Stepsize Sequence Design for Distributed Average Consensus

Carlos Mosquera*, *Member, IEEE*, Roberto López-Valcarce*, *Member, IEEE*, Sudharman K. Jayaweera†, *Senior Member, IEEE*

* Departamento de Teoría de la Señal y Comunicaciones, Universidad de Vigo, 36310 Vigo, Spain. E-mail: {mosquera,valcarce}@gts.tsc.uvigo.es.

† Department of Electrical and Computer Engineering, University of New Mexico, Albuquerque, NM 87131-0001, USA. E-mail: jayaweera@ece.unm.edu.

***Abstract*— Starting from local observations, iterative consensus algorithms attempt to drive a sensor network to a common estimate in a decentralized, incremental manner. When additive noise perturbs the sensor exchanges, a decreasing stepsize guarantees convergence under certain conditions, although the design of such stepsize sequence for fastest convergence is an unsettled issue. We present a greedy approach to stepsize sequence design, which minimizes the mean squared error at each iteration. This design requires knowledge of the network topology; in order to overcome this drawback, a modified design based only on average descriptors of the network is also developed.**

***Index Terms*— Consensus algorithms, distributed estimation, parameter estimation, sensor networks.**

I. Introduction

We address the distributed estimation of a scalar parameter by a sensor network. Each node updates its estimate by exchanging values with its neighbors, driving the whole network to a common estimate or consensus through an iterative process; see e.g. [1] for a review of consensus algorithms. The imperfect nature of the exchanges and, in particular, unavoidable additive noise, makes it necessary to adopt some cautions, as already noted in [2]. For example, [3]-[4] adopt a PLL-like approach, updating the state (phase) and using the incremental differences (frequency) as estimate. With this approach, the required noise resilience comes at the price of unbounded growth of the state variable. In addition, consensus is not strictly achieved unless the stepsize is allowed to be time-varying and asymptotically vanishing. In fact, a decreasing stepsize sequence was used in [5], [6] and [7], which proposed sequences of the type a/k^b, where a, b are constants and k is the iteration number. Although by proper choice of b convergence can be guaranteed, good performance in terms of final mean squared error (MSE) and speed is not guaranteed. In this letter we propose a greedy approach by which the stepsize is dynamically chosen to minimize the MSE at each iteration. The resulting stepsize sequence yields fast convergence, although it requires knowledge of the complete network topology. In practice, however, *a priori* knowledge about the network beyond general descriptors, such as, e.g., sensor density, may not be available. Hence, in order to sidestep this problem, we develop a modified design in which only the average number of neighbors is needed, together with the observation and communication noise powers.

In Section II we introduce the framework. The distributed estimation method and the greedy stepsize sequence design are presented in Section III. Section IV presents the modified design, and numerical simulations and conclusions are provided in Sections V and VI respectively.

University of Vigo's contribution was supported by the Spanish Government under projects SPROACTIVE (ref. TEC2007-68094-C02-01/TCM) and COMONSENS (CONSOLIDER-INGENIO 2010 CSD2008-00010). S. K. Jayaweera was supported in part by the National Science foundation (NSF) under grant CCF-0830545. Parts of this work were previously presented at the 9th IEEE International Workshop on Signal Processing Advances for Wireless Communications, Recife, Brasil, July 2008.

II. Sensor Network Model

Let $G = (V, E)$ be a graph with M nodes (sensors) $v_m \in V$ and edges $e_{ij} \in E$ if there is a path from v_i to v_j. The adjacency matrix $\boldsymbol{\mathcal{A}}$ is defined by $[\boldsymbol{\mathcal{A}}]_{ij} = 1$ if $e_{ji} \in E$ and zero otherwise. We consider graphs that are undirected ($\boldsymbol{\mathcal{A}} = \boldsymbol{\mathcal{A}}^T$) and connected (there exists a sequence of edges from any v_i to any v_j). The degree matrix $\boldsymbol{D}$ is a diagonal matrix such that $[\boldsymbol{D}]_{ii}$ is equal to the number of connections entering node i. The Laplacian matrix $\boldsymbol{L}$ is defined as $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{\mathcal{A}}$; for undirected graphs, $\boldsymbol{L} = \boldsymbol{L}^T$. Note that $\boldsymbol{L}\mathbf{1} = \mathbf{0}$, where $\mathbf{1}$ is the all-ones vector. The number of total connections of G is $\Delta(G) \doteq \mathbf{1}^T\boldsymbol{D}\mathbf{1}$.

