

DISTRIBUTED ESTIMATION: DO NOT TRUST GOSSIPS

*Carlos Mosquera**

Dept. de Teoría de la Señal y Comunicaciones
Universidad de Vigo, 36310 Vigo, Spain
E-mail: mosquera@gts.tsc.uvigo.es

Sudharman K. Jayaweera

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

ABSTRACT

Distributed estimation and detection is of interest for those situations for which the sensor net must achieve an agreement by exchanging information without resorting to the use of an external fusion center. In this paper we deal with the distributed estimation of a parameter for both static and time-varying cases, for which it is important to have similar estimates as accurate as possible. The cooperation is performed in a distributed way to guarantee scalability and robustness to failures, and it is designed to reduce the detrimental effects of the channel noise on the sensor exchanges.

1. INTRODUCTION

Deployment of sensors for monitoring, collaborative information processing and control has gained a considerable attention and research in recent years. If a wireless sensor network can operate autonomously, that is, without a central repository or a fusion center collecting and processing all measurements, important advantages become evident such as scalability and robustness against node failure. The coordinated action of different sensors requires the local exchange of information to improve on their individual estimates. A good deal of research has been done in the early years of this decade, well exposed in [1] and [2], exploiting the appealing mathematical properties of consensus analysis, that is, the

study of algorithms driving all the sensors to a common value. The extension of the basic theory to more realistic scenarios if of paramount importance, for which packet losses, quantization noise or additive channel noise need to be considered to different extent. In this work we focus our attention on the independent noise case, that is, on the role played by the additive noise on the processing of the received values. Since practical sensors impose tight restrictions on battery usage, it is desirable to minimize the energy allocated to coordination and, in particular, to the required exchanges for distributed estimation. We will expose the basic principles of noise resilient schemes for both static and dynamic cases.

2. NETWORK MODEL

We consider a graph $G = (V, E)$, with N nodes (sensors) $v_n \in V$ and edges $e_{ij} \in E$ if there is a path from node v_i to node v_j . The elements of the adjacency matrix \mathcal{A} are defined as $[\mathcal{A}]_{ij} = 1$ if e_{ji} belongs to E , otherwise they are zero. In this study the graphs are undirected ($\mathcal{A} = \mathcal{A}^T$) and connected, so there is a sequence of edges to go from any node i to any other node j . The degree matrix \mathbf{D} is a diagonal matrix such that $[\mathbf{D}]_{ii}$ is equal to the number of connections entering node i . With that, the Laplacian matrix \mathbf{L} is defined as $\mathbf{L} = \mathbf{D} - \mathcal{A}$. In other words, the elements $[\mathbf{L}]_{ij}$ of the Laplacian matrix \mathbf{L} are defined as

$$[\mathbf{L}]_{ij} = \begin{cases} -1, & e_{ji} \in E \\ \mathbf{D}_{ii}, & i = j. \end{cases} \quad (1)$$

*Work performed while at University of New Mexico funded by the Spanish Government under grant JC2007-00250 and MEC project SPROACTIVE, reference TEC2007-68094-C02-01/TCM.

In addition, we define the number of total connections of graph G as $\Delta(G) \doteq \mathbf{1}^T \mathbf{D} \mathbf{1}$. The eigenvalues λ_n of \mathbf{L} contain significant information about the topology of the graph G . In fact, if they are ordered as $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_N$, we have that $\lambda_1 = 0$, and $\lambda_2 > 0$ for a connected graph. This second eigenvalue λ_2 is known as the algebraic connectivity of the graph, and its value plays a major role in the speed at which information can be diffused through the network [2]. The corresponding eigenvectors will be denoted by \mathbf{u}_n , with $\mathbf{u}_1 = \mathbf{1}$. The additive noise in the signal received at the j -th sensor from the i -th sensor is zero-mean with variance σ_w^2 , and is independent for all channel realizations and among different sensor links. The received noise values at the exchange associated with the k th iteration are collected in $\mathbf{W}(k)$:

$$\mathbf{W}(k) = \begin{pmatrix} 0 & w_{12}(k) & \cdots & w_{1N}(k) \\ w_{21}(k) & 0 & \cdots & w_{2N}(k) \\ \vdots & \vdots & \ddots & \vdots \\ w_{N1}(k) & w_{N2}(k) & \cdots & 0 \end{pmatrix}. \quad (2)$$

