

Distributed Sequential Estimation With Noisy, Correlated Observations

Sudharman K. Jayaweera, *Member, IEEE*, and Carlos Mosquera, *Member, IEEE*

Abstract—The problem of distributed sequential estimation of a nonrandom parameter over noisy communication links is considered, with observations that are correlated spatially across the sensor field. A recursive algorithm for updating the sequential estimator is derived assuming wide-sense stationary observations. It is shown that the performance of the sequential estimator is a trade-off between the quality of node observations and communication channels. A sufficient condition for the convergence of the estimator variance is derived, and asymptotic expressions for variance when this condition is met are obtained for both iid and correlated observations.

Index Terms—Distributed estimation, distributed sequential estimation, sensor networks, sequential estimation.

I. INTRODUCTION

IN THIS letter, we address a problem of distributed sequential estimation of a nonrandom parameter. It is assumed that a network of sensor nodes is linked via noisy communication channels. Each node makes an observation that is statistically related to a parameter θ of interest. In addition, it receives the estimator at the previous node in the network (assuming a particular predetermined ordering of nodes, for example) over a noisy communication link. Each node forms an updated estimator by combining these two observations optimally (in the sense of unbiased linear estimators), before passing it on to the next node down the network.

If retransmissions are allowed between two nodes, the problem can be approximated by distributed estimation with error-free links. Thus, our interest is when retransmissions between nodes are not allowed. This may be the case, for example, when node power consumption is an important consideration. Another situation in which our formulation could be justified is in taking into account finite precision effects in encoding and decoding, and quantization errors. We formulate this problem based on best linear unbiased estimation (BLUE). In general, we allow node observations to be spatially correlated. A recursive algorithm for sequential updating of the estimator at distributed nodes is derived. It is shown that due to the trade-off

between channel and observation noise, sequential updating may not always lead to improved estimator performance. A sufficient condition for this to happen is given, and the asymptotic variance achieved by the distributed sequential estimator is derived.

Although there is an extensive literature on distributed detection [1]–[3], sequential detection [4], [5], and distributed estimation [6]–[8], the problem of distributed sequential estimation, especially over noisy communication channels, seems to have received little attention. In this letter, our objective is to formulate this problem and analyze the estimator performance under several interesting scenarios. A particular work that has considered distributed recursive estimation over adaptive networks is [9]. However, it is to be noted that [9] differs from this letter in several aspects. First and foremost, it does not consider noisy communications between nodes, which is the main focus of this letter. In addition, in this letter, we consider generally correlated node observations.

II. SENSOR NET MODEL AND PROBLEM FORMULATION

We consider the problem of sequentially estimating a nonrandom parameter θ in an n -node distributed sensor system. The observation at the k th sensor, for $k = 0, 1, 2, \dots, n-1$, is modeled as $w_k = \theta + v_k$, where observation noise v_k is assumed to be a sequence of possibly non-iid, zero-mean random variables with variances σ_k^2 , and θ is the fixed parameter of interest to be estimated. It is assumed that each node forms its estimator sequentially (in a predetermined order) and forwards its estimator to the next node in the network. The next node forms an updated estimator of θ by combining its own observation with the noise-corrupted estimator of the previous node received over a noisy communication link. Denoting the noise corrupted estimator of the previous node received at the k th node by y_k , the augmented observation vector at the k th node can be written as

$$\mathbf{z}_k = \begin{bmatrix} w_k \\ y_k \end{bmatrix} = \begin{bmatrix} \theta + v_k \\ \hat{\theta}_{k-1}(\mathbf{z}_{k-1}) + n_k \end{bmatrix} \quad (1)$$

where $\hat{\theta}_k(\mathbf{z}_k)$ and n_k denote the BLUE at node k based on the augmented observation \mathbf{z}_k and the zero-mean, additive communication (receiver) noise with variance σ_c^2 , respectively. It is assumed that communication noise n_k is a sequence of iid random variables that is independent of v_k for all k . For completeness, it should be noted that $\mathbf{z}_0 = z_0 = w_0 = \theta + v_0$. It follows from the unbiasedness of the BLUE estimator $\hat{\theta}_{k-1}(\mathbf{z}_{k-1})$ at node $k-1$ that $\mathbb{E}\{\mathbf{z}_k\} = \mathbf{1}\theta$, where $\mathbf{1}$ denotes the vector $[1, 1]^T$.

