_k\right] + \mathbb{E}\left[\left\|\bar{\theta}\mathbf{1}\right\|^2\right] \tag{23}$$

which gives trivially:

$$\mathbb{E}\left[\left\|X_k - \bar{\theta}\mathbf{1}\right\|^2\right] = \mathbb{E}\left[\left\|X_k\right\|^2\right] - n\bar{\theta}^2 \tag{24}$$

Thus, the goal is now to prove that $\mathbb{E}\left[\left\|X_k\right\|^2\right]$ converges to $n\bar{\theta}^2$ as k increases. We rewrite the recursive system (13) into a more efficient way. For this, we define $\Psi(k,i)$ and $\Phi(k,i)$ in $\mathfrak{M}_n(\mathbb{R})$[4] by:

$$\Psi(k,i) \triangleq \begin{cases} I & \text{if } i \geq k \\ W_{k-1}W_{k-2}\dots W_{i+1}W_i & \text{otherwise} \end{cases} \tag{25}$$

$$\Phi(k,i) \triangleq \Psi(k,i) - \Psi(k,i+1) \tag{26}$$

By putting $\Phi(k,i)$ and $\Psi(k,i)$ into system (13), one obtains a more efficient formulation for X_k:

$$\begin{cases} X_k = Z_k + \sum_{i=1}^{k-1} \Phi(k,i) Z_i \\ Z_0 = X_0 = [0\dots 0]^\mathsf{T} \end{cases} \tag{27}$$

This notation helps proving convergence of the second order moment of X_k where classical upper bounding (see [14]) would fail. As one should expect from expression (27), $\mathbb{E}\left[\left\|X_k\right\|^2\right]$ could be written with second-order moments of measures:

$$\mathbb{E}\left[\left\|X_k\right\|^2\right] = \mathbb{E}\left[\left\|Z_k\right\|^2\right] + 2\mathbb{E}\left[Z_k^\mathsf{T}\sum_{i=1}^{k-1}\Phi(k,i)Z_i\right] + \mathbb{E}\left[\left\|\sum_{i=1}^{k-1}\Phi(k,i)Z_i\right\|^2\right] \tag{28}$$

The limit of the first term is easily derived as:

Proposition 5.

$$\lim_{k\to\infty} \mathbb{E}\left[\left\|Z_k\right\|^2\right] = \left\|\bar{Z}\right\|^2 \tag{29}$$

[4] $\mathfrak{M}_n(\mathbb{R})$ denotes the set of real $n \times n$ matrices

Proof.

$$\mathbb{E}\left[\|Z_k\|^2\right] = \sum_{p=1}^{n} \mathbb{E}\left[\left(\theta_p^{(k)}\right)^2\right] \tag{30}$$

$$\xrightarrow[k\to\infty]{} \sum_{p=1}^{n} \mathbb{E}\left[\theta_p\right]^2 = \left\|\bar{Z}\right\|^2 \tag{31}$$

□

In order to make the proof clearer to the reader, the two last terms of equation (28) can be decomposed by separating noisy and deterministic components:

$$\mathbb{E}\left[Z_k^{\mathsf{T}} \sum_{i=1}^{k-1} \Phi(k,i) Z_i\right] = \mathbb{E}\left[\bar{Z}^{\mathsf{T}} \sum_{i=1}^{k-1} \Phi(k,i)\bar{Z}\right] + \mathbb{E}\left[B_k^{\mathsf{T}} \sum_{i=1}^{k-1} \Phi(k,i) B_i\right] + \mathbb{E}\left[B_k^{\mathsf{T}} \sum_{i=1}^{k-1} \Phi(k,i)\bar{Z}\right] + \mathbb{E}\left[\bar{Z}^{\mathsf{T}} \sum_{i=1}^{k-1} \Phi(k,i) B_i\right] \tag{32}$$

$$\mathbb{E}\left[\left\|\sum_{i=1}^{k-1} \Phi(k,i) Z_i\right\|^2\right] = \mathbb{E}\left[\left\|\sum_{i=1}^{k-1} \Phi(k,i)\bar{Z}\right\|^2\right] + \mathbb{E}\left[\left\|\sum_{i=1}^{k-1} \Phi(k,i) B_i\right\|^2\right] + 2\mathbb{E}\left[\left(\sum_{i=1}^{k-1} \Phi(k,i)\bar{Z}\right)^{\mathsf{T}} \left(\sum_{i=1}^{k-1} \Phi(k,i) B_i\right)\right] \tag{33}$$

As noise vectors B_i are centered, cross terms of equations (32) and (33) vanishes through linear combinations:

$$\mathbb{E}\left[Z_k^{\mathsf{T}} \sum_{i=1}^{k-1} \Phi(k,i) Z_i\right] = \mathbb{E}\left[\bar{Z}^{\mathsf{T}} \sum_{i=1}^{k-1} \Phi(k,i)\bar{Z}\right] + \mathbb{E}\left[B_k^{\mathsf{T}} \sum_{i=1}^{k-1} \Phi(k,i) B_i\right] \tag{34}$$

$$\mathbb{E}\left[\left\|\sum_{i=1}^{k-1} \Phi(k,i) Z_i\right\|^2\right] = \mathbb{E}\left[\left\|\sum_{i=1}^{k-1} \Phi(k,i)\bar{Z}\right\|^2\right] + \mathbb{E}\left[\left\|\sum_{i=1}^{k-1} \Phi(k,i) B_i\right\|^2\right] \tag{35}$$

The limit of the four terms remaining in (34) and (35) are obtained separately:

Proposition 6.

$$\lim_{k\to\infty} \mathbb{E}\left[\bar{Z}^{\mathsf{T}} \sum_{i=1}^{k-1} \Phi(k,i)\bar{Z}\right] = n\bar{\theta}^2 - \left\|\bar{Z}\right\|^2 \tag{36}$$

Proposition 7.

$$\lim_{k\to\infty} \mathbb{E}\left[\left\|\sum_{i=1}^{k-1} \Phi(k,i)\bar{Z}\right\|^2\right] = \left\|\bar{Z}\right\|^2 - n\bar{\theta}^2 \tag{37}$$

Proposition 8.

$$\mathbb{E}\left[B_k^\mathsf{T}\sum_{i=1}^{k-1}\Phi(k,i)B_i\right]\xrightarrow[k\to+\infty]{}0$$

Proposition 9.

$$\lim_{i\to+\infty}\mathbb{E}\left[\left\|\sum_{i=1}^{k-1}\Phi(k,i)B_i\right\|^2\right]=0$$

The proof of these four propositions is the fastidious task we announced for proving proposition (4), and is provided in appendix. Propositions 6 and 7 are a reformulation of classical results on gossip averaging algorithms ([16]) adaptated to the terminology of this article, and do not contains major difficulties. On the contrary, the proof of propositions 8 and 9 is more technical and relies on the fact that the perturbations due to the measurement/estimation noise vanish if the assumption made on covariances in (12) is valid.