3. ENTANGLED KALMAN FILTERS

We need to track an autoregressive (AR) process $x(k)$ given by

$$x(k+1) = Fx(k) + u(k)$$

where process noise $u(k)$ is assumed to be white with constant covariance $E\{u(k)u^H(k)\} = Q_u$. In a centralized setting, a unique node collects a set of N noisy observations of $x(k)$ as

$$\mathbf{y}(k) = \mathbf{H}x(k) + \mathbf{v}(k). \quad (3)$$

If the measurement noise $\mathbf{v}(k)$ is white and independent of $\mathbf{u}(k)$, with covariance matrix denoted as \mathbf{Q}_v , then the Kalman filtered estimator, that is, the recursive relation for the update of the linear estimation $\hat{x}(k)$ of $x(k)$ based on the observations $\mathbf{y}(0), \dots, \mathbf{y}(k)$, is given by [3]

$$\hat{\mathbf{x}}(k) = (1 - \mathbf{K}(k)\mathbf{H})F\hat{\mathbf{x}}(k-1) + \mathbf{K}(k)\mathbf{y}(k) \quad (4)$$

with the corresponding recursion for the gain $\mathbf{K}(k)$ detailed in [3]. Now, let us consider that instead of a central observer, we have a set of N sensors, each getting a value $y_n(k)$ of the vector $\mathbf{y}(k)$ in (3):

$$y_n(k) = H_n x(k) + v_n(k), n = 1, \dots, N$$

with H_n and $v_n(k)$ the n th elements of \mathbf{H} and $\mathbf{v}(k)$ respectively. After the update shown in (4), all sensors exchange their estimates before the next measurement. We put all the N estimates at step k in the vectors $\bar{\mathbf{x}}(k)$ (before exchanging information) and $\hat{\mathbf{x}}(k)$ (after exchanging information) respectively:

$$\begin{aligned} \bar{\mathbf{x}}(k) &\doteq [\bar{x}_1(k) \quad \bar{x}_2(k) \quad \cdots \quad \bar{x}_N(k)]^T \\ \hat{\mathbf{x}}(k) &\doteq [\hat{x}_1(k) \quad \hat{x}_2(k) \quad \cdots \quad \hat{x}_N(k)]^T. \end{aligned} \quad (5)$$

The Kalman update and the merging states are given respectively by

$$\bar{x}_n(k) = K_1(k)\hat{x}_n(k-1) + K_2^{(n)}(k)y_n(k), n = 1, \dots, N$$

$$\hat{\mathbf{x}}(k) = \mathbf{A}(k)\bar{\mathbf{x}}(k) + \text{diag}\{\mathbf{A}(k)\mathbf{W}(k)\}$$

where $\text{diag}\{\mathbf{B}\}$ is a column vector collecting the main diagonal elements of the matrix \mathbf{B} , and $\mathbf{A}(k)$ is detailed later in (8). If we group all the sensors coefficients $K_2^{(n)}(k)$ as

$$\mathbf{K}_2(k) \doteq \begin{pmatrix} K_2^{(1)}(k) & 0 & \cdots & 0 \\ 0 & K_2^{(2)}(k) & \ddots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & K_2^{(N)}(k) \end{pmatrix}$$

and define a diagonal matrix \mathbf{H}_d out of \mathbf{H} in (3)

$$\mathbf{H}_d \doteq \begin{pmatrix} H_1 & 0 & \cdots & 0 \\ 0 & H_2 & \ddots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & H_N \end{pmatrix}$$

