VI. Conclusion

The sampled-data design of dual filter banks for dual-tree complex wavelet transforms is studied in this correspondence. The present work improves the sampled-data design approach of [15] in two aspects: 1) it directly gives FIR solutions, and 2) the solutions possess prescribed numbers of zeros at $z = -1$. We achieve the first improvement by reducing the design as an LMI optimization problem; the second is guaranteed by introducing a new filter in the sampled-data system configuration. By arranging the coefficients of the unknown filter in the sampled-data system accordingly, the presented design procedure can be used to find symmetric solutions.

References

[1] A. Baradarani and R. Yu, "A dual-tree complex wavelet with application in image denoising," in *Proc. 2007 IEEE Int. Conf. Signal Processing Communications*, Dubai, UAE, Nov. 2007, pp. 1203–1206.

[2] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Imaging coding using wavelet transform," *IEEE Trans. Image Process.*, vol. 1, no. 2, pp. 205–220, Apr. 1992.

[3] S. Boyd, L. E. Ghaoui, E. Feron, and V. Balakrishnan, *Linear Matrix Inequalities in System and Control Theory*. Philadelphia, PA: SIAM, 1994.

[4] T. Chen and B. A. Francis, *Optimal Sampled-data Control Systems*. New York: Springer-Verlag, 1995.

[5] N. G. Kingsbury, "Design of Q-shift complex wavelets for image processing using frequency domain energy minimization," in *Proc. IEEE Int. Conf. Image Processing*, Barcelona, Spain, Sep. 2003, vol. 1, pp. 1013–1016.

[6] S. Mallat, *A Wavelet Tour of Signal Processing*. New York: Academic, 1999.

[7] M. Nagahara and Y. Yamamoto, "Optimal design of fractional delay filters," in *Proc. 42nd IEEE Conf. Decision Control*, Dec. 2003, vol. 6, pp. 6539–6544.

[8] M. Nagahara and Y. Yamamoto, "Optimal design of fractional delay FIR filters without band-limiting assumption," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, Philadelphia, PA, Mar. 2005, vol. 4, pp. 221–224.

[9] I. W. Selesnick, "Hilbert transform pairs of wavelet bases," *IEEE Signal Process. Lett.*, vol. 8, no. 6, pp. 170–173, Jun. 2001.

[10] I. W. Selesnick, "The design of approximate Hilbert transform pairs of wavelet bases," *IEEE Trans. Signal Process.*, vol. 50, no. 5, pp. 1144–1152, May 2002.

[11] I. W. Selesnick, R. G. Baraniuk, and N. G. Kingsbury, "The dual-tree complex wavelet transform—A coherent framework for multiscale signal and image processing," *IEEE Signal Process. Mag.*, vol. 6, pp. 123–151, Nov. 2005.

[12] D. B. H. Tay and M. Palaniswami, "Design of approximate Hilbert transform pair of wavelets with exact symmetry," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Montreal, QC, Canada, May 2004, vol. 2, pp. 921–924.

[13] P. P. Vaidyanathan, "The discrete-time bounded-real lemma in digital filtering," *IEEE Trans. Circuit Syst.*, vol. 32, no. 9, pp. 918–924, Sep. 1985.

[14] C. F. Van Loan, "Computing integrals involving the matrix exponential," *IEEE Trans. Autom. Control*, vol. 23, no. 1, pp. 395–404, Jan. 1978.

[15] R. Yu, "Characterization and sampled-data design of dual-tree filter banks for Hilbert transform pairs of wavelet bases," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2458–2471, Jun. 2007.

[16] R. Yu and A. Baradarani, "Design of Hilbert transform pairs of wavelet bases via LMI optimization," in *Proc. 17th Int. Symp. Mathematical Theory Networks Systems*, Kyoto, Japan, Jul. 2006, pp. 1793–1796.

[17] R. Yu and H. Ozkaramanli, "Hilbert transform pairs of orthogonal wavelet bases: Necessary and sufficient conditions," *IEEE Trans. Signal Process.*, vol. 53, no. 12, pp. 4723–4725, Dec. 2005.

[18] R. Yu and H. Ozkaramanli, "Hilbert transform pairs of biorthogonal wavelet bases," *IEEE Trans. Signal Process.*, vol. 54, no. 6, pp. 2119–2125, Jun. 2006.

Enforcing Consensus While Monitoring the Environment in Wireless Sensor Networks

Paolo Braca, Stefano Marano, and Vincenzo Matta

***Abstract*—In many environmental monitoring applications of Wireless Sensor Networks (WSNs), safe information retrieval from any subset of sensors, at an arbitrary instant of time, should be guaranteed. Accordingly, we study the behavior of a WSN that continuously senses the surrounding environment, while consensus among its nodes is simultaneously enforced. For this *running consensus* scheme, analytical bounds in terms of consensus degree and compsarison with an ideal centralized system are provided, and example of applications are presented.**

***Index Terms*—Gossip algorithms, pairwise averaging, running consensus, wireless sensor networks.**

I. Introduction

The theory of consensus has a long history, especially in fields like distributed computation, multiagent systems, and cooperative control, see, e.g., [1]–[3]. With the advent of wireless sensor networks (WSNs) there has been a growing interest in fully decentralized systems—i.e., networks without fusion center—in which the nodes (as synonymous of sensors) reach agreement about a common value in a cooperative and decentralized fashion. Investigations in this direction can be traced back to [4] and, even earlier, to [5]. Useful entry points to the pertinent literature can be found in [6]–[14]; it is also worth mentioning the emerging approach inspired to physical (e.g., coupled oscillators) or biological systems, see [15] and references therein.