Now, thanks to propositions 5 to 9 the limit of $\mathbb{E}\left[\|X_k\|^2\right]$ when k goes to $+\infty$ can be derived:

$$\mathbb{E}\left[X_k^\mathsf{T}X_k\right]\xrightarrow[k\to\infty]{}\left\|\bar{Z}\right\|+2(n\bar{\theta}^2-\left\|\bar{Z}\right\|)+(\left\|\bar{Z}\right\|-n\bar{\theta}^2) \tag{38}$$

$$=n\bar{\theta}^2 \tag{39}$$

Equation (24) and relation (38) let us finish the proof of proposition 4:

$$\lim_{k\to\infty}\mathbb{E}\left[\left\|X_k-\bar{\theta}\mathbf{1}\right\|^2\right]=0 \tag{40}$$

□

3.5 Convergence in probability

Proposition 4 ensures the convergence in probability of X_k toward the uniform vector $\bar{\theta}\mathbf{1}$. This is the probabilistic counterpart of the convergence of sequences on a normed vector space. Proposition 10 states that the probability of having an error greater than a given threshold can be made arbitrary small, the time index being greater than a second threshold depending on the error.

Proposition 10. *For any positive real numbers α and δ, there is an integer N, such that:*

$$\forall k\geq N,\ \Pr\left[\left\|X_k-\bar{\theta}\mathbf{1}\right\|\geq\delta\right]\leq\alpha \tag{41}$$

Proof. If $(\Omega,\mathcal{B},\mathbb{P})$ is a probabilty space, and f is a measurable real-valued function on Ω, Markov's inequality states that for any $t\in\mathbb{R}^+$:

$$\mathbb{P}\left(\{\omega\in\Omega:|f(\omega)|\geq t\}\right)\leq\frac{1}{t}\mathbb{E}_{\mathbb{P}}\left[|f|\right] \tag{42}$$

In particular, for any real random variable X and $\delta>0$, this is equivalent to:

$$\Pr\left[|X|\geq\delta\right]\leq\frac{\mathbb{E}\left[|X|\right]}{\delta} \tag{43}$$

Applying inequality (43) to $\left\|X_k - \bar{\theta}\mathbf{1}\right\|$, one obtains:

$$\Pr\left[\left\|X_k - \bar{\theta}\mathbf{1}\right\| \geq \delta\right] = \Pr\left[\left\|X_k - \bar{\theta}\mathbf{1}\right\|^2 \geq \delta^2\right] \tag{44}$$

$$\leq \delta^{-2}\mathbb{E}\left[\left\|X_k - \bar{\theta}\mathbf{1}\right\|^2\right] \tag{45}$$

By proposition (4), on can find an integer N such that

$$\forall k \geq N,\ \mathbb{E}\left[\left\|X_k - \bar{\theta}\mathbf{1}\right\|^2\right] \leq \alpha\epsilon^2 \tag{46}$$

Together, equations (45) and (46) ensure that $\forall k \geq N$, one can guarantee that:

$$\Pr\left[\left\|X_k - \bar{\theta}\mathbf{1}\right\| \geq \delta\right] \leq \alpha \tag{47}$$

□

Corollary 1. $\forall\alpha > 0,\ \delta > 0,\ \exists N \in \mathbb{N}$, *such that* $\forall k \geq N$, *the two following inequalities hold:*

i) $\Pr\left[\max\limits_{1\leq i\leq n} |x_i^{(k)} - \bar{\theta}| \geq \delta\right] \leq \alpha$

ii) $\forall i \in [1, n],\ \Pr\left[|x_i^{(k)} - \bar{\theta}| \geq \delta\right] \leq \alpha$

Proof. This is a consequence of the classical norm inequality stating that

$$\forall X \in \mathbb{R}^n, \|X\|_\infty \leq \|X\|_2 \tag{48}$$

Convergence in probability is then trivially deduced from this inequality. □

4 Simulation and analysis

No formal result is still available on the rate of convergence of JEGA. Some results may be explicitly deduced from the proof, but it seems more interesting to observe directly the behaviour of quadratic error through time under the influence of parameters (noise, mean parameters, ...). Our goal is to get intuitive and empirical knowledge of global tendencies. For this purpose, we consider a wireless network modeled as a random unit disc graph[5] $\mathcal{G}$ where n nodes are distributed unifomily on a square simulation plane of width d. For the sake of simplicity, each node updates its local estimate at each iteration k (it gets one sample). However, only one pair of nodes will interact. For this purpose, an initiator i is chosen randomly among the set of vertices of $\mathcal{G}$ according to uniform distribution, while the destinator j is chosen uniformily randomly too in the set of neighbors of i. We then compute an estimate of the average value of a parameter. Measurement noise are i.i.d. zero-mean gaussian random variables with standard deviation σ_i at node i.

[5] i.e. two vertices share an edge if and only if their euclidean distance is 1

4.1 Impact of noise variance

In this set of simulations, the entries of the steady state vector $\bar{Z}$ were taken once at random uniformily in the interval $[0, 3]$[6] and kept constant during the simulations. We analyse the evolution of the mean square deviation under several values of σ in the set $\{10, 1, 0.1, 0.01\}$, taken identically for all nodes , and also for noiseless measurements ($\sigma = 0$). This last case correspond to the standard gossip averaging process. We run the algorithm for N_{iter} iterations and process N_{avg} simulations. The term MSE denotes the mean squared error norm of the state vector X_k, i.e. $\mathbb{E}\left[\left\|X_k - \bar{\theta}\mathbf{1}\right\|^2\right]$.

n	40
d	100
r	2

σ	$\{0, 0.01, 0.1, 1, 10\}$
N_{iter}	1000
N_{avg}	3000

Table 1: Simulation parameters

In the noiseless case, one recognize the classical sum-exponential convergence rate of gossip averaging algorithms, dominated for large k as follows:

$$\mathbb{E}\left[\left\|X_k - \bar{\theta}\mathbf{1}\right\|^2\right] \leq [\lambda_2(W)]^k \mathbb{E}\left[\left\|X_0 - \bar{\theta}\mathbf{1}\right\|^2\right] \tag{49}$$

Proof of this inequality can be found in [14]. In the general case, i.e. when $\sigma \neq 0$, the error due to noise cancellation should be superposed to the noiseless error. We used standard sample mean estimators on each node, which are known to converge proportionnaly to the inverse of the number of samples. This trend is conserved in the global mean squared error (MSE) as seen on figure 2, where we plotted the inverse normalized MSE. Thus, for small σ, the convergence should be seen as experiencing two differents states. The first state corresponds to the coarse homogeneization of the components of X_k just as if all estimates to be averaged were constant (classical gossip averaging, see [6] and [16]), while the second corresponds to the slow cancellation of estimation noise, as if spatial averaging was instantaneous. However one should think that there are asymptotes towards any non-zero values. This false impression is given by the logarithmic scale and the fact that estimator variances decreases proportionnaly to k^{-1}).