The eigenvalues of $\boldsymbol{L}$, $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_M$, contain significant information about the topology of G: in particular, $\lambda_1 = 0$, whereas for a connected graph, $\lambda_2 > 0$. The so-called *algebraic connectivity* λ_2 plays a major role in the speed at which information can diffuse through the network [1].

The additive noise $n_{ij}(k)$ in the signal received at sensor j from sensor i in the k-th information exchange (as described below) is assumed zero-mean, independent among different sensor links, and temporally white. These noise samples are collected in the matrix $\boldsymbol{N}(k) = [n_{ij}(k)]$, where $n_{ii}(k) = 0$ for all i, k. We assume the same noise variance σ_n^2 in all links, although this assumption can be relaxed as we will conclude later.

III. Distributed Estimation

Let $\boldsymbol{x}(0) \doteq [\ x_1 \quad x_2 \quad \cdots \quad x_M\ ]^T$ collect the initial observations of an unkown value x at the M sensors. These estimates are assumed unbiased ($\mathbb{E}\{x_i\} = x$ for all i), uncorrelated, and with the same variance ($\mathbb{E}\{(x_i - x)(x_j - x)\} = \sigma^2$ if $i = j$ and zero otherwise). All sensors update their estimates

in a process driven by exchanges with their neighbors, in order to reduce the estimation error and to eventually agree on a final value. At step k the vector of estimates is $\boldsymbol{x}(k)$, whereas the error vector is

$$\boldsymbol{e}(k) \doteq \boldsymbol{x}(k) - x\mathbf{1}. \tag{1}$$

The error covariance matrix $\boldsymbol{R}(k) \doteq \mathbb{E}\{\boldsymbol{e}(k)\boldsymbol{e}^T(k)\}$ satisfies $\boldsymbol{R}(0) = \sigma^2\boldsymbol{I}$. At iteration k, the signal from node i received at sensor j is corrupted by communication noise, and is given by $x_i(k) + n_{ij}(k)$. For some weight matrix $\boldsymbol{W}(k)$, the update is given by

$$\boldsymbol{x}(k+1) = \boldsymbol{W}(k)\boldsymbol{x}(k) + \text{diag}\{\boldsymbol{W}(k)\boldsymbol{N}(k)\} \tag{2}$$

where $\text{diag}\{\boldsymbol{C}\}$ denotes a vector with the diagonal elements of $\boldsymbol{C}$. In the absence of noise, $\boldsymbol{x}(k+1) = \boldsymbol{W}(k)\boldsymbol{x}(k)$, and asymptotic alignment of all the elements of $\boldsymbol{x}(k)$ can be guaranteed with a constant Perron matrix $\boldsymbol{W} = \boldsymbol{I} - \gamma\boldsymbol{L}$ if $0 < \gamma < 1/\max\{[\boldsymbol{D}]_{ii}\}$ [1]. In such a case $\boldsymbol{W}$ has an eigenvalue equal to 1 with corresponding eigenvector $\mathbf{1}$, and $\lim_{k\to\infty}\boldsymbol{W}^k = \mathbf{1}\mathbf{1}^T/M$. Then the sequence of estimates converges to $\lim_{k\to\infty}\boldsymbol{x}(k) = [\mathbf{1}^T\boldsymbol{x}(0)]\cdot\mathbf{1}$, meaning that all nodes asymptotically agree on the same value, which is the Best Linear Unbiased Estimate (BLUE) for this problem and has variance σ^2/M. However, in the presence of noisy exchanges this scheme blows up [2]. To sidestep this problem, the weight matrix can be made time-varying, i.e., $\boldsymbol{W}(k) = \boldsymbol{I} - \gamma(k)\boldsymbol{L}$, with $\gamma(k)$ a stepsize sequence to be designed. It is well known from the stochastic approximation literature [8], [9] that, in order to ensure asymptotic convergence of $\boldsymbol{x}(k)$ to a constant vector, the positive sequence $\gamma(k)$ must satisfy

$$\sum_{k=0}^{\infty}\gamma(k) = \infty, \quad \sum_{k=0}^{\infty}\gamma^2(k) < \infty. \tag{3}$$