In the current literature, the assumption is usually made that the network senses the environment *before* implementing the consensus algorithm. In some applications, however, it is desirable that the sensing task is performed *while* some form of consensus is simultaneously enforced (*running consensus*). This may be the case, for instance, of sensor networks operating in dangerous environments in which a significant number of nodes can be suddenly impaired. Since usually it is not known when the sensors will be damaged (and which sensors will be), it is difficult to decide when the sensing task should be terminated to begin the consensus phase. Therefore, the running consensus scheme may represent a valid alternative, being conceived to allow information retrieval from any subset of nodes at any instant of time.

The nodes of the network sense the surrounding environment by collecting observations, here modeled as random variables. With reference to the so-called *gossip* algorithms [7], we assume that such observations are averaged over different sensors with the final aim of making available, to any node, the arithmetic mean of all the observations globally collected by the entire network. Different from the classical consensus schemes, however, this arithmetic mean is time varying due to the new measurements collected by the network while the averaging process is carried on. Thus, two opposite tendencies can be identified. From one hand, the averaging cuts down the differences between the sensors (consensus). On the other hand, due to the new observations, a form of diversity among the nodes (namely, something opposite to the consensus) is enforced, as well. Which of the two effects will prevail?

Manuscript received June 14, 2007; revised December 31, 2007. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Erchin Serpedin.

The authors are with the Department of Information and Electrical Engineering (DIIIE), University of Salerno, 84084, Fisciano (SA), Italy (e-mail: pbraca@unisa.it; marano@unisa.it; vmatta@unisa.it).

Digital Object Identifier 10.1109/TSP.2008.917855

Being the first investigation of the subject, we abstract the problem: no attention is paid to practical aspects including synchronism issues (this is deeply discussed, e.g., in [7]), the optimization of the energy consumption, the communications protocols, errors, failures, and so forth. Also the independence among different observations (see below) can be considered a simplifying assumption.

The remainder of the paper is so organized. The problem is precisely formalized in Section II, and the main results are presented in Section III. Examples of applications are provided in Section IV, while in Section V we summarize.

II. Problem Formalization

We consider a slotted system where in the tth time slot each sensor of the network collects a new measurement, and these measurements are modeled as iid (independent, identically distributed) random variables, with statistical mean μ and with variance σ^2. The common probability distribution of these variables is arbitrary and plays no role in this work.[1] The collected random values are stored for successive processing; the value stored by a node is hereafter also referred to as the *state* of that sensor. We denote by $x(t)$ the column vector whose entries are the measurements collected by the sensors at time t, and whose size is n, the number of nodes in the network.

Let $s(t)$ be the n-vector whose entries represent the state of the nodes at the tth time slot. We assume that this (random) state vector is updated as follows:

$$s(t) = \frac{t-1}{t} W(t) s(t-1) + \frac{x(t)}{t}. \tag{1}$$

The matrices $W(t)$, $t = 1, 2, \ldots$, are assumed iid and doubly stochastic, so that the connection protocol among nodes (formalized by the product $W(t)s(t-1)$ in the above equation) amounts to a weighted average of the nodes' states [7]. The $W(t)$s are also assumed statistically independent of the sensors' observations. Note that the first term on the right-hand side (RHS) of (1) enforces consensus among the nodes, while the second accounts for the new measurements. These two terms are weighted by $(t-1)/t$ and $1/t$, respectively, a choice that will be justified shortly [see discussion following (3)].

The basic assumption made throughout this work is that the statistical average $E[W(t)W^T(t)]$ (which is obviously doubly stochastic as well) has unitary eigenvalue with algebraic multiplicity 1, see also [7]. Due to the identical distribution with respect to the time slot t, to simplify the notation, in the following we write $E[WW^T]$ for $E[W(t)W^T(t)]$.

The updating (1) can be also recast as

$$s(t) = \frac{1}{t} \sum_{h=1}^{t} \phi(t,h) x(h) \tag{2}$$

where $\phi(t,h)$ is defined as the product of $W(t)$ matrices as follows:

$$\phi(t,h) := \begin{cases} I, & \text{if } h = t, \\ W(t)W(t-1)\cdots W(h+1), & \text{if } h < t \end{cases} \tag{3}$$

with I denoting the identity matrix. The properties of $\phi(t,h)$ are investigated in [7], here we only stress that $\phi(t,h)$ is doubly stochastic, being the product of doubly stochastic matrices.

As a consequence of the weights chosen in the updating (1), from (2) we immediately recognize that the statistical average of the network state stays constant with t. Indeed, denoting with $\mathbf{1}$ the vector of all ones, we have $E[s(t)] = E[\sum_{h=1}^{t} \phi(t,h)x(h)/t] = E[\sum_{h=1}^{t} \phi(t,h)\mathbf{1}\mu/t]$. Since $\phi(t,h)$ is doubly stochastic, the same property holds for its statistical expectation, thus implying that $E[\phi(t,h)] = \mathbf{1}$, yielding $E[s(t)] = E[x(t)] = \mathbf{1}\mu$. The mean value of the state vector is a constant vector whose entries coincide with the statistical average of the observations and this allows us to assume hereafter $\mu = 0$ without any loss of generality, and motivates the following definition of consensus.

[1]Actually, the approach could be extended to observations with only the first two moments equal, and the assumption of independence could be relaxed to that of uncorrelatness; in the following we only refer to iid variables for simplicity.