A. Uncorrelated Node Observations

If we denote the variance of the BLUE estimator $\hat{\theta}_k$ at node k by P_k and the 2×2 covariance matrix of \mathbf{z}_k by Σ_k , it is easily

Manuscript received April 23, 2008; revised June 24, 2008. The work of C. Mosquera was supported by the Spanish Government under project TEC2007-68094-C02-01/TCM. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Mihaela van der Scaar.

S. K. Jayaweera is with the Department of Electrical and Computer Engineering, University of New Mexico, Albuquerque, NM 87131-0001 USA (e-mail: jayaweera@ece.unm.edu).

C. Mosquera is with the Departamento de Teoría de la Señal y Comunicaciones, Universidad de Vigo, 36310 Vigo, Spain (e-mail: mosquera@gts.tsc.uvigo.es).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>

Digital Object Identifier 10.1109/LSP.2008.2003990

seen that $\Sigma_k = Cov(\mathbf{z}_k) = diag(\sigma_k^2, P_{k-1} + \sigma_c^2)$. It follows that the BLUE estimator [10] at node k is

$$\hat{\theta}_k(\mathbf{z}_k) = \frac{\mathbf{1}^T \Sigma_k^{-1} \mathbf{z}_k}{\mathbf{1}^T \Sigma_k^{-1} \mathbf{1}} = \frac{P_{k-1} + \sigma_c^2}{G(k)} w_k + \frac{\sigma_k^2}{G(k)} y_k \quad (3)$$

where, for brevity, we have defined $G(k) = P_{k-1} + \sigma_c^2 + \sigma_k^2$. If we define the estimator error as $\tilde{\theta}_k \triangleq \hat{\theta}_k(\mathbf{z}_k) - \theta$, then the estimator error dynamics are governed by the following non-constant coefficient difference equation:

$$\tilde{\theta}_k = \frac{\sigma_k^2}{G(k)} \tilde{\theta}_{k-1} + \frac{P_{k-1} + \sigma_c^2}{G(k)} v_k + \frac{\sigma_k^2}{G(k)} n_k. \quad (4)$$

The error variance P_k of the BLUE estimator at node k is given by

$$P_k = \frac{1}{\mathbf{1}^T \Sigma_k^{-1} \mathbf{1}} = \left(\frac{1}{P_{k-1} + \sigma_c^2} + \frac{1}{\sigma_k^2} \right)^{-1} \quad (5)$$

which can be arranged as

$$P_k = P_{k-1} - \frac{(P_{k-1} + \sigma_c^2) P_{k-1} - \sigma_c^2 \sigma_k^2}{P_{k-1} + \sigma_c^2 + \sigma_k^2} \quad (6)$$

with the initialization $P(0) = Var(\hat{\theta}_0) = \sigma_0^2$. From (6), it can be seen that $P_k \leq P_{k-1}$ if and only if the k th node's observation quality is above a certain threshold, i.e.,

$$P_k \leq P_{k-1} \text{ if and only if } \sigma_k^2 \leq \left(1 + \frac{P_{k-1}}{\sigma_c^2}\right) P_{k-1}. \quad (7)$$

This condition is of course a consequence of noisy communication channels. If the links between nodes were to be perfect, regardless of the quality of the observation at each node, the updated estimator would be better than that at the previous node.

1) *iid Observations*: An example that is of particular interest is the identical observation variances across all nodes, $\sigma_k^2 = \sigma_0^2$ for all k . Note that even in this case, the estimator dynamics are still governed by a nonconstant coefficient difference equation. It is however easy to show that in this case, $P_k \leq P_{k-1}$ for all $k \geq 1$ (see Appendix A). Since sequence $\{P_k\}$ is lower bounded by zero, then as k tends to infinity, the asymptotic variance converges to $P_k = P_{k-1} = P_\infty$. Hence, from (6), we can show that

$$P_\infty = \frac{\sigma_c^2}{2} \left(\sqrt{1 + 4 \frac{\sigma_0^2}{\sigma_c^2}} - 1 \right). \quad (8)$$

It is easy to verify that $P_\infty \leq \sigma_0^2$, as it should be. It is also of interest to note that

$$\text{if } \sigma_c^2 \gg \sigma_0^2: P_\infty \approx \sigma_0^2 \quad (9)$$

$$\text{if } \sigma_c^2 \ll \sigma_0^2: P_\infty \approx \sigma_c \sigma_0 \left(1 - \frac{1}{2} \frac{\sigma_c}{\sigma_0}\right). \quad (10)$$

Intuitively, (9) says that when the communication channel is unreliable, the performance attained is determined mainly by node observation quality. In other words, we essentially disregard the previous nodes estimate received through the unreliable link. On the other hand, (10) states that when the channel is reliable (compared to the uncertainty in node observations), the asymptotically achievable performance is given by the geometric mean of the two uncertainties, σ_0^2 and σ_c^2 . Fig. 1 shows the convergence of the estimator variance P_k to the asymptotic P_∞ in (8) for iid observations.