then it can be proved that the covariance matrix $\mathbf{P}(k)$ of the error $\mathbf{E}(k) = \hat{\mathbf{x}}(k) - x(k)\mathbf{1}$ evolves as

$$\begin{aligned} \mathbf{P}(k) &= K_1^2(k)(\mathbf{I} - \gamma(k)\mathbf{L})\mathbf{P}(k-1)(\mathbf{I} - \gamma(k)\mathbf{L}^T) + Q_u \\ &\cdot (\mathbf{I} - \gamma(k)\mathbf{L})(\mathbf{K}_2(k)\mathbf{H}_d - \mathbf{I})\mathbf{1}\mathbf{1}^T(\mathbf{K}_2(k)\mathbf{H}_d - \mathbf{I})(\mathbf{I} - \gamma(k)\mathbf{L}^T) \\ &+ (\mathbf{I} - \gamma(k)\mathbf{L})\mathbf{K}_2(k)\mathbf{Q}_v\mathbf{K}_2(k)(\mathbf{I} - \gamma(k)\mathbf{L}^T) + \gamma^2(k)\sigma_w^2\mathbf{D}. \end{aligned}$$

with

$$\mathbf{K}_2(k) = \left(1 - \frac{K_1(k)}{F}\right) \mathbf{H}_d^{-1}. \quad (6)$$

At each step k we must find the set of parameters $K_1(k)$, $\mathbf{K}_2(k)$ and $\gamma(k)$ jointly minimizing the trace of $\mathbf{P}(k)$. Although a solution cannot be found in closed-form, and

the corresponding function is not convex, we can apply a Gauss-Seidel iteration to find the solution of the set of non-linear algebraic equations $\nabla \text{tr}\{\mathbf{P}(k)\} = 0$, which has shown to converge in a low number of steps for all tested cases. An off-line computation of this sequence of parameters makes it possible to find the asymptotic values for K_1 , K_2 and γ for a given topology. Thus, local information can diffuse through the network, although there is a gap with respect to the corresponding centralized Kalman filter performance due to the non-ideal exchanges of values among different sensors and the imposed limitations to the communication, which prevents the network from achieving a consensus as to the value of $x(k)$. If a higher rate of exchanges can be applied, then other solutions such as that shown in [7] can improve the performance by doing several rounds of merging between two consecutive Kalman updates.

4. DISTRIBUTED ESTIMATION OF STATIC PARAMETERS

For the static case we have that $H = 1$, $F = 1$ and $Q_u = 0$, and the corresponding weights boil down to $K_1(k) = 1$ and $K_2(k) = \mathbf{0}$. In such a case, we have that the sensors exchange their estimates with non-constant weights:

$$\mathbf{x}(k+1) = \mathbf{A}(k)\mathbf{x}(k) + \text{diag}\{\mathbf{A}(k)\mathbf{W}(k)\}. \quad (7)$$

In the absence of noise, alignment of all the elements of $\mathbf{x}(k)$ can be guaranteed in the limit $k \rightarrow \infty$ with a constant Perron matrix $\mathbf{A} = \mathbf{I} - \gamma\mathbf{L}$, and $0 < \gamma < 1/\max[D_{ii}]$ [2]. In such a case \mathbf{A} has an eigenvalue equal to 1 with the corresponding eigenvector $\mathbf{1}$, and $\lim_{k \rightarrow \infty} \mathbf{A}^k = \mathbf{1}\mathbf{1}^T/N$. This parameterization is especially suited for *circulant topologies*, that is, those for which the Laplacian \mathbf{L} can be expressed as a circulant matrix, since all nodes have the same connections. The presence of additive noise in the exchanges makes this scheme blow up [4], unless the weights decrease. In consequence, we use a time-varying step size parameter $\gamma(k)$ to be discussed later such that the sequence of matrices is written as