A. Definition of Running Consensus

Let $C(t) = E[s(t)s^T(t)]$ be the covariance matrix of the state vector. Denoting by $C_{ij}(t)$ its entries, we define

$$\rho_c(t) := \frac{C_{ij}(t)}{\sqrt{C_{ii}(t)\, C_{jj}(t)}}$$

and

$$\rho_e(t) := \frac{\sqrt{C_{ii}(t)\, C_{jj}(t)}}{\frac{C_{ii}(t)+C_{jj}(t)}{2}}.$$

The first quantity is the standard statistical correlation coefficient, and the second is the ratio between the geometric and the arithmetic mean of two diagonal entries of the matrix $C(t)$.

Definition: The consensus coefficient between the nodes i and j is defined as

$$\rho(t) := \rho_c(t)\rho_e(t) = \frac{2C_{ij}(t)}{C_{ii}(t) + C_{jj}(t)}. \tag{4}$$

Note that the dependence upon (i,j) is omitted for notational convenience. The rationale of the above definition is now explained. As well known, $\rho_c(t)$ quantifies the degree of statistical dependence between random variables (state of the nodes, in our case), and attains its maximum value of 1 only if the variables are linearly dependent almost everywhere (a.e.). On the other hand, $\rho_e(t)$ is an index of equality between two positive numbers (here variances), and takes the value 1 only when these numbers are equal. Consequently, the product $\rho(t) = \rho_c(t)\,\rho_e(t)$ belongs to $(-1,1)$ and the value 1 is attained only when the two random variables coincide a.e.: having zero-mean and being linearly related ($\rho_c(t) = 1$) they can only differ for a scale factor, which must by unitary because the variables share the same variance ($\rho_e(t) = 1$). This legitimates the adoption of $\rho(t)$ as a quantitative measure of the consensus degree: when $\rho(t) = 1$ the state of the two nodes is a.e. identical.

While a unitary consensus coefficient for all the pair of nodes would mean that all nodes share the same state (a.e.), it is also necessary to check that such state is that desired. In the next section we elaborate on this issue.

B. Comparison With the Optimal Centralized Scheme

If all the observations collected by the entire network up to time slot t were simultaneously available to a single device, this latter could compute the following arithmetic mean, which represents the state of an ideal centralized system:

$$\eta_{\mathrm{ctr}}(t) := \frac{\sum_{h=1}^{t} \mathbf{1}^T x(h)}{nt} = \frac{\mathbf{1}^T s(t)}{n}$$

where the equality follows from (2), since $\mathbf{1}^T\phi(t,h) = \mathbf{1}^T$.

Computing $\eta_{\mathrm{ctr}}(t)$ turns out to be the main goal of many inference problems in WSNs where the optimal decision/estimation statistic is obtained by averaging the network observations, or functions thereof, see, e.g., [15]. Accordingly, in our fully decentralized architecture, the ideal goal would be that any node attained the same performances of the centralized scheme.

To this aim, let us refer to the mean square error. As to the statistical mean, $E[\eta_{\mathrm{ctr}}(t)]$ is just μ (that we have set to zero for simplicity), and

being $E[s(t)] = \mathbf{1}\mu$, we recognize that the state of all the nodes takes the same mean value of the optimal centralized scheme, at any time t. Consequently, all the nodes should hopefully share the same *variance* of the centralized scheme $\sigma^2_{\rm ctr}(t) := \sigma^2/(tn)$. Accordingly, we introduce the quantity

$$\gamma(t) = \frac{C_{ii}(t)}{\sigma^2_{\rm ctr}(t)} \tag{5}$$

and the condition of approaching the *optimal* centralized performance is met, for a generic node i, when $\gamma(t) \to 1$. We stress that the above performance index does depend upon i, but such dependence is not made explicit.

The sought connections between the state vector $s(t)$ and the state of the centralized device, regarded for convenience as an n-vector with equal entries, can be conveniently expressed by the error term:

$$e(t) := s(t) - \mathbf{1}\eta_{\rm crt}(t) = \left(I - \frac{\mathbf{1}\mathbf{1}^T}{n}\right) s(t). \tag{6}$$

Note that $\mathbf{1}\mathbf{1}^T/n$ in (6) is the matrix that projects vectors onto the subspace of equal-entry vectors, and $I - \mathbf{1}\mathbf{1}^T/n$ is the projection matrix for the subspace orthogonal to that. Now, projecting $s(t)$ over the subspace spanned by the columns of $I - \mathbf{1}\mathbf{1}^T/n$, we get an updating equation for $e(t)$ similar to (2). Indeed, from (1) and (6), we can write

$$\begin{aligned} e(t) &= \frac{t-1}{t}\left(I - \frac{\mathbf{1}\mathbf{1}^T}{n}\right) W(t) s(t-1) \\ &\quad + \frac{1}{t}\underbrace{\left(I - \frac{\mathbf{1}\mathbf{1}^T}{n}\right) x(t)}_{:=\widetilde{x}(t)} \\ &= \frac{t-1}{t}\underbrace{\left(W(t) - \frac{\mathbf{1}\mathbf{1}^T}{n}\right)}_{:=\widetilde{W}(t)}\underbrace{\left(I - \frac{\mathbf{1}\mathbf{1}^T}{n}\right) s(t-1)}_{e(t-1)} + \frac{1}{t}\widetilde{x}(t) \end{aligned}$$

where we used $(I - \mathbf{1}\mathbf{1}^T/n)W(t) = (W(t) - \mathbf{1}\mathbf{1}^T/n)(I - \mathbf{1}\mathbf{1}^T/n)$, that follows from the equality $W(t)\mathbf{1}\mathbf{1}^T = \mathbf{1}\mathbf{1}^T$.