The sequence $\gamma(k)$ will influence the final error: with noisy exchanges and a time-varying stepsize, and provided that convergence is achieved, the consensus value will not in general be equal to the BLUE, and will depend on the noise power and stepsize sequence. In terms of estimation variance, communication noise results in a gap with respect to the performance of the BLUE as detailed next.

It follows from (2) that the error updates as

$$\boldsymbol{e}(k+1) = \boldsymbol{W}(k)\boldsymbol{e}(k) + (\boldsymbol{W}(k) - \boldsymbol{I})x\mathbf{1} + \text{diag}\{\boldsymbol{W}(k)\boldsymbol{N}(k)\}. \tag{4}$$

Taking expectations in (4), it is seen that the estimates remain unbiased along the iterations provided that $(\boldsymbol{W}(k) - \boldsymbol{I})\mathbf{1} = \mathbf{0}$, which indeed holds for $\boldsymbol{W}(k) = \boldsymbol{I} - \gamma(k)\boldsymbol{L}$. Then

$$\boldsymbol{e}(k+1) = (\boldsymbol{I} - \gamma(k)\boldsymbol{L})\boldsymbol{e}(k) + \gamma(k)\text{diag}\{\boldsymbol{\mathcal{A}}\boldsymbol{N}(k)\}. \tag{5}$$

Since $\boldsymbol{N}(k)$ and $\boldsymbol{e}(k)$ are uncorrelated, $\boldsymbol{R}(k)$ obeys the recursion

$$\boldsymbol{R}(k+1) = (\boldsymbol{I} - \gamma(k)\boldsymbol{L})\boldsymbol{R}(k)(\boldsymbol{I} - \gamma(k)\boldsymbol{L}^T) + \gamma^2(k)\sigma_n^2\boldsymbol{D}, \tag{6}$$

from which

$$\begin{aligned}\boldsymbol{R}(k+1) = \boldsymbol{R}(k) - \gamma(k)\boldsymbol{R}(k)\boldsymbol{L}^T - \gamma(k)\boldsymbol{L}\boldsymbol{R}(k) \\ + \gamma^2(k)\boldsymbol{L}\boldsymbol{R}(k)\boldsymbol{L}^T + \gamma^2(k)\sigma_n^2\boldsymbol{D}.\end{aligned} \tag{7}$$

If consensus is achieved, $\boldsymbol{R}(k)$ must asymptotically approach $\sigma_\infty^2\mathbf{1}\mathbf{1}^T$, where σ_∞^2 denotes the asymptotic estimation error variance. Thus, in that case, $\sigma_\infty^2 = \lim_{k\to\infty}(\mathbf{1}^T\boldsymbol{R}(k)\mathbf{1})/M^2$. This value can be readily obtained from (7) when $\boldsymbol{R}(0) = \sigma^2\boldsymbol{I}$, and is given by

$$\sigma_\infty^2 = \frac{\sigma^2}{M} + \frac{\Delta(G)}{M^2}\sigma_n^2\sum_{j=0}^{\infty}\gamma^2(j). \tag{8}$$