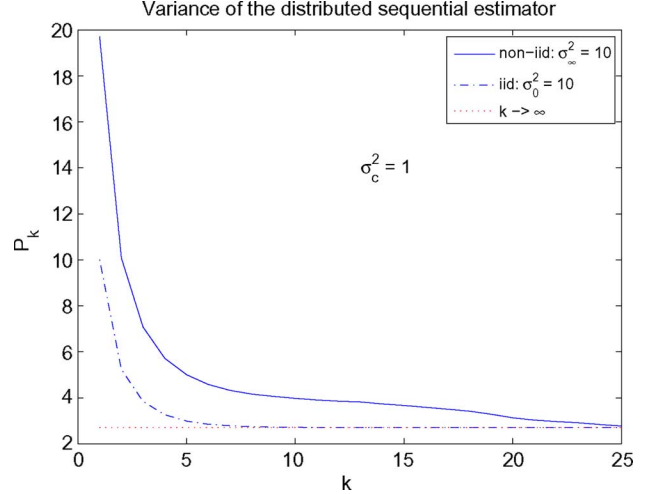


Fig. 1. Distributed sequential estimator error variance P_k and its convergence with iid versus non-iid observations ($\sigma_c^2 = 1$).

2) *Non-Identical Observations With Monotonically Increasing Quality*: A situation where the convergence of the sequential estimator will be guaranteed with non-iid observations is when the sequence of observation noise variances $\{\sigma_k^2\}$ is a monotonically decreasing sequence such that $\lim_{k \rightarrow \infty} \sigma_k^2 = \sigma_\infty^2$. In that case, it is straightforward to show that

$$P_\infty = \frac{\sigma_c^2}{2} \left(\sqrt{1 + 4 \frac{\sigma_\infty^2}{\sigma_c^2}} - 1 \right). \quad (11)$$

Comparison of (11) with (8) shows that sequential distributed estimator's achievable performance in this case is as if all the observations were iid with the minimum observation noise variance σ_∞^2 . Fig. 1 shows the variance of the distributed sequential estimator as a function of the node index with non-iid observations with $\sigma_\infty^2 = 10$. The observation noise variances $\{\sigma_k^2\}$ were randomly drawn from the interval $(\sigma_\infty^2, 2\sigma_\infty^2)$ but ordered to be a monotonically decreasing sequence. Also shown in Fig. 1 is the variance evolution if the observation were to be iid with $\sigma_0^2 = \sigma_\infty^2$. Fig. 1 confirms the asymptotic convergence obtained above. The non-iid observations require more nodes for convergence to the asymptotic performance as expected.

B. Correlated Node Observations

In practice, it is likely that node observations are spatially correlated. When this is the case, the optimal distributed sequential estimator $\hat{\theta}_{k-1}(\mathbf{z}_{k-1}) = (\mathbf{1}^T \Sigma_{k-1}^{-1} \mathbf{z}_{k-1} / \mathbf{1}^T \Sigma_{k-1}^{-1} \mathbf{1})$ obtained at node $k-1$ becomes also correlated with the observation at node k , leading to a nondiagonal covariance matrix Σ_k at node k

$$\Sigma_k = \begin{bmatrix} \sigma_k^2 & r_{k,k-1} \\ r_{k,k-1} & P_{k-1} + \sigma_c^2 \end{bmatrix} \quad (12)$$

where we have defined

$$r_{k,k-m} \triangleq E\{v_k \tilde{\theta}_{k-m}\} \quad \text{for } m = 1, 2, \dots, k \quad (13)$$

with, as before, $\tilde{\theta}_k = \hat{\theta}_k(\mathbf{z}_k) - \theta$. The sequential estimator at node k can thus be obtained as

$$\hat{\theta}_k(\mathbf{z}_k) = \frac{P_{k-1} + \sigma_c^2 - r_{k,k-1}}{G'(k)} w_k + \frac{\sigma_k^2 - r_{k,k-1}}{G'(k)} y_k \quad (14)$$