The error can be thus rewritten as

$$e(t) = \frac{1}{t}\sum_{h=1}^{t} \widetilde{\phi}(t,h)\widetilde{x}(h) \tag{7}$$

where $\widetilde{\phi}(t,h)$ is defined as the product of $\widetilde{W}(t)$s matrices, exactly as $\phi(t,h)$ was defined in terms of the $W(t)$s, see (3). The covariance matrix can be expressed as

$$\begin{aligned} C(t) &= E\left[\left(e(t) + \frac{\mathbf{1}\mathbf{1}^T}{n}s(t)\right)\left(e^T(t) + s^T(t)\frac{\mathbf{1}\mathbf{1}^T}{n}\right)\right] \\ &= E[e(t)e^T(t)] + \sigma^2_{\rm ctr}(t)\mathbf{1}\mathbf{1}^T \end{aligned} \tag{8}$$

where the last equality follows from straightforward algebra.

We see that $C(t)$ is the sum of the covariance matrix of the error term $e(t)$ plus the covariance matrix of the centralized system. The above relationships will be exploited to demonstrate the Proposition in the next section.

III. Main Results

We are now ready to characterize the performances of the proposed running consensus scheme. Recall that the largest eigenvalue of $E[WW^T]$ is 1; let us denote by λ_U the second largest eigenvalue of that matrix, and by λ_L the smallest. Let us also introduce the quantities

$$\psi_U(t) := \frac{1}{t}\frac{1-\lambda_U^t}{1-\lambda_U}$$

and

$$\psi_L(t) := \frac{1}{t}\frac{1-\lambda_L^t}{1-\lambda_L}.$$

Proposition: The following bounds hold:

$$(n-1)\psi_L(t) \le \gamma(t) - 1 \le (n-1)\psi_U(t) \tag{9}$$

$$\frac{n\psi_L(t)}{1+(n-1)\psi_U(t)} \le 1 - \rho(t) \le \frac{n\psi_U(t)}{1+(n-1)\psi_L(t)}. \tag{10}$$

■

Note that $\rho(t)$ and $\gamma(t)$ depend upon the sensor indices i and j, while the bounds do not: the above relationships hold for any choice of i and j.

Proof: Let us start from computing the term $E[e(t)e^T(t)]$ appearing in (8). According to (7) and recalling that the matrices $\widetilde{W}(t)$s are iid and independent of the sensors' observations, we have

$$\begin{aligned} E[e(t)e^T(t)] &= \frac{1}{t^2}\sum_{h=1}^{t} E[\widetilde{\phi}(t,h)E[\widetilde{x}(h)\widetilde{x}^T(h)]\widetilde{\phi}^T(t,h)] \\ &= \frac{\sigma^2}{t^2}\sum_{h=1}^{t} E[\widetilde{\phi}(t,h)\widetilde{\phi}^T(t,h)] \\ &:= \frac{\sigma^2}{t^2}\sum_{h=1}^{t} \widetilde{\Phi}(t-h) \end{aligned} \tag{11}$$

having exploited $E[\widetilde{x}(h)\widetilde{x}^T(h)] = (I - \mathbf{1}\mathbf{1}^T/n)\sigma^2$.

To elaborate, we need the following results. First, recalling the definition of $\widetilde{\phi}(t,h)$, we have

$$\begin{aligned} \widetilde{\Phi}(t-h) &= E[\widetilde{W}(t)\cdots\widetilde{W}(h+2)\widetilde{W}(h+1) \\ &\qquad \times \widetilde{W}^T(h+1)\widetilde{W}^T(h+2)\cdots\widetilde{W}^T(t)] \\ &\overset{(i)}{=} E[\widetilde{W}(1)\cdots\widetilde{W}(t-h-1)\widetilde{W}(t-h) \\ &\qquad \times \widetilde{W}^T(t-h)\widetilde{W}^T(t-h-1)\cdots\widetilde{W}^T(1)] \\ &\overset{(ii)}{=} E[\widetilde{W}(1)\cdots\widetilde{W}(t-h-1)E[\widetilde{W}(t-h) \\ &\qquad \times \widetilde{W}^T(t-h)]\widetilde{W}^T(t-h-1)\cdots\widetilde{W}^T(1)] \\ &\overset{(iii)}{=} E[\widetilde{W}(1)\cdots\widetilde{W}(t-h-1)E[\widetilde{W}\widetilde{W}^T] \\ &\qquad \times \widetilde{W}^T(t-h-1)\cdots\widetilde{W}^T(1)]. \end{aligned} \tag{12}$$

In the above: (i) follows from the independence of the matrices $W(t)$s: for any selection of $t-h$ different time indices, the statistical expectation is the same; (ii) is a straightforward application of the iteration property of conditional expectation [16], that also exploits again the independence of the $\widetilde{W}(t)$s; in (iii) we only simplify the notation since the inner expectation is independent upon the time index.