This equation can be extended to time-variant cases for which the Laplacian matrix $\boldsymbol{L}$ changes with time [9]: in essence, the noise contribution must be modulated at each step k as a function of its average power on the active links. It follows that, when choosing the stepsize sequence, there is a tradeoff between convergence speed and excess MSE. For a given convergence rate, $\sum_{j=0}^{\infty}\gamma^2(j)$ should be as low as possible; at the same time, $\sum_{k=0}^{\infty}\gamma(k) = \infty$ must hold: this fact was also raised in [9]. Unfortunately, it seems difficult to translate all these considerations into an optimization problem for the selection of the stepsize sequence. Instead, we propose a greedy approach in which $\gamma(k)$ is selected in order to minimize $\text{tr}\{\boldsymbol{R}(k+1)\}$. From (7), this value is given by

$$\gamma(k) = \frac{\text{tr}\{\boldsymbol{R}(k)\boldsymbol{L}\}}{\text{tr}\{\boldsymbol{L}\boldsymbol{R}(k)\boldsymbol{L} + \sigma_n^2\boldsymbol{D}\}}, \; k = 0, 1, \ldots \tag{9}$$

Together, the coupled recursions (7) and (9) allow the computation of the covariance and stepsize sequences. Since this method just tries to minimize the MSE from one iteration to the next, we will refer to it as *Myopic BLUE* (M-BLUE). The stepsize sequence can be precomputed, provided that the graph topology is known and does not change with time. The case of different communication noise powers in different links can be easily accommodated by replacing $\sigma_n^2\boldsymbol{D}$ by a diagonal matrix such that its i-th diagonal entry is given by the total noise contribution entering the i-th node.

Fig. 1 shows the evolution of the MSE of M-BLUE, given by $\text{tr}\{\boldsymbol{R}(k)\}/M$, for a network of $M = 100$ nodes with a uniform spatial distribution, such that $\Delta(G) = 602$, $\sigma^2 = 1$, and ideal noise-free connections ($\sigma_n^2 = 0$). Also shown is the MSE for a constant stepsize approach using $\gamma = 2/(\lambda_2 + \lambda_M)$, which yields fastest convergence for this class of methods [10]. It is seen that allowing for a time-varing stepsize results in an improved convergence rate. Note that in this example $\gamma(k)$ does not settle down to a constant value, but rather it sustains a period-2 oscillation around the optimal constant stepsize.

IV. Sequence Design

In order for the different nodes to compute the M-BLUE stepsize sequence, knowledge of the network topology (i.e. the Laplacian matrix $\boldsymbol{L}$) is required. In practice, however, this *a priori* information may not be available, and in fact the network topology could change over time. Thus, it is of interest to develop stepsize sequence designs that do not require this knowledge.

Regarding M-BLUE, note that the cost function $\text{tr}\{\boldsymbol{R}(k)\}$ gives equal weight to errors in all sensors (which eventually will agree on the same estimated value), and therefore it implicitly assumes that all nodes play exactly the same role for

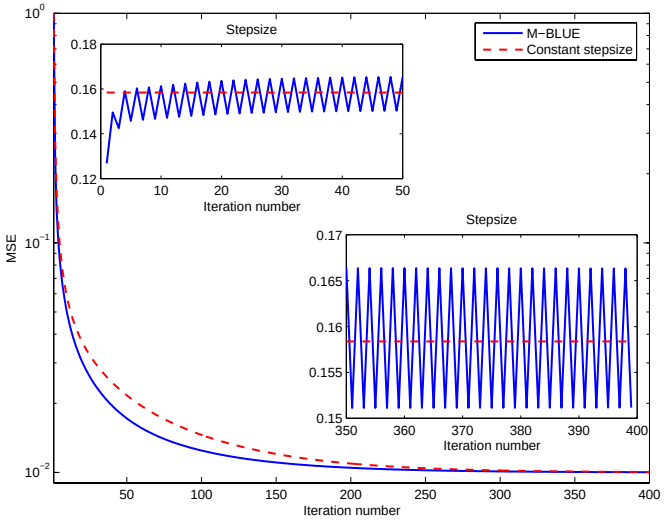

Fig. 1. MSE and stepsize evolution of the consensus iterative processes. Random network of 100 nodes, noiseless case.