4.2 Impact of message exchange rate

The first set of simulations shows an increase of convergence speed when σ decreases. Thus, one should consider two way of reducing virtually σ. On one hand, the message exchange rate could be reduced by some factor, say K. On the other hand, the averaging process flows normally but data samples are

[6] this arbitrary value was chosen to exhibit to desired phenomenon

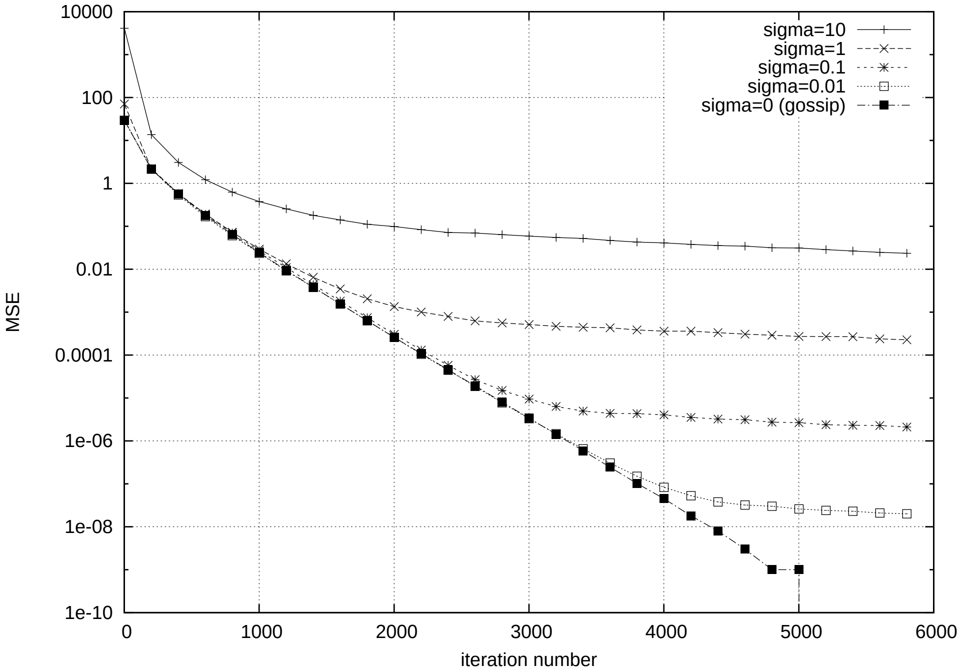

Figure 1: MSE $\mathbb{E}\left[\left\|X_k - \bar{\theta}\mathbf{1}\right\|^2\right]$ (log-scale)

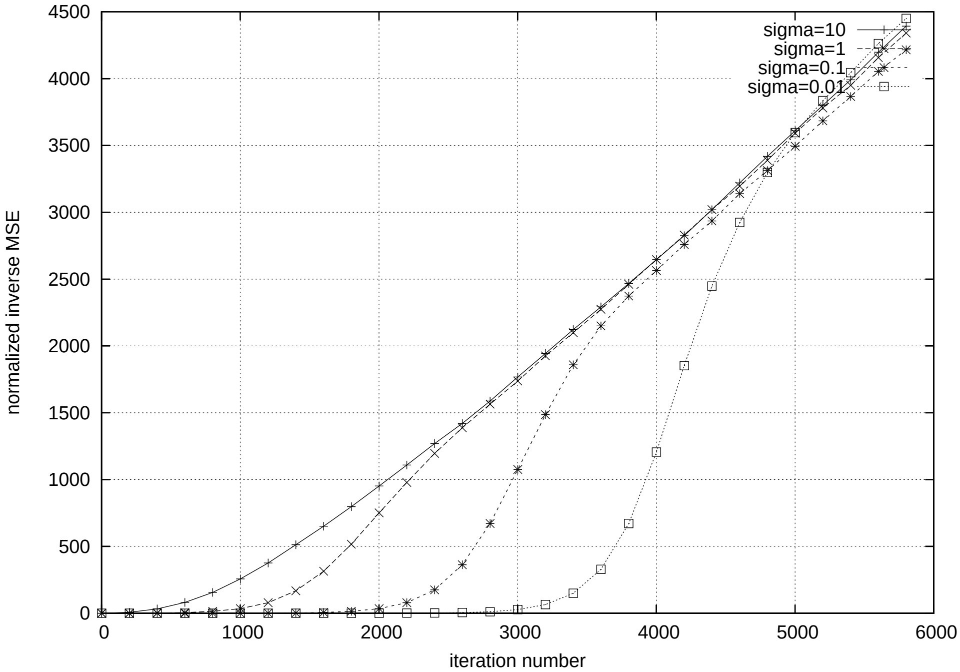

Figure 2: Inverse normalized MSE $\mathbb{E}\left[\frac{1}{\sigma^2}\left\|X_k - \bar{\theta}\mathbf{1}\right\|^2\right]^{-1}$ (log-scale)

buffered and sent in bursts to estimators: this should avoid the propagation of information of poor quality. In fact, the first proposal induces a latency proportional to K but leads to energy savings for achieving a given error (see figures 3 & 4: less exchanges are needed for a achieving a given error), while the second schemes seems not to improve convergence. In fact, simulations show that buffering creates jigsaw oscillations around the error achieved by standard scheme [7] (see figures 5 & 6).

5 Conclusion

In this paper, we introduced a new distributed algorithm for joint estimation and averaging which generalizes the space-time diffusion scheme presented in [9], and named it JEGA. We proved the convergence of our solution in terms of first and second order moments of deviation to the true average, and then deduced the convergence in probability of each local averaged estimate. As it is here based on peer-to-peer interactions, this algorithm is clearly adapted for sensor networks applications. However, we propose to generalize the proofs given here to the ase of synchronous interactions characterized by a constant transition matrix: such an approach relies on finding necessary conditions for ergodicity and verifying their consequences on networking models. The ability offered by JEGA of performing estimation and averaging in parallel gives rise to applications in cartography, localization, or synchronization in wireless sensor networks. Despite its simplicity, the main default of this algorithm is the difficulty to find a closed expression for its convergence rate. Nevertheless, by mean of simulation, our analysis provides a good heurisitic for qualitatively predict the mean behaviour of deviation through time. In a future work, we will focus our attention on a deepest analysis of convergence rate and on finding better solutions with the help of predictive/polynomial filters ([21]).