where $G'(k) = P_{k-1} + \sigma_c^2 + \sigma_k^2 - 2r_{k,k-1}$. The variance P_k of the estimator $\hat{\theta}_k(\mathbf{z}_k)$ is $P_k = (\sigma_k^2(P_{k-1} + \sigma_c^2) - r_{k,k-1}^2)/G'(k)$, with the initialization $P_0 = \sigma_0^2$. As a result, the recursion for $\hat{\theta}_k$ becomes

$$\tilde{\theta}_k = \frac{\sigma_k^2 - r_{k,k-1}}{G'(k)} \tilde{\theta}_{k-1} + \frac{P_{k-1} + \sigma_c^2 - r_{k,k-1}}{G'(k)} v_k + \frac{\sigma_k^2 - r_{k,k-1}}{G'(k)} n_k. \quad (15)$$

Hence, from (13) and (15), we have the following recursive expression for computing $r_{k,k-1}$ at node k :

$$r_{k,k-1} = \frac{\sigma_{k-1}^2 - r_{k-1,k-2}}{G'(k-1)} r_{k,k-2} + \frac{P_{k-2} + \sigma_c^2 - r_{k-1,k-2}}{G'(k-1)} \mathbb{E}\{v_k v_{k-1}\}. \quad (16)$$

Without any further assumptions on the specific correlation structure of the observations, computing $r_{k,k-1}$ via (16) requires a recursion that spans over all $r_{j,j-m}$ values for $m = 1, 2, \dots, j$ for each $j = k, k-1, \dots, 1$. However, in Appendix B, we have shown that when v_k is a wide-sense stationary (wss) sequence with identical variance σ_0^2 such that

$$\mathbb{E}\{v_k v_{k-m}\} = \rho^m \sigma_0^2 = \rho^m \sigma_0^2 \quad (17)$$

(16) can be manipulated to obtain a recursive algorithm for computing $r_{k,k-1}$ at node k based only on the $r_{k-1,k-2}$ computed at the previous node $k-1$. This leads to the following recursive distributed sequential estimation algorithm.

1) Initialization:

$$\hat{\theta}_0(\mathbf{z}_0) = w_0 \\ P_0 = \sigma_0^2 \quad \text{and} \quad r_{1,0} = \rho \sigma_0^2$$

2) For $k = 1, 2, 3, \dots$:

$$G'(k) = \sigma_c^2 + P_{k-1} + \sigma_0^2 - 2r_{k,k-1} \\ \hat{\theta}_k(\mathbf{z}_k) = \frac{P_{k-1} + \sigma_c^2 - r_{k,k-1}}{G'(k)} w_k + \frac{\sigma_k^2 - r_{k,k-1}}{G'(k)} y_k \\ F(k) = \frac{\sigma_c^2 + P_{k-1} - \sigma_0^2}{G'(k)} \\ P_k = \frac{\sigma_0^2 (\sigma_c^2 + P_{k-1}) - r_{k,k-1}^2}{G'(k)} \\ r_{k+1,k} = \frac{\rho}{2} [(1 + F(k)) \sigma_0^2 + (1 - F(k)) r_{k,k-1}]. \quad (18)$$

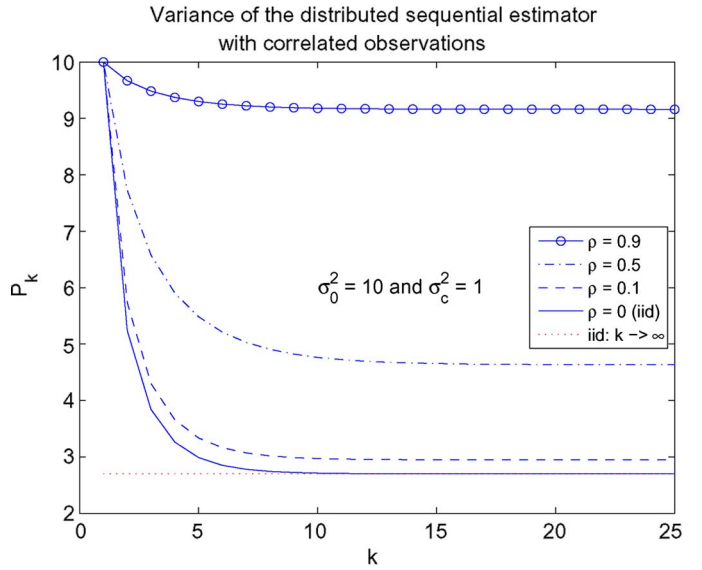


Fig. 2. Distributed sequential estimator error variance P_k with correlated observations. $\sigma_0^2 = 10$ and $\sigma_c^2 = 1$.