recursion guidance purposes. Carrying this argument one step further, one could assume as a first approximation, and lacking any other *a priori* information, that the network topology is location invariant, i.e., all nodes have the same number of neighbors, say d_c, and thus the corresponding Laplacian matrix $\boldsymbol{L}_c$ (but not necessarily the *true* Laplacian $\boldsymbol{L}$) is circulant under proper node numbering. Then $\boldsymbol{L}_c = \boldsymbol{F}\boldsymbol{\Lambda}_c\boldsymbol{F}^H$, with $\boldsymbol{F}$ the orthonormal M-DFT matrix and $\boldsymbol{\Lambda}_c$ a diagonal matrix with the M-point DFT of the first row of $\boldsymbol{L}_c$ on its diagonal. Note that for this circular topology, the corresponding covariance update is obtained by substituting $\boldsymbol{L}$ and $\boldsymbol{D}$ in (7) by $\boldsymbol{L}_c$ and $d_c\boldsymbol{I}$ respectively. From this, and the fact that $\boldsymbol{R}(0) = \sigma^2\boldsymbol{I}$, it is readily seen that this update preserves circularity of the covariance matrices: only the eigenvalues of $\boldsymbol{R}(k)$ change with time, while the eigenvectors are given by the columns of $\boldsymbol{F}$. This results in the following compact form of the M-BLUE recursions for circular topologies:

$$\gamma_{\text{circ}}(k) = \frac{\boldsymbol{\ell}^T\boldsymbol{r}(k)}{(\boldsymbol{\ell}\cdot\boldsymbol{\ell})^T\boldsymbol{r}(k) + d_c\sigma_n^2 M}, \quad (10)$$

$$r_n(k+1) = (1-\gamma_{\text{circ}}(k)\ell_n)^2 r_n(k) + d_c\sigma_n^2\gamma_{\text{circ}}^2(k), \quad (11)$$

for $n = 1,\ldots,M$, $k = 0, 1, \ldots$, where $\boldsymbol{\ell} \doteq [\ell_1 \cdots \ell_M]^T$ and $\boldsymbol{r}(k) = [r_1(k) \cdots r_M(k)]^T$ denote the M-point DFT of the first row of $\boldsymbol{L}_c$ and $\boldsymbol{R}(k)$, respectively, with $\boldsymbol{r}(0) = (\sigma^2/\sqrt{M})\mathbf{1}$, and $\cdot$ denotes the entrywise (Hadamard) product.

The design parameter d_c in this "virtual" circulant network is taken as the closest integer to the average node degree $d \doteq \Delta(G)/M$ of the original network, so the eigenvalues sum in both cases is as close as possible. Then, the circulant Laplacian $\boldsymbol{L}_c$ is determined by its first row $[\boldsymbol{L}_c]_1$, given by

$$[\boldsymbol{L}_c]_1 = (d_c, \overbrace{-1,\cdots,-1}^{d_c/2}, 0, \cdots, 0, \overbrace{-1,\cdots,-1}^{d_c/2}), \quad (12)$$

for d_c even and

$$[\boldsymbol{L}_c]_1 = (d_c, \overbrace{-1,\cdots,-1}^{(d_c-1)/2}, -\tfrac{1}{2}, 0, \cdots, 0, -\tfrac{1}{2}, \overbrace{-1,\cdots,-1}^{(d_c-1)/2}), \quad (13)$$

for d_c odd. The $-\frac{1}{2}$ terms in (13) make it possible to have a circulant symmetric matrix by generalizing Laplacian entries: in a way, they represent connections with weights halved. The vector $\boldsymbol{\ell}$ is then obtained as the M-point DFT of $[\boldsymbol{L}_c]_1$. In addition, for networks with different noise powers in different links, the product $d_c\sigma_n^2$ can be replaced in (10)-(11) by the average noise power entering the nodes.