[7] no buffering, no exchange delay

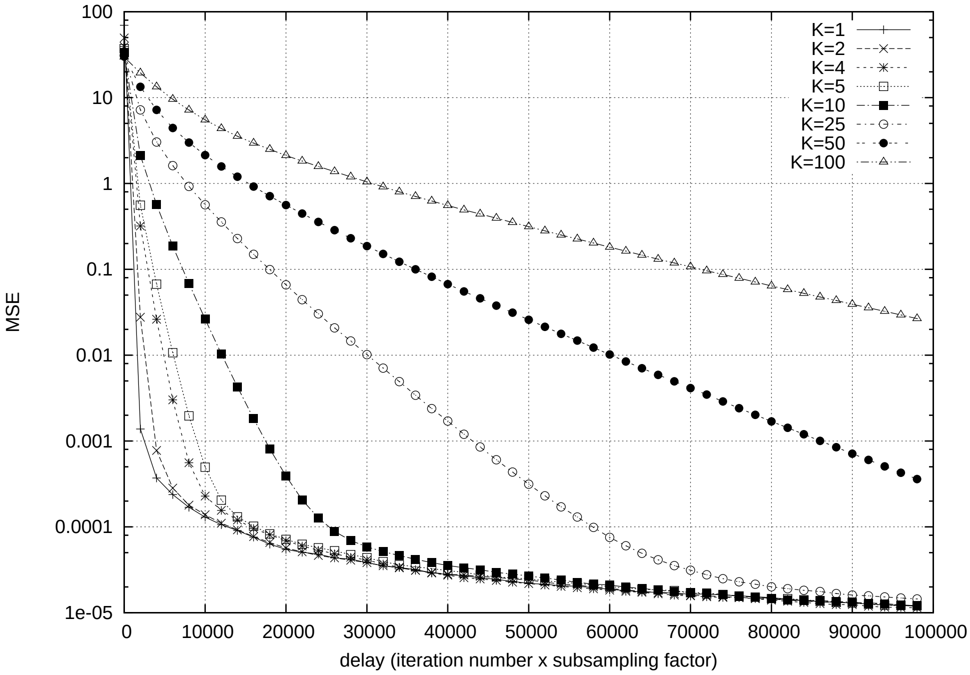

Figure 3: MSE $\mathbb{E}\left[\left\|X_k - \bar{\theta}\mathbf{1}\right\|^2\right]$ (log-scale)

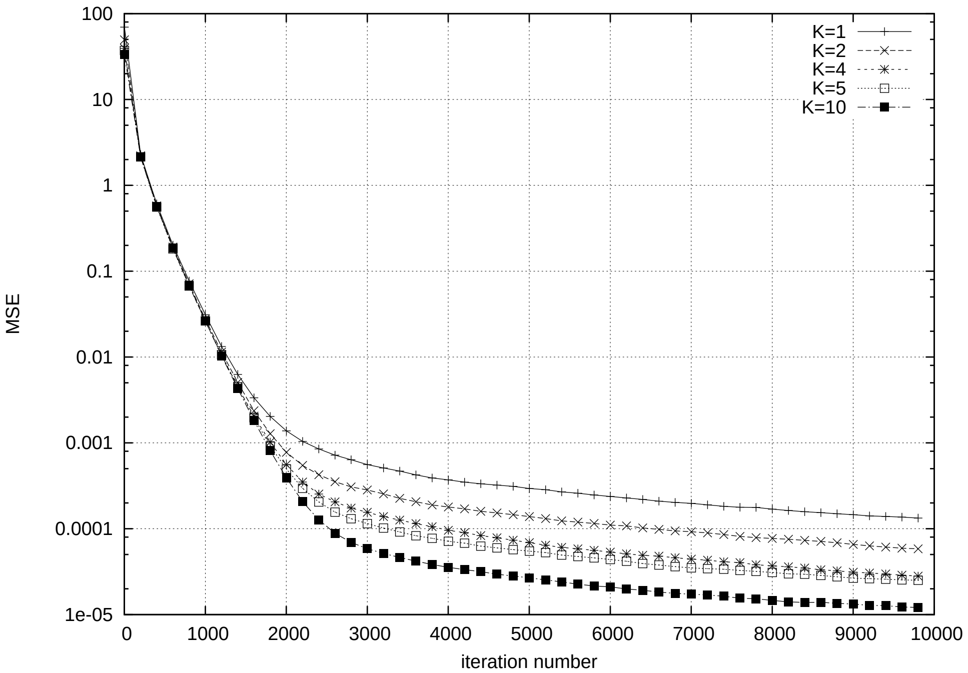

Figure 4: MSE $\mathbb{E}\left[\left\|X_k - \bar{\theta}\mathbf{1}\right\|^2\right]$ (log-scale)

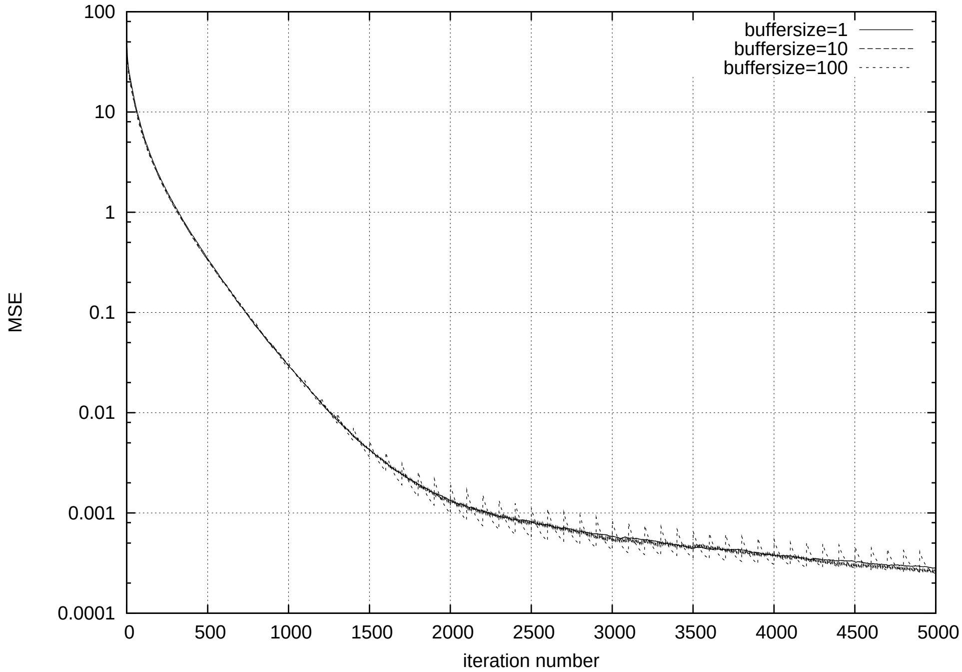

Figure 5: MSE $\mathbb{E}\left[\left\|X_k - \bar{\theta}\mathbf{1}\right\|^2\right]$ (log-scale)