$$\mathbf{A}(k) = \mathbf{I} - \gamma(k)\mathbf{L}. \quad (8)$$

As stated in [5], the sequence $\gamma(k)$ needs to be positive and such that

$$\sum_{k=0}^{\infty} \gamma(k) = \infty, \quad \sum_{k=0}^{\infty} \gamma^2(k) < \infty \quad (9)$$

to ensure the asymptotic convergence of $\mathbf{x}(k)$ to a constant vector. If we express the error vector $\mathbf{E}(k) = \mathbf{x}(k) - x\mathbf{1}$ as a function of $\gamma(k)$,

$$\mathbf{E}(k+1) = (\mathbf{I} - \gamma(k)\mathbf{L})\mathbf{E}(k) + \gamma(k)\text{diag}\{\mathbf{A}\mathbf{W}(k)\}$$

then the error covariance $\mathbf{R}(k)$ follows the recursion

$$\mathbf{R}(k+1) = (\mathbf{I} - \gamma(k)\mathbf{L})\mathbf{R}(k)(\mathbf{I} - \gamma(k)\mathbf{L}^T) + \gamma^2(k)\sigma_w^2\mathbf{D}$$

from which

$$\begin{aligned} \mathbf{R}(k+1) &= \mathbf{R}(k) - \gamma(k)\mathbf{R}(k)\mathbf{L}^T - \gamma(k)\mathbf{L}\mathbf{R}(k) \\ &\quad + \gamma^2(k)\mathbf{L}\mathbf{R}(k)\mathbf{L}^T + \gamma^2(k)\sigma_w^2\mathbf{D}. \end{aligned} \quad (10)$$

If consensus is achieved then the covariance matrix $\mathbf{R}(k)$ will asymptotically approach a constant matrix equal to $\sigma_\infty^2\mathbf{1}\mathbf{1}^T$, where σ_∞^2 denotes the limit estimation error. In order to find its value, we write the recursion for the average value of the elements of $\mathbf{R}(k)$:

$$\frac{1}{N^2}\mathbf{1}^T\mathbf{R}(k+1)\mathbf{1} = \frac{1}{N^2}\mathbf{1}^T\mathbf{R}(k)\mathbf{1} + \frac{\Delta(G)}{N^2}\gamma^2(k)\sigma_w^2 \quad (11)$$

where we have used the symmetry of the graph, that is, $\mathbf{L} = \mathbf{L}^T$ and that $\mathbf{L}\mathbf{1} = \mathbf{0}$. In consequence, for a starting covariance matrix $\mathbf{R}(0) = \sigma^2\mathbf{I}$,

$$\frac{1}{N^2}\mathbf{1}^T\mathbf{R}(k+1)\mathbf{1} = \frac{1}{N}\sigma^2 + \frac{\Delta(G)}{N^2}\sigma_w^2 \sum_{j=0}^k \gamma^2(j). \quad (12)$$

In the limit, $\gamma(k)$ gets to zero and $\mathbf{R}(k)$ becomes the constant matrix $\sigma_\infty^2\mathbf{1}\mathbf{1}^T$ with

$$\sigma_\infty^2 = \frac{\sigma^2}{N} + \frac{\Delta(G)}{N^2}\sigma_w^2 \sum_{j=0}^{\infty} \gamma^2(j). \quad (13)$$

The value of $\gamma(k)$ minimizing the trace of the error covariance matrix $\mathbf{R}(k+1)$ at each step can be seen to be given by

$$\gamma(k) = \frac{\text{tr}\{\mathbf{R}(k)\mathbf{L}^T + \mathbf{L}\mathbf{R}(k)\}}{2\text{tr}\{\mathbf{L}\mathbf{R}(k)\mathbf{L}^T + \sigma_w^2\mathbf{D}\}}. \quad (14)$$

and will be used in the simulations in next section.