Furthermore, let us consider an arbitrary (deterministic) vector, say $y(1)$, orthogonal to $\mathbf{1}$, and define $y(2) := \widetilde{W}^T(1)y(1)$, $y(3) := \widetilde{W}^T(2)y(2), \ldots, y(t-h) := \widetilde{W}^T(t-h-1)y(t-h-1)$. All these vectors still remain orthogonal to $\mathbf{1}$, and we have the following properties: for $k = 2, \ldots, t-h$

$$E\left[\|y(k)\|^2\right] = E[y^T(k-1)E[\widetilde{W}\widetilde{W}^T]y(k-1)] \tag{13}$$

and, for $k = 1, \ldots, t-h$

$$\lambda_L E[\|y(k)\|^2] \le E\left[y^T(k)E[\widetilde{W}\widetilde{W}^T]y(k)\right] \le \lambda_U E[\|y(k)\|^2]. \tag{14}$$

Equation (13) is obvious. As to (14), note first that the matrix $E[\widetilde{W}\widetilde{W}^T]$ has all its eigenvalues equal to those of $E[WW^T]$, but for the eigenvalue 1, which is replaced by 0. Thus, the maximum eigenvalue of $E[\widetilde{W}\widetilde{W}^T]$ is λ_U, whence the upper bound in (14). Then, since $E[\widetilde{W}\widetilde{W}^T]$ has a zero eigenvalue with eigenvector $\mathbf{1}$ and recalling that $y(k)\perp\mathbf{1}$, we can lower bound the quadratic form with the successive eigenvalue of $E[\widetilde{W}\widetilde{W}^T]$ (which is not necessarily nonzero), namely with λ_L, see, e.g., [17, Th. 4.2.2].

Armed with the above results, we can write

$$\begin{aligned}
&y^T(1)\widetilde{\Phi}(t-h)y(1)\\
&\quad\overset{(i)}{=}\ y^T(1)E[\widetilde{W}(1)\cdots\widetilde{W}(t-h-1)E[\widetilde{W}\widetilde{W}^T]\\
&\qquad\widetilde{W}^T(t-h-1)\cdots\widetilde{W}^T(1)]y(1)\\
&\quad\overset{(ii)}{=}E[y^T(2)E[\widetilde{W}(2)\cdots\widetilde{W}(t-h-1)E[\widetilde{W}\widetilde{W}^T]\\
&\qquad\widetilde{W}^T(t-h-1)\cdots\widetilde{W}^T(2)]y(2)]\\
&\qquad\cdots\\
&\quad=E[y^T(t-h)E[\widetilde{W}\widetilde{W}^T]y(t-h)]\\
&\quad\overset{(iii)}{\leq}\lambda_U E[\|y(t-h)\|^2]\\
&\quad\overset{(iv)}{=}\lambda_U E\left[y^T(t-h-1)E[\widetilde{W}\widetilde{W}^T]y(t-h-1)\right]\\
&\quad\leq\lambda_U^2 E\left[\|y(t-h-1)\|^2\right]\\
&\qquad\cdots\\
&\quad=\lambda_U^{t-h}\|y(1)\|^2.
\end{aligned}$$

In the above, (i) is simply (12); (ii) and the successive equalities result form the definition of the sequence $y(k)$; the upper bound in (14) implies the inequality (iii); (iv) is a consequence of (13); the remainder of the chain is obtained by repeating the last two steps.

A similar result is obtained by using the lower bound in (14), instead of the upper bound. Thus, denoting $y(1)$ simply by y, one obtains:

$$\lambda_L^{t-h}\|y\|^2\leq y^T\widetilde{\Phi}(t-h)y\leq\lambda_U^{t-h}\|y\|^2. \tag{15}$$

As y is an arbitrary vector orthogonal to $\mathbf{1}$, we can set $y=e_i-\mathbf{1}/n$ in (15), where e_i denotes a vector of zeros with only the ith entry equal to 1. This yields

$$\lambda_L^{t-h}(n-1)/n\leq\widetilde{\Phi}_{ii}(t-h)\leq\lambda_U^{t-h}(n-1)/n.$$

Now, as seen in (11) the diagonal entries $C_{ii}(t)$ of the covariance matrix in (8) involve just the terms $\widetilde{\Phi}_{ii}(t-h)$, which can be bounded as above. In this way, after straightforward algebra, we finally get

$$\begin{aligned}
\sigma^2_{\mathrm{ctr}}(t)[1+(n-1)\psi_L(t)]\leq C_{ii}(t)\\
\leq\sigma^2_{\mathrm{ctr}}(t)[1+(n-1)\psi_U(t)]
\end{aligned} \tag{16}$$

and the claim in (9) immediately follows.

Similarly, setting $y=e_i-e_j(i\neq j)$ in (15), yields

$$\lambda_L^{t-h}\leq\frac{\widetilde{\Phi}_{ii}(t-h)+\widetilde{\Phi}_{jj}(t-h)}{2}-\widetilde{\Phi}_{ij}(t-h)\leq\lambda_U^{t-h}$$

from which we obtain

$$n\sigma^2_{\mathrm{ctr}}(t)\psi_L(t)\leq\frac{C_{ii}(t)+C_{jj}(t)}{2}-C_{ij}(t)\leq n\sigma^2_{\mathrm{ctr}}(t)\psi_U(t).$$

Recalling the definition of $\rho(t)$, and further using (16), the above immediately gives (10). The proof is now complete. $\triangle$

The Proposition proved in this section deserves some comments.