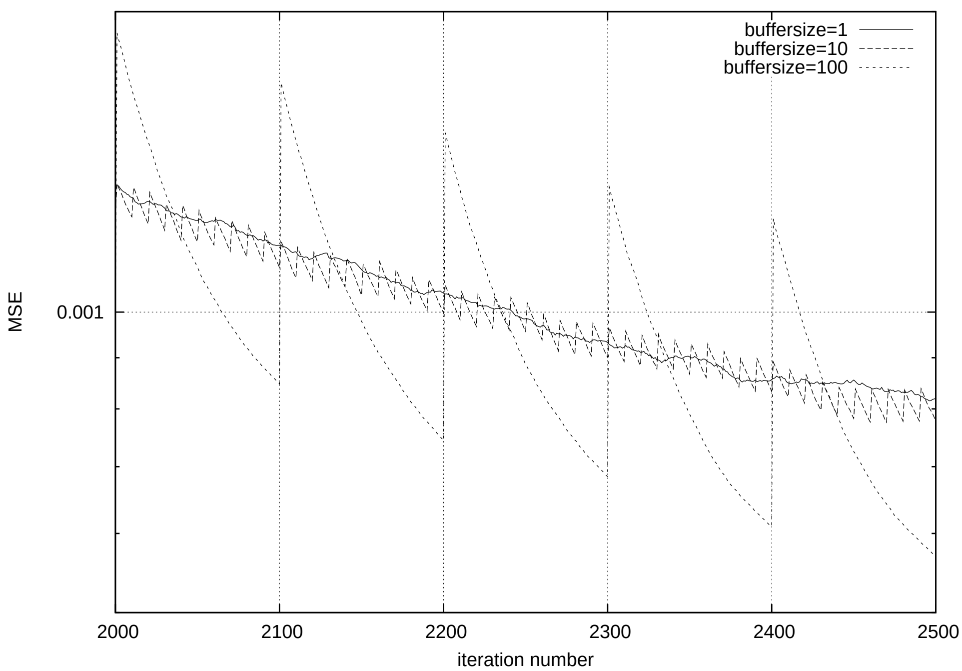

Figure 6: MSE $\mathbb{E}\left[\left\|X_k - \bar{\theta}\mathbf{1}\right\|^2\right]$ (log-scale): zoom of fig 5

Appendix A: Proofs for propositions in part 3.4

Proof of proposition 6

$$\mathbb{E}\left[\bar{Z}^{\mathsf{T}}\sum_{i=1}^{k-1}\Phi(k,i)\bar{Z}\right] = \bar{Z}^{\mathsf{T}}\mathbb{E}\left[\sum_{i=1}^{k-1}\Phi(k,i)\right]\bar{Z} \tag{50}$$

where the middle sum can be easily simplified:

$$\begin{aligned}
\mathbb{E}\left[\sum_{i=1}^{k-1}\Phi(k,i)\right] &= \mathbb{E}\left[\sum_{i=1}^{k-1}(\Psi(k,i)-\Psi(k,i+1))\right] && (51)\\
&= \mathbb{E}\left[\Psi(k,1)-\Psi(k,k)\right] && (52)\\
&= \mathbb{E}\left[W_{k-1}W_{k-2}\dots W_1\right] - I && (53)
\end{aligned}$$

As the matrices W_k are i.i.d., one can average them separately and obtains:

$$\mathbb{E}\left[\sum_{i=1}^{k-1}\Phi(k,i)\right] = W^{k-1} - I \tag{54}$$

This last expression can be reinjected in eq. (50):

$$\mathbb{E}\left[\bar{Z}^{\mathsf{T}}\sum_{i=1}^{k-1}\Phi(k,i)\bar{Z}\right] = \bar{Z}^{\mathsf{T}}\left(W^{k-1}-I\right)\bar{Z} \tag{55}$$

By mean of ergodicity of W, one have:

$$\bar{Z}^{\mathsf{T}}\left(W^{k-1}-I\right)\bar{Z} \xrightarrow[k\to\infty]{} \bar{Z}^{\mathsf{T}}\left(\frac{\mathbf{1}\mathbf{1}^{\mathsf{T}}}{n}-I\right)\bar{Z} = n\bar{\theta}^2 - \left\|\bar{Z}\right\|^2 \tag{56}$$

Proof of proposition 7

$$\begin{aligned}
\mathbb{E}\left[\left\|\sum_{i=1}^{k-1}\Phi(k,i)\bar{Z}\right\|^2\right] &= \mathbb{E}\left[\left\|\left(\sum_{i=1}^{k-1}\Phi(k,i)\right)\bar{Z}\right\|^2\right] && (57)\\
&= \mathbb{E}\left[\left\|\left(\Psi(k,1)-I\right)\bar{Z}\right\|^2\right] && (58)\\
&= \mathbb{E}\left[\bar{Z}^{\mathsf{T}}\Psi(k,1)^{\mathsf{T}}\Psi(k,1)\bar{Z}\right] && (59)\\
&\quad - 2\mathbb{E}\left[\bar{Z}^{\mathsf{T}}\Psi(k,1)\bar{Z}\right] + \left\|\bar{Z}\right\|^2 && (60)
\end{aligned}$$

Equation (59) is equivalent to the squared norm of the state vector in standard gossip averaging algorithms [16], and thus tends naturally toward the norm of the average consensus vector:

$$\mathbb{E}\left[\bar{Z}^{\mathsf{T}}\Psi(k,1)^{\mathsf{T}}\Psi(k,1)\bar{Z}\right] \xrightarrow[k\to\infty]{} \left\|\bar{\theta}\mathbf{1}\right\|^2 = n\bar{\theta}^2 \tag{61}$$

As eq. (60) is equivalent to eq. (56), we can compute the limit of eq. (57) for an infinite k:

$$\mathbb{E}\left[\left\|\sum_{i=1}^{k-1}\Phi(k,i)\bar{Z}\right\|^2\right] \xrightarrow[k\to\infty]{} \left\|\bar{Z}\right\|^2 - n\bar{\theta}^2 \tag{62}$$

Proof of proposition 8

This proof relies on the use of the following statement, which is a special case of the theorems proved in [22]:

Proposition 11 ([22]). *Let E be either $\mathbb{R}$ or $\mathbb{C}$, $(u_n)_{n\in\mathbb{N}}$ a sequence of elements of E such that $\lim_{n\to\infty} \|u_n\| = 0$, and $z \in E$ with $|z| < 1$. Then,*

$$\lim_{n\to+\infty} \sum_{k=1}^{n} u_k z^{n-k} = 0 \tag{63}$$

W is clearly symmetric and positive-semidefinite. It implies that W can be diagonalized by an orthogonal matrix. In the following, P_j denotes the eigenvector of W associated with the j-th eigenvalue λ_j.