The case of *random geometric graphs* is of special interest to describe practical deployments in which sensors are randomly scattered across large spatial extensions. If two nodes connect when they fall within range of each other, then the resulting network is described by a *random geometric graph*, whose properties are widely studied in the literature [11]. In such random graphs, the average node degree $d = \Delta(G)/M$, representing the average number of neighbors, can be approximated by the expected degree of any given node, determined by the node spatial distribution and the connectivity radius R. For example, if a large number of nodes is uniformly distributed with density α on a square significantly larger than the connectivity radius, then $d \approx \alpha\pi R^2$ [11]; see, e.g. [12] for results regarding Poisson or Gaussian distributions. Random link failures can also be considered by modifiying accordingly the average number of neighbours. The asymptotic expressions corresponding to (10) and (11), if we let $M \to \infty$ given that sensor networks are usually large, can be readily proved to be

$$\gamma_{\text{circ}}(k) = \frac{\int_0^{2\pi} T(\omega)R^{(k)}(\omega)d\omega}{\int_0^{2\pi} T^2(\omega)R^{(k)}(\omega)d\omega + 2\pi d_c\sigma_n^2}, \quad (14)$$

$$R^{(k+1)}(\omega) = (1-\gamma_{\text{circ}}(k)T(\omega))^2 R^{(k)}(\omega) + d_c\sigma_n^2\gamma_{\text{circ}}^2(k)$$

where $R^{(k)}(\omega)$ denotes the Fourier transform of the first row of $\boldsymbol{R}(k)$, and $T(\omega)$ is the Fourier transform of (12) for d_c even:

$$T(\omega) \doteq d_c - 2\cos(\omega((d_c+2)/4))\sin(\omega d_c/4)/\sin(\omega/2), \quad (15)$$

and (13) for d_c odd:

$$T(\omega) \doteq d_c - 2\cos(\omega((d_c+1)/4))\sin(\omega(d_c-1)/4)/\sin(\omega/2) - \cos(\omega(d_c+1)/2). \quad (16)$$

These asymptotic expressions will be used in the next section for moderate size networks, requiring only the knowledge of d_c, σ^2 and σ_n^2, and using basic numerical integration with rectangles to compute (14).

V. Numerical Illustration

Figs. 2 and 3 depict the MSE evolution for two networks following uniform and Gaussian spatial distributions, respectively, with the signal-to-noise ratio given by $\rho \doteq \sigma^2/\sigma_n^2$. The performance gap between M-BLUE and its circular approximation, which is the price to pay for not using the complete topology information in the stepsize sequence design, is smaller for the uniform case, as could be expected since the number of neighbors has a lower variation throughout the network.

As a final remark, in order to reduce memory needs, [5], [6] and [7] presented some rules which can be generalized and

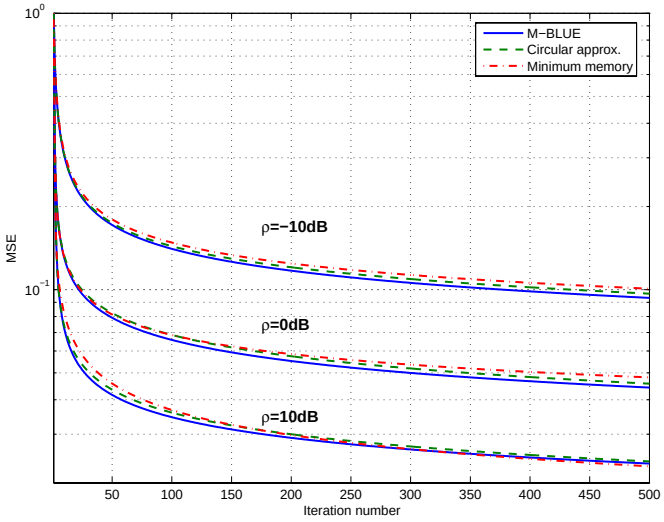

Fig. 2. MSE evolution: $M = 100$ sensors, uniform distribution. The average number of connections per node is 6.