- Since $\psi_{U,L}(t)\to 0$ when $t\to\infty$, we have that $\rho(t)\to 1$, and $\gamma(t)\to 1$: asymptotically, the consensus is reached, and the performance of the optimal centralized system is attained. Furthermore, $\psi_{U,L}(t)$ both go to zero essentially as t^{-1}, implying the same speed of convergence to 1 for $\rho(t)$ and $\gamma(t)$.
- For large t, we have $\lambda_{U,L}^t\ll 1$. Then, confusing $n-1$ with n, which is the case of interest, we obtain the approximate bounds

$$\frac{n}{t}\frac{1}{1-\lambda_L}\leq\epsilon(t)\leq\frac{n}{t}\frac{1}{1-\lambda_U} \tag{17}$$

  where $\epsilon(t)$ is a compact notation for both the error figures $\gamma(t)-1$ and $1-\rho(t)$. Note that the bounds on $\epsilon(t)$ depend upon the normalized time t/n, and large values of n slow down the convergence, as one might expect. Note also that n/t seems like the rate at which the new observations become negligible to the average made up to time t, and the system performance seems hence to be dominated by such rate.
- The upper bound in the previous equation gives a conservative (worst case) estimate of the rate of convergence, namely

$$r:=\lim_{t\to\infty}t\epsilon(t)=\frac{n}{1-\lambda_U}. \tag{18}$$

- Useful insights about the differences with classical consensus algorithms can be gained by allowing multiple averaging steps per time slot, namely, by assuming that the weighting matrix $W(t)$ in (1) is actually the product of $v>1$ iid doubly stochastic matrices, say $W(t)=M_1(t)\cdots M_v(t)$. By a straightforward modification of the previous proof, we can obtain bounds like those in the Proposition, in terms of the smallest and the second largest eigenvalues of $E[MM^T]$, say $\xi_L\geq 0$ and $\xi_U<1$: the final result amounts to substitute in the claims of the Proposition $\lambda_{U,L}$ with $\xi_{U,L}^v$, respectively. In the classical consensus scenario, see e.g., [7], v diverges so that $\xi_{U,L}^v$ become negligible with exponential rate.
- In the same spirit of the previous item, one may also refer to the case that each sensor just spends time t to gather data and then exchange the locally averaged data using classic algorithms [7] which reach consensus in a time negligible compared to the data gathering time t. Of course, this scheme does not take into account possible sensors' failures occurring before the data exchange process, which is the main motivation of our analysis.
- A distinct feature of the running consensus scheme is the speed of convergence of the performance figures. In fact, this is substantially different from the exponential law that governs the classic consensus algorithms [7]. Furthermore, in our setup, the specific network topology/connectivity (which rule the system eigenvalues) is less crucial with respect to the classical case. Indeed, in this latter, a network design yielding a larger value of ξ_U is expected to *exponentially* outperform a system with a smaller eigenvalue. Oppositely, in our case, the *universal* scaling law is t^{-1}, and only the value of the rate coefficient can be tuned by the eigenvalues of the system.

IV. Examples: Pairwise Averaging

A popular protocol prescribes that at any time slot a single pair of nodes exchanges their data, leading to the so-called pairwise algorithm. In this case $W(t)$ takes the form (borrowing notations from [7])

$$W_{ij}=I-\frac{(e_i-e_j)(e_i-e_j)^T}{2}. \tag{19}$$

Here the randomly selected pair is just (i,j), and in fact multiplying W_{ij} by a state vector amounts to replace the states i and j with their arithmetic mean. In this section we limit the analysis to a simple pairwise protocol. Specifically, we make reference to the network topologies schematically depicted n Fig. 1, and we assume that only the *admissible* pairs of nodes (those connected by straight lines) can be selected for the pairwise averaging. Any such pair is selected with one

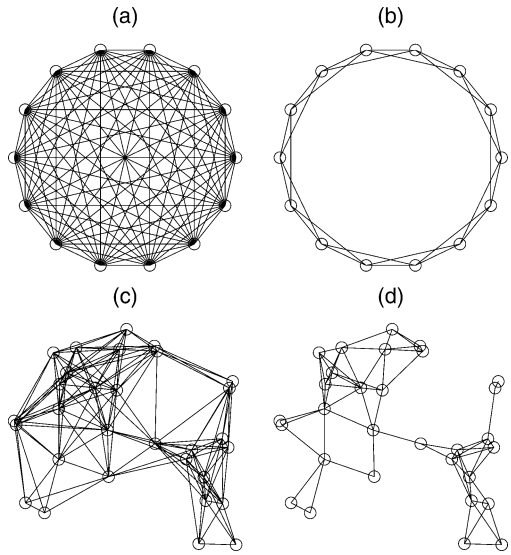

Fig. 1. Architectures of the networks used in the examples. Panel (a) represents a completely connected ring of $n = 15$ sensors. The same ring is considered in panel (b) where each node is connected only to four neighbors. In panels (c) and (d), which refer to $n = 30$, the node position is that typical of randomly deployed sensors, with the network in (c) having a larger number of admissible pairs than that depicted in (d).

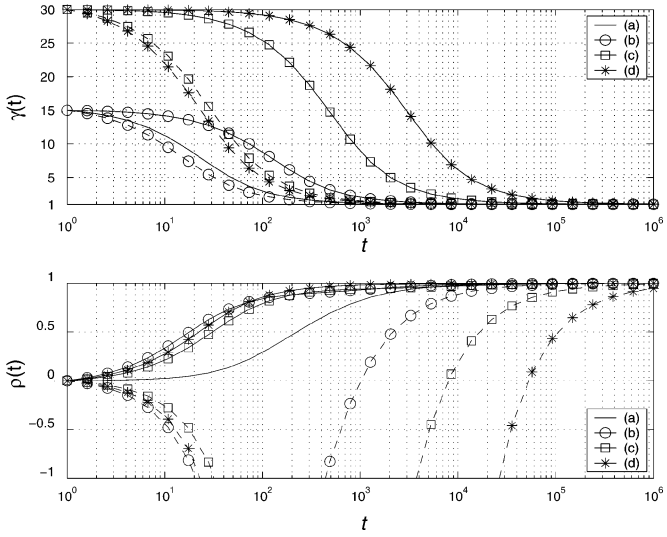

Fig. 2. Bounds for the normalized variance $\gamma(t)$ and for the consensus coefficient $\rho(t)$, as provided by the Proposition in Section III. The upper bounds are drawn as solid curves, while the lower bounds as dashed lines. The labeling (a)–(d) refers to the networks in Fig. 1 (in (a) the two bounds coincide). Recall that (a) and (b) refer to $n = 15$, while (c) and (d) refer to $n = 30$, whence the differences of $\gamma(t)$ at the initial point $t = 1$.

and the same probability, so that $W(t) = W_{ij}$ and any realization of such random matrix (any choice of an admissible pair (i, j)) has the same chance of occurrence.