$$\begin{aligned}
\mathbb{E}\left[B_k^{\mathsf{T}} \sum_{i=1}^{k-1} \Phi(k,i) B_i\right] &= \sum_{i=1}^{k-1} \mathbb{E}\left[B_k^{\mathsf{T}} \mathbb{E}\left[\Phi(k,i)\right] B_i\right] &(64)\\
&= \sum_{i=1}^{k-1} \mathbb{E}\left[B_k^{\mathsf{T}} \left(W^{k-i} - W^{k-i-1}\right) B_i\right] &(65)\\
&= \sum_{j=1}^{n} \sum_{i=1}^{k-1} \left(\lambda_j^{k-i} - \lambda_j^{k-i-1}\right) \mathbb{E}\left[\left(B_k^{\mathsf{T}} P_j\right)\left(P_j^{\mathsf{T}} B_i\right)\right] &(66)
\end{aligned}$$

where equation (65) results from the independance of estimation noises and exchange matrices. For $\lambda_j = 1$, $\lambda_j^{k-i} - \lambda_j^{k-i-1}$ vanishes. On the other side, the covariance term can be bounded by a non-negative sequence having a zero limit at infinity:

$$\begin{aligned}
\left|\mathbb{E}\left[\left(B_k^{\mathsf{T}} P_j\right)\left(P_j^{\mathsf{T}} B_i\right)\right]\right| &= \left|\sum_{u,v} \left[P_j P_j^{\mathsf{T}}\right]_{uv} C_{uv}^{ki}\right| &(67)\\
&\leq n^2 \max_{u,v} \max_{\substack{k\in\mathbb{N}\\ k>i}} |C_{uv}^{ki}| \triangleq n^2 C_i &(68)
\end{aligned}$$

The prerequisite condition on C_{uv}^{ki} ensures that C_i converges to 0. Thus, $\mathbb{E}\left[\left(B_k^{\mathsf{T}} P_j\right)\left(P_j^{\mathsf{T}} B_i\right)\right]$ does similarly for all normalized vector P_j. On another side, $|\lambda_j| < 1$ whenever $\lambda_j \neq 1$. By splitting the summation over eigenvalues and developping the multiplication by $\lambda_j^{k-i} - \lambda_j^{k-i-1}$, equation (66) can be easily expressed as a linear combination of $2(n-1)$ sums, each of which satifies proposition 11. This implies that (66) goes to 0 as k grows, independantly of the decrease rate of covariances[8].

Proof of proposition 9

Let us start by rewritting equation (9):

$$\mathbb{E}\left[\left\|\sum_{i=1}^{k-1} \Phi(k,i) B_i\right\|^2\right] = \sum_{i=1}^{k-1} \sum_{j=1}^{k-1} \mathbb{E}\left[B_j^{\mathsf{T}} \Phi(k,j)^{\mathsf{T}} \Phi(k,i) B_i\right] \tag{69}$$

[8] the partial sum (65) must be carefully manipulated in order to avoid divergence at large k

As we develop the term $\Phi(k,j)^{\mathsf{T}}\Phi(k,i)$, we observe once again that the components of each B_i and B_j that are colinear to $\mathbf{1}$ (associated with eigenvalue 1) are not transmitted through $\Phi(k,i)$:

$$\begin{aligned}\Phi(k,j)^{\mathsf{T}}\Phi(k,i) = {} & \Psi(k,j)^{\mathsf{T}}\Psi(k,i) - \Psi(k,j+1)^{\mathsf{T}}\Psi(k,i) \\ & - \Psi(k,j)^{\mathsf{T}}\Psi(k,i+1) + \Psi(k,j+1)^{\mathsf{T}}\Psi(k,i+1) \quad (70)\end{aligned}$$

For the sake of simplicity, $\Xi_k(i,j)$ denotes $\mathbb{E}\left[\Psi(k,j)^{\mathsf{T}}\Psi(k,i)\right]$. It is useful to notice that $\Xi_k(i,j)$ can be factorized by externalizing terms in i:

$$\begin{aligned}\Xi_k(i,j) &= \mathbb{E}\left[\Psi(k,j)^{\mathsf{T}}\Psi(k,i)\right] && (71)\\ &= \mathbb{E}\left[\Psi(k,j)^{\mathsf{T}}\Psi(k,j)\Psi(j,i)\right] && (72)\\ &= \mathbb{E}\left[\Psi(k,j)^{\mathsf{T}}\Psi(k,j)W^{j-i}\right] && (73)\\ &= \Xi_k(j,j)W^{j-i} && (74)\end{aligned}$$

Following the approach of [14] helps bounding the spectral radius of $\Xi_k(j,j)$. For any n-dimensional vector $X \perp \mathbf{1}$, the following inequalities hold:

$$\begin{aligned}X^{\mathsf{T}}\Xi_k(j,j)X &= X^{\mathsf{T}}\mathbb{E}\left[\Psi(k,j)^{\mathsf{T}}\Psi(k,j)\right]X && (75)\\ &= X^{\mathsf{T}}\mathbb{E}\left[\Psi(k-1,j)^{\mathsf{T}}W\Psi(k-1,j)\right]X && (76)\\ &\leq \lambda_2 X^{\mathsf{T}}\mathbb{E}\left[\Psi(k-1,j)^{\mathsf{T}}\Psi(k-1,j)\right]X && (77)\\ &\leq \lambda_2^{k-j}\left\|X\right\|^2 && (78)\end{aligned}$$

Using this inequality in conjunction with the Rayleigh-Ritz characterization theorem, one obtains:

$$\begin{aligned}\rho_\Psi(k,j) &\triangleq \rho\left(\Xi_k(j,j) - \frac{\mathbf{1}\mathbf{1}^{\mathsf{T}}}{n}\right) && (79)\\ &= \max_{\substack{X\perp\mathbf{1}\\ \|X\|=1}}\left\{X^{\mathsf{T}}\mathbb{E}\left[\Psi(k,j)^{\mathsf{T}}\Psi(k,j)\right]X\right\} && (80)\\ &\leq \lambda_2^{k-j} && (81)\end{aligned}$$