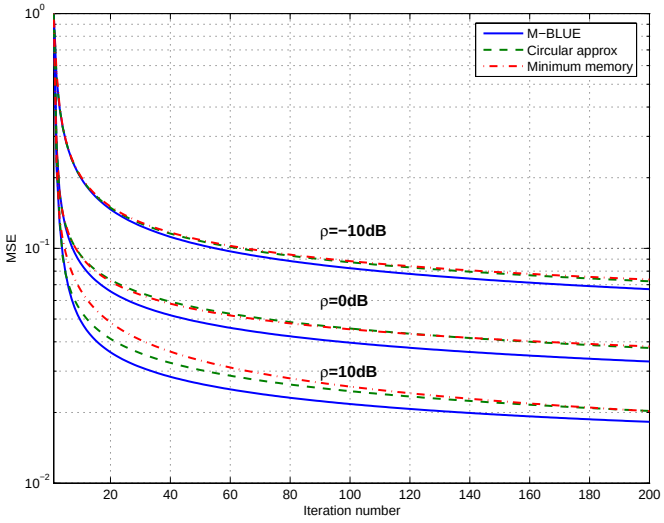

Fig. 3. MSE evolution: $M = 100$ sensors, Gaussian distribution. The average number of connections per node is 12.

jointly described as $\gamma(k) = a/k^b$, with $0.5 < b \leq 1$ in order to satisfy (3). The required memory to store the corresponding sequence would reduce now to the two parameters a and b, although there is no design rule to obtain them. However, for a target number of iterations, one can minimize off-line $||\gamma_{\text{circ}}(k) - a/k^b||^2$ with respect to a and b. Figs. 2 and 3 include also the corresponding performance under the label of *minimum memory* for comparison.

VI. Conclusions

The performance of distributed estimation methods depends on initial estimate accuracy, number of sensors and inter-sensor link qualities. A noise-resilient greedy stepsize sequence design for the iterative consensus algorithm has been presented, together with a modification which uses only average network descriptors and thus does not require full network topology knowledge. In this way, robustness against link failures is also obtained. Future work will address the potential benefits of using different stepsizes at each node in terms of MSE and convergence speed. Note that the M-BLUE design can be readily generalized to this case.

References

[1] R. Olfati-Saber, J. A. Fax, and R. M. Murray, "Consensus and cooperation in networked multi-agent systems," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 215–233, jan 2007.

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

[3] S. Barbarossa and G. Scutari, "Decentralized Maximum Likelihood Estimation for Sensor Networks Composed of Nonlinearly Coupled Dynamical Systems," *IEEE Transactions on Signal Processing*, vol. 55, no. 7, pp. 3456–3470, july 2007.

[4] I. Schizas, A. Ribeiro, and G. B. Giannakis, "Consensus in Ad Hoc WSNs with Noisy Links - Part I: Distributed Estimation of Deterministic Signals," *IEEE Transactions on Signal Processing*, vol. 56, no. 1, pp. 350–364, jan 2008.

[5] Y. Hatano, A. Das, and M. Mesbahi, "Agreement in presence of noise: pseudogradients on random geometric networks," in *Proceedings of the 44th Conference on Decision and Control*, Sevilla, Spain, dec 2005, pp. 6382–6387.

[6] L. Pescosolido, S. Barbarossa, and G. Scutari, "Average consensus algorithms robust against channel noise," in *IEEE 9th Workshop on Signal Processing Advances in Wireless Communications*, Recife, Brazil, july 2008, pp. 261–265.

[7] S. Kar and J. Moura, "A mixed time-scale algorithm for distributed parameter estimation: Nonlinear observation models and imperfect communication," in *Proceedings International Conference on Acoustics, Speech and Signal Processing*, 2009, pp. 3669–3672.

[8] H. Kushner and G. Yin, *Stochastic Approximation and Recursive Algorithms and Appliations*, 2nd ed. Springer, 2003.

[9] S. Kar and J. Moura, "Distributed consensus algorithms in sensor networks with imperfect communication: Link failures and channel noise," *IEEE Transactions on Signal Processing*, vol. 57, no. 1, pp. 355–369, 2009.

[10] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *Systems & Control Letters*, vol. 53, pp. 65–78, 2004.

[11] M. Penrose, *Random Geometric Graphs*. Oxford University Press, 2003.

[12] C. Bettstetter, "On the connectivity of ad hoc networks," *The Computer Journal*, vol. 47, no. 4, pp. 432–447, 2004.