In this case, the eigenvalues appearing in the Proposition proved in Section III admit a simple interpretation. Indeed, from (19) we immediately see that W_{ij} is doubly stochastic, symmetric and idempotent. The last two properties imply that $E[WW^T] = E[W]$, with the consequence that the eigenvalues λ_L and λ_U can be equivalently referred to this latter matrix. In addition, it can be easily shown that our basic requirement, namely $\lambda_U < 1$, is fulfilled provided that the graph associated to $E[W]$ is strongly connected [17], and is certainly true in the architectures of Fig. 1.

In Fig. 1(a), $n = 15$ sensors are arranged to form a ring, and all the pairs of node are admissible. In panel Fig. 1(b) the same ring topology of Fig. 1(a) is considered, with the difference that each sensor can only communicate with four neighbors, two in one direction and two in the opposite one, and this result in a lower number of admissible pairs. Similarly, Figs. 1(c)–Fig. 1(d) refer to a network made of $n = 30$ sensors with a topology typical of randomly deployed sensors, and in (c) more pairs of nodes are admissible than in (d).

The considered architectures determine $E[W]$ and, specifically, its eigenvalues λ_U and λ_L. For the scenario in Fig. 1(a), we have $\lambda = \lambda_U = \lambda_L = (n-2)/(n-1) (\approx 0.9286$ in our case). In fact, we can easily find $E[W] = \lambda I + (1-\lambda)/(n)\mathbf{1}\mathbf{1}^T$, and the eigenvalues of such matrix are $\lambda, \lambda, \ldots, \lambda, 1$. The bounds in both the (9) and (10) coincide, implying that $\gamma(t)$ and $\rho(t)$ can be computed exactly. These functions are drawn in Fig. 2 as solid lines without markers in top and bottom panels, respectively. We see that $\gamma(t)$ starts from $n = 15$ and decreases monotonically toward 1, while $\rho(t)$ grows monotonically from 0 to 1.

Consider now the network of Fig. 1(b): here we have $\lambda_U \approx 0.9868$ and $\lambda_L \approx 0.8921$. The bounds on $\gamma(t)$, see (9), are shown in the top plot of Fig. 2, while the bounds in (10) for the consensus coefficient $\rho(t)$ are drawn in the bottom plot. The vertical axis of this latter is limited to the meaningful range $(-1, 1)$; in fact the lower bound for $\rho(t)$ may occasionally fall below -1, thus loosing significance.

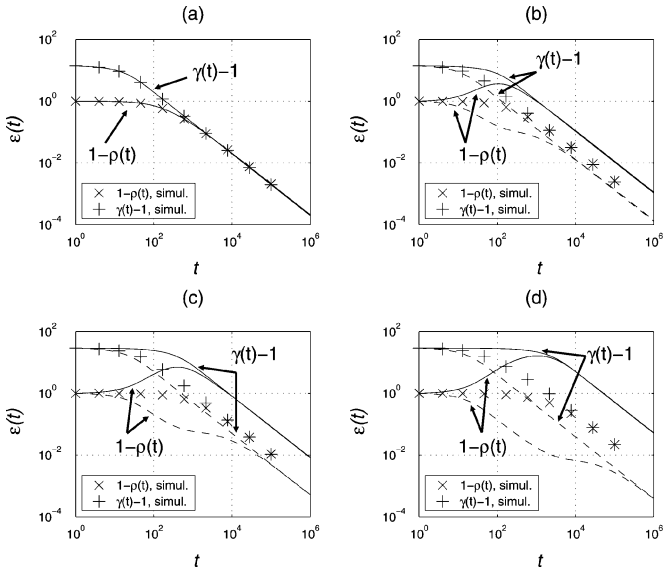

Fig. 3. Bounds for the performance indices $\epsilon(t) = \gamma(t) - 1$ and $\epsilon(t) = 1 - \rho(t)$ as function of t, for the same four scenarios (a)–(d) of Fig. 1. For large t, the bounds of the two performance figures coincide, as prescribed by (17). The upper bounds are drawn as solid curves, while the lower bounds as dashed lines. The points marked with "+" and "×" result from computer simulations for estimating $\gamma(t)$ and $\rho(t)$, respectively.

The eigenvalues (λ_U, λ_L) of the networks depicted in Fig. 1(c) and (d) are $(0.9964, 0.9426)$ and $(0.9994, 0.9262)$, respectively. The bounds in (9) and (10) are also drawn in Fig. 2. Note that the performance index $\gamma(t)$ now starts from $n = 30$.

The asymptotic behavior of the network performances is better highlighted in Fig. 3, where the panels (a)–(d) refer to the their analogue in Fig. 1, and the bounds for $\gamma(t) - 1$ and $1 - \rho(t)$ are drawn on the same plot. Note that, for large t, the bounds simplify as in (17), giving the portion of the curves marked by dots. Clearly, in Fig. 3(a) the upper and the lower bound coincide so that only two curves are drawn.