Recalling that any component of B_i colinear to $\mathbf{1}$ can be cancelled, the modulus of its coordinates on a basis of eigenvectors of W are then rescaled by a factor less than or equal to λ_2^{j-i} when we apply W^{j-i}. This gives rise to the following inequality:

$$\begin{aligned}\left|\mathbb{E}\left[B_j^{\mathsf{T}}V_kV_k^{\mathsf{T}}W^{j-i}B_i\right]\right| &= \left|\sum \lambda_p^{j-i}\mathbb{E}\left[B_j^{\mathsf{T}}V_kV_k^{\mathsf{T}}\underline{V}_p\underline{V}_p^{\mathsf{T}}B_i\right]\right| && (82)\\ &\leq (n-1)n^2\lambda_2^{j-i}\max_{u,v}|C_{uv}^{ij}| && (83)\end{aligned}$$

where V_k is any unitary vector of $\mathbb{R}^n$, and $\underline{V}_p$ is an eigenvector of W (and then of W^{j-i}) associated with eigenvalue λ_p. Let $\mathcal{S}_k^{ij}$ be the spectrum of $\Xi_k(i,j)$, constituted of eigenvalues λ_ψ of associated eigenvector V_{λ_Ψ}. Using this notation, $\mathbb{E}\left[B_j^\mathsf{T}\Xi_k(i,j)B_i\right]$ can be brought into a useful form for bounding purposes:

$$\left|\mathbb{E}\left[B_j^\mathsf{T}\Xi_k(i,j)B_i\right]\right| = \left|\mathbb{E}\left[B_j^\mathsf{T}\Xi_k(j,j)W^{j-i}B_i\right]\right| \tag{84}$$

$$\leq \sum_{\substack{\lambda_\Psi\in\mathcal{S}_k^{ij}\\ \lambda_\psi\neq 1}} \left|\lambda_\Psi\mathbb{E}\left[\left(B_j^\mathsf{T}V_{\lambda_\Psi}V_{\lambda_\Psi}^\mathsf{T}W^{j-i}B_i\right)\right]\right| \tag{85}$$

$$\leq (n-1)^2\rho_\Psi(k,j)n^2\lambda_2^{j-i}\max_{u,v}|C_{uv}^{ij}| \tag{86}$$

Coupling equations (81) and (83), the following majorization states:

$$\left|\mathbb{E}\left[B_j^\mathsf{T}\Xi_k(i,j)B_i\right]\right| \leq (n-1)^2n^2\lambda_2^{k-i}\max_{u,v}|C_{uv}^{ij}| \tag{87}$$

Let us define $\xi \triangleq \sqrt{\lambda_2}$, and remember that $i<j$: immediately $0\leq\lambda_2^{k-i} = \xi^{2k-2i}\leq\xi^{2k-i-j}$. This inequality hence implies:

$$\left|\sum_{i<j}\mathbb{E}\left[B_j^\mathsf{T}\Xi_k(i,j)B_i\right]\right| \leq (n-1)^2n^2\sum_{i<j}\lambda_2^{k-i}\max_{u,v}|C_{uv}^{ij}| \tag{88}$$

$$\leq \underbrace{(n-1)^2n^2\sum_{i<j}\xi^{2k-i-j}\max_{u,v}|C_{uv}^{ij}|}_{\xrightarrow[k\to\infty]{}0} \tag{89}$$

The convergence toward 0 of the right-hand side is ensured by proposition 12:

Proposition 12. *([22]) Let E be either $\mathbb{R}$ or $\mathbb{C}$, $(u_{ij})_{(i,j)\in\mathbb{N}^2}\in E^{\mathbb{N}^2}$ such that $\lim\limits_{(i+j)\to\infty}\|u_{ij}\|=0$, and $z\in E$ with $|z|<1$. Then*

$$\lim_{n\to+\infty}\sum_{1\leq i,j\leq n}u_{ij}z^{2n-i-j}=0 \tag{90}$$

When $i=j$, a classical result from [14] states that:

$$\mathbb{E}\left[B_i^\mathsf{T}\Xi_k(i,i)B_i\right] \leq \lambda_2^{k-i}\mathbb{E}\left[\|B_i\|^2\right] \tag{91}$$

Then, summing over i gives the diagonal limit:

$$0\leq\sum_{i=1}^{k-1}\mathbb{E}\left[B_i^\mathsf{T}\Xi_k(i,i)B_i\right] \leq \underbrace{\sum_{i=1}^{k-1}\lambda_2^{k-i}\mathbb{E}\left[\|B_i\|^2\right]}_{\xrightarrow[k\to\infty]{}0} \tag{92}$$

In the same way, it is easy to bound terms with $\Xi_k(i+1,j)$, $\Xi_k(i,j+1)$ or $\Xi_k(i+1,j+1)$ in equation (69). We obtain 4 sums in the spirit of the combination of equations (88) and (92) that tend to 0 as k goes to infinity.

Appendix B: Optimality of $1/2$ weights

Given a weight $\alpha \in [0,1]$ for exchanges, the transition matrix W_{ij} takes the form:

$$W_{ij} = I - \alpha(e_i - e_j)(e_i - e_j)^\mathsf{T} \tag{93}$$

Under the assumption that W_k i.i.d. random matrices drawn from the set of W_{ij}, performance bounds are derived given the second smallest eigenvalue of $\mathbb{E}\left[W_k^\mathsf{T} W_k\right]$ (see [15]), i.e.:

$$\lambda_2^{(\alpha)} \triangleq \rho\left(\mathbb{E}\left[W_k^\mathsf{T} W_k\right] - \frac{\mathbf{1}\mathbf{1}^\mathsf{T}}{n}\right) \tag{94}$$

$$= \max_{\substack{X \perp \mathbf{1} \\ \|X\|=1}} X^\mathsf{T} \mathbb{E}\left[W_k^\mathsf{T} W_k\right] X \tag{95}$$

One can try to optimize this criterion according to parameter α. Proposition 13 answers simply to this problem.

Proposition 13. *For peer-to-peer exchanges with fixed link probabilities and fix exchange weight $\alpha \in [0,1]$, $\lambda_2^{(\alpha)}$ is optimal for $\alpha = 1/2$.*

Proof. Let p_{ij} be the probabilty that, if node i is chosen as initiator, it contacts node j. Then $\mathbb{E}\left[W_k^\mathsf{T} W_k\right]$ can be easily rewritten with the same notations as in equation (93):

$$\mathbb{E}\left[W_k^\mathsf{T} W_k\right] = \frac{1}{n} \sum_{i,j} p_{ij} W_{ij}^\mathsf{T} W_{ij} \tag{96}$$

$$= I - 2\alpha(1-\alpha)\left(e_i - e_j\right)\left(e_i - e_j\right)^\mathsf{T} \tag{97}$$

Then, writting $\lambda_i(M)$ the i-th eigenvalue of M in decreasing order, one has:

$$\lambda_2^{(\alpha)} = \rho\left(\frac{1}{n}\sum_{i \sim j} p_{ij} {W_{ij}}^\mathsf{T} W_{ij}\right) \tag{98}$$

$$= \lambda_2\left(I - 2\alpha(1-\alpha)\left(e_i - e_j\right)\left(e_i - e_j\right)^\mathsf{T}\right) \tag{99}$$

$$= 1 - 2\alpha(1-\alpha)\lambda_{n-1}\left(\sum_{i \sim j} p_{ij} E_{ij}\right) \tag{100}$$

$$= 1 - 2\alpha(1-\alpha)\mu \tag{101}$$

Now, derivating $\lambda_2^{(\alpha)}$ w.r.t. weight α gives:

$$\frac{\partial \lambda_2^{(\alpha)}}{\partial \alpha} = (-2 + 4\alpha)\mu \tag{102}$$

Since $\mathbb{E}\left[W_k\right]$ is ergodic, μ cannot be null . Thus the only way for $\lambda_2^{(\alpha)}$ to have a null derivative is that $\alpha = 1/2$. The matrix $\mathbb{E}\left[W_k^\mathsf{T} W_k\right]$ is trivially doubly stochastic and then $\lambda_2^{(\alpha)} \leq 1 = \lambda_2^{(0)}$. As a consequence, $\lambda_2^{(\alpha)}$ is minimum for $\alpha = 1/2$. $\square$

References

[1] K. Sohraby, D. Minoli, and T. Znati, *Wireless Sensor Networks: Technology, Protocols and Applications.* Wiley-Interscience, 2007.