5. NUMERICAL RESULTS

We have compared the performance of different distributed estimation schemes for the static case: (i) Fixed-weights, first-order (CO-BLUE, [6]); (ii) Fixed-weights,

second-order (RD-BLUE, [7]); (iii) M-BLUE. For analogy reasons, we use the term M-BLUE to describe the myopic strategy jointly described by Equations (7), (8) and (14). In the first two cases, the sensors exchange their internal states, collected in the vector $\Phi(k)$, as

$$\begin{aligned} \Phi(k+1) &= f(\mathbf{x}(0)) + \mathbf{A}_1 \Phi(k) + \mathbf{A}_2 \Phi(k-1) \\ &+ \text{diag}\{\mathbf{A}_1 \mathbf{W}(k) + \mathbf{A}_2 \mathbf{W}(k-1)\} \end{aligned}$$

and compute the change in their states to obtain the estimates

$$\mathbf{x}(k) = \Phi(k) - \Phi(k-1) \quad (15)$$

with $\mathbf{A}_2 = 0$ in the first-order case. The weights of \mathbf{A}_1 and \mathbf{A}_2 have been chosen to yield the same asymptotic error, whereas the function $f(\mathbf{x}(0))$ must guarantee the unbiasedness of the estimates. We have run 10^3 realizations in a fully connected network with $N = 10$ sensors, for two different values of $\rho \doteq \sigma^2/\sigma_w^2$. The results are included in Figure 1, which shows how the convergence speed depends on the level of additive noise as expected.

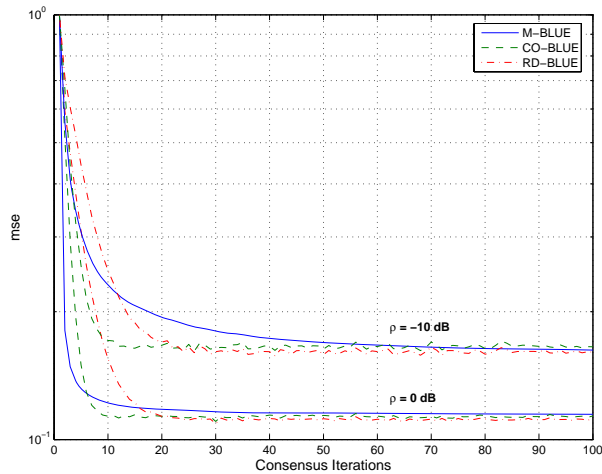


Fig. 1. Average stationary mean square error (mse) performance for a network of 10 sensors.

6. CONCLUDING REMARKS

We have shown how different sensors can follow the evolution of a common parameter by sharing their successive estimates through noisy exchanges. The sensors share their estimates within the limits of the network topology, in an attempt to improve the global performance. A consensus strategy has also been presented

for the static scenario, for which sensors are expected to achieve a common estimate after a series of exchanges. In both cases there is a trade-off between the need to share information to improve the estimates accuracy and the quality of the noisy links.

7. REFERENCES

- [1] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, “Randomized gossip algorithms,” *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2508–2530, jun 2006.
- [2] Reza Olfati-Saber, J. Alex Fax, and Richard M. Murray, “Consensus and cooperation in networked multi-agent systems,” *Proceedings of the IEEE*, vol. 95, no. 1, pp. 215–233, jan 2007.
- [3] Thomas Kailath, Ali H. Sayed, and Babak Hassibi, *Linear Estimation*, Prentice Hall, 2000.
- [4] Lin Xiao, Stephen Boyd, and Seung-Jean Kim, “Distributed average consensus with least-mean-square deviation,” *Journal of Parallel and Distributed Computing*, vol. 67, no. 1, pp. 33–46, jan 2007.
- [5] S. Kar and J.M.F. Moura, “Distributed consensus algorithms in sensor networks with communication channel noise and random link failures,” in *Proceedings of the 41st Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, USA, nov 2007.
- [6] 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.
- [7] S.I. Roumeliotis I.D. Schizas, G.B. Giannakis and A. Ribeiro, “Anytime Optimal Distributed Kalman Filtering and Smoothing,” in *Proceedings 14th Workshop on Statistical Signal Processing*, Madison, WI, USA, aug 2007, pp. 368–372.