In Fig. 3 we also check the derived bounds by means of computer simulations based on a standard Monte Carlo counting procedure. For each time slot $t = 1, 2, \ldots$, the shown simulation points involve 10^3 program runs for computing the entries $C_{ij}(t)$ of the covariance matrix. Then, the estimated values of $\rho(t)$ and $\gamma(t)$ (recall that these indices depend upon i and j) are obtained as arithmetic averages of the pertinent entries; for instance, the values of $\gamma(t)$ result from averaging the n diagonal entries $C_{ii}(t)$, see (5).

V. Summary

In some applications of WSNs operating in dangerous environments, safe information retrieval must be guaranteed from an arbitrary node (or subset of nodes) at an arbitrary time. If the information to be retrieved is the arithmetic mean of all the observations collected by the whole network, this problem can be conveniently approached by a *running consensus* scheme. This paradigm prescribes that the sensors of the network simultaneously collect new observations from the environment and exchange data with neighbor nodes to perform averaging. The design goal is that all the nodes should reach agreement about a common value, that is just the total arithmetic mean of all the collected observations. With reference to a suitable updating rule for the network state [see (1)] we show that, in fact, this goal is achieved asymptotically with the time slot t.

More in detail, the concept of consensus is formalized by defining $\rho(t)$, a slight modification of the standard correlation coefficient; in addition, the performance of a node normalized to the optimal performance of an ideal centralized system is also considered and denoted by $\gamma(t)$. The theorem proved in this paper gives upper and lower bounds for the system performances $\rho(t)$ and $\gamma(t)$, as function of the main system parameters [see (9) and (10)]. The analysis of these bounds allows us to conclude that, in the limit of $t \to \infty$, the consensus is always reached $(\rho(t) \to 1)$, and exactly the same performance of an ideal centralized scheme is attained $(\gamma(t) \to 1)$. The indices $\rho(t)$ and $\gamma(t)$ both approach their limit value 1 according to the same scaling law rt^{-1}, with the network topology and its connectivity properties ruling the rate coefficient r, but not the functional dependence upon t. One consequence is that the network characteristics impact more marginally the asymptotic behavior of the system, with respect to what happens in classical consensus scenarios.

The setup considered in this paper deliberately disregards a number of practical concerns including synchronism issues, communication errors, bandwidth limitations. In addition, with running data it would be of interest to consider the case where underlying statistics of the measured data are changing over time. Relaxing some of the adopted working hypotheses could be the subject of future investigations.

Acknowledgment

The authors would like to thank the anonymous reviewers for their hints. One of them, in particular, suggested a generalization to a broader class of consensus algorithms.

References

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

[2] W. Ren, R. W. Beard, and E. M. Atkins, "A survey of consensus problems in multi-agent coordination," in *Proc. Amer. Control Conf.*, Portland, OR, Jun. 2005, pp. 1859–1864.

[3] N. A. Lynch, *Distributed Algorithms*. San Mateo, CA: Morgan Kaufmann, 1997.

[4] P. F. Swaszek and P. Willett, "Parley as an approach to distributed detection," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 31, no. 1, pp. 447–457, Jan. 1995.

[5] J. Tsitsiklis and M. Athans, "Convergence and asymptotic agreement in distributed decision problems," *IEEE Trans. Autom. Control*, vol. 29, no. 1, pp. 42–50, Jan. 1984.

[6] S. S. Boyd, A. Ghosh, and B. S. D. Prabhakar, "Gossip algorithms: Design, analysis and applications," in *Proc. INFOCOM*, Miami, FL, Mar. 13–17, 2005, vol. 3, pp. 1653–1664.

[7] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2508–2530, Jun. 2006.

[8] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *Syst. Control Lett.*, vol. 53, pp. 65–78, 2004.

[9] A. Kashyap, T. Basar, and R. Srikant, "Quantized consensus," in *Proc. Int. Symp. Inf. Theory (ISIT)*, Seattle, WA, Jul. 9–14, 2006, pp. 635–639.

[10] L. Scardovi and R. Sepulchre, "Collective optimization over average quantities," in *Proc. 45th IEEE Conf. Decision Control*, San Diego, CA, Dec. 13–15, 2006, pp. 3369–3374.

[11] A. Olshevsky and J. N. Tsitsiklis, "Convergence rates in distributed consensus and averaging," in *Proc. 45th IEEE Conf. Decision Control*, San Diego, CA, Dec. 13–15, 2006, pp. 3387–3392.

[12] Y.-W. Hong, A. Scaglione, and P. Varshney, "A communication architecture for reaching consensus in decision for a large network," in *Proc. IEEE/SP 13th Workshop on Statist. Signal Process.*, Jul. 17–20, 2005, pp. 1220–1225.

[13] Z. Jin and R. M. Murray, "Multi-hop relay protocols for fast consensus seeking," in *Proc. 45th IEEE Conf. Decision Control*, San Diego, CA, Dec. 13–15, 2006, vol. 1, pp. 1001–1006.

[14] R. Olfati-Saber, "Distributed Kalman filter with embedded consensus filters," in *Proc. 44th IEEE Conf. Decision Control*, 2005, pp. 8179–8184.

[15] S. Barbarossa and G. Scutari, "Bio-inspired sensor network design," *IEEE Signal Process. Mag.*, vol. 24, no. 3, pp. 26–35, May 2007.

[16] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3rd ed. New York: McGraw-Hill, 1991.

[17] R. Horn and C. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.