[2] P. Juang, H. Oki, Y. Wang, M. Martonosi, L. Peh, and D. Rubenstein, "Energy-efficient computing for wildlife tracking: Design tradeoffs and early experiences with zebranet," in *ASPLOS, San Jose, CA*, oct 2002.

[3] A. Mainwaring, D. Culler, J. Polastre, R. Szewczyk, and J. Anderson, "Wireless sensor networks for habitat monitoring," in *WSNA '02: Proceedings of the 1st ACM international workshop on Wireless sensor networks and applications.* New York, NY, USA: ACM, 2002, pp. 88–97.

[4] A. Terzis, A. Anandarajah, K. Moore, and I.-J. Wang, "Slip surface localization in wireless sensor networks for landslide prediction," in *IPSN '06: Proceedings of the fifth international conference on Information processing in sensor networks.* New York, NY, USA: ACM, 2006, pp. 109–116.

[5] M. V. Ramesh, R. Raj, J. U. Freeman, S. Kumar, and P. V. Rangan, "Factors and approaches towards energy optimized wireless sensor networks to detect rainfall induced landslides," in *ICWN*, 2007, pp. 435–438.

[6] L. Xiao, S. Boyd, and S. Lall, "A scheme for robust distributed sensor fusion based on average consensus," in *IPSN '05: Proceedings of the 4th international symposium on Information processing in sensor networks.* Piscataway, NJ, USA: IEEE Press, 2005, p. 9.

[7] A. Jadbabaie, J. Lin, and A. Morse, "Coordination of groups of mobile autonomous agents using nearest neighbor rules," *IEEE Transactions on Automatic Control*, vol. 48, no. 6, pp. 988 – 1001, 2002.

[8] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," 2003. [Online]. Available: citeseer.comp.nus.edu.sg/xiao03fast.html

[9] L. Xiao, S. Boyd, and S. Lall, "A space-time diffusion scheme for peer-to-peer least-squares estimation," in *IPSN '06: Proceedings of the fifth international conference on Information processing in sensor networks.* New York, NY, USA: ACM Press, 2006, pp. 168–176.

[10] N. Maréchal, J. B. Pierrot, and J. M. Gorce, "Fine synchronization for wireless sensor networks using gossip averaging algorithms," 2007, [Submitted to IEEE ICC'08].

[11] M. Jelasity, A. Montresor, and O. Babaoglu, "Gossip-based aggregation in large dynamic networks," *ACM Trans. Comput. Syst.*, vol. 23, no. 3, pp. 219–252, 2005.

[12] D. Kempe, A. Dobra, and J. Gehrke, "Gossip-based computation of aggregate information," *focs*, vol. 00, p. 482, 2003.

[13] R. Olfati-Saber, J. A. Fax, and R. M. Murray, "Consensus and cooperation in networked multi-agent systems," in *Proc. of the IEEE*, Jan 2007.

[14] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Gossip algorithms: Design, analysis and applications," pp. 1653–1664, 2005.

[15] ——, "Analysis and optimization of randomized gossip algorithms," in *Proc. of 44th IEEE Conference on Decision and Control (CDC 2004)*, Dec 2004, pp. 5310 – 5315.

[16] ——, "Randomized gossip algorithms," *IEEE/ACM Trans. Netw.*, vol. 14, no. SI, pp. 2508–2530, 2006.

[17] A. G. Dimakis, A. D. Sarwate, and M. J. Wainwright, "Geographic gossip: Efficient aggregation for sensor networks," 2006.

[18] M. E. Yildiz and A. Scaglione, "Differential nested lattice encoding for consensus problems," in *IPSN '07: Proceedings of the 6th international conference on Information processing in sensor networks*. New York, NY, USA: ACM Press, 2007, pp. 89–98.

[19] L. Xiao, S. Boyd, and S.-J. Kim, "Distributed average consensus with least-mean-square deviation," *J. Parallel Distrib. Comput.*, vol. 67, no. 1, pp. 33–46, 2007.

[20] W. Beyn and L. Elsner, "Infinite products and paracontracting matrices," *Electronic Transactions on Numerical Analysis*, vol. 2, pp. 183–193, 1997. [Online]. Available: citeseer.ist.psu.edu/elsner97infinite.html

[21] E. Kokiopoulou and P. Frossard, "Polynomial Filtering for Fast Convergence in Distributed Consensus," *IEEE Transactions on Signal Processing*, 2008.

[22] N. Maréchal, "On the action of a convolution-like operator on spaces of multidimensional sequences having isotropic limits," 2007, [CEA internal report - available on demand].

[23] R. A. Horn and C. R. Johnson, *Matrix analysis*. New York, NY, USA: Cambridge University Press, 1986.

[24] S. N. Elaydi, *An introduction to difference equations*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 1996.

[25] C. Godsil and G. Royle, *Algebraic Graph Theory*, ser. Graduate Texts in Mathematics. Springer-Verlag, 2001.

[26] D. S. Scherber and H. C. Papadopoulos, "Locally constructed algorithms for distributed computations in ad-hoc networks," in *IPSN '04: Proceedings of the third international symposium on Information processing in sensor networks*. New York, NY, USA: ACM Press, 2004, pp. 11–19.

[27] D. L. Hall and J. Llinas, *Handbook of Multisensor Data Fusion*. CRC Press, 2001.

[28] D. Estrin, R. Govindan, J. Heidemann, and S. Kumar, "Next century challenges: scalable coordination in sensor networks," in *MobiCom '99: Proceedings of the 5th annual ACM/IEEE international conference on Mobile computing and networking*, New York, NY, USA, 1999, pp. 263–270.

Unité de recherche INRIA Rhône-Alpes
655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier (France)

Unité de recherche INRIA Futurs : Parc Club Orsay Université - ZAC des Vignes
4, rue Jacques Monod - 91893 ORSAY Cedex (France)
Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)
Unité de recherche INRIA Sophia Antipolis : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399