The above algorithm allows node k to update its sequential estimator based only on information received from node $k-1$. Specifically, node $k-1$ passes on its estimator $\hat{\theta}_{k-1}(\mathbf{z}_{k-1})$ and the associated statistics P_{k-1} and $r_{k,k-1}$ to node k . It should be noted that we are assuming that the second-order statistics P_{k-1} and $r_{k,k-1}$ are received error-free. This is a reasonable assumption when the statistics change slowly overtime so that they need only to be exchanged periodically once in a block. Note that, as long as second-order statistics of observations and channel noise variances are fixed, these quantities stay fixed so that they need to be exchanged only once. In that situation, these quantities can be exchanged among nodes with extra protection against noise, and since it is done once, the extra resources required can be neglected. It can also be shown that as $k \rightarrow \infty$, the final convergent estimator variance P_∞ is given by the solution to the coupled equations, shown at the bottom of the page, where we have denoted $r_\infty \triangleq \lim_{k \rightarrow \infty} r_{k,k-1}$. It can be shown that when $\sigma_0^2 \ll \sigma_c^2$, regardless of the value of ρ , the estimator variance converges to $P_\infty \rightarrow \sigma_0^2$ as in the case of uncorrelated observations, i.e.,

$$\text{if } \sigma_0^2 \ll \sigma_c^2 : r_\infty \rightarrow \rho \sigma_0^2 \quad \text{and} \quad P_\infty \rightarrow \sigma_0^2.$$

Fig. 2 shows the variance P_k of the distributed sequential estimator in a network with correlated observations.

$$r_\infty = \frac{\sigma_c^2 + P_\infty + \sigma_0^2 - \sqrt{(\sigma_c^2 + P_\infty + \sigma_0^2)^2 - 4\rho\sigma_0^2(2-\rho)(\sigma_c^2 + P_\infty)}}{2(2-\rho)} \\ P_\infty = \frac{-(\sigma_c^2 - 2r_\infty) + \sqrt{(\sigma_c^2 - 2r_\infty)^2 + 4(\sigma_c^2\sigma_0^2 - r_\infty^2)}}{2}$$

III. SUMMARY

In this letter, we considered the problem of distributed sequential estimation of a fixed parameter over noisy communication links. The performance of the optimal linear estimator was shown to be limited by the relative values of channel and observation noise variances. In particular, a sufficient condition for the convergence of the estimator variance with iid observations was derived. In the case of correlated observations, a recursive algorithm was provided to update the sequential estimator only based on information obtained at the previous node.

APPENDIX A

When node observations are iid, i.e., $\sigma_k^2 = \sigma_0^2$, we will use mathematical induction to show that $P_k \leq P_{k-1}$ for all $k \geq 1$. Let $f_k = (1/P_k) - (1/P_{k-1})$ for $k = 1, 2, \dots$. Since $P_0 = \sigma_0^2$ and $P_1 = (1/\sigma_1^2) + (1/P_0 + \sigma_c^2) = (1/\sigma_0^2) + (1/\sigma_0^2 + \sigma_c^2)$, it follows that $f_1 \geq 0$. Now suppose that $f_k \geq 0$, implying that $P_{k-1} \geq P_k$. Then, from (5)

$$f_{k+1} = \frac{1}{P_{k+1}} - \frac{1}{P_k} = \frac{1}{P_k + \sigma_c^2} - \frac{1}{P_{k-1} + \sigma_c^2} \geq 0.$$

By the principle of mathematical induction, then $f_k \geq 0$ for all $k \geq 1$. In other words, $P_k \leq P_{k-1}$ for all $k \geq 1$.

APPENDIX B DERIVATION OF ALGORITHM (18)

First note from (13) and (15) that, for $m = 1, \dots, k-1$

$$r_{k,k-m} = \frac{\sigma_0^2 - r_{k-m,k-m-1}}{G'(k-m)} r_{k,k-m-1} + \frac{P_{k-m-1} + \sigma_c^2 - r_{k-m,k-m-1}}{G'(k-m)} \rho_m \sigma_0^2 \quad (19)$$

where $G'(k-m) = \sigma_c^2 + P_{k-m-1} + \sigma_0^2 - 2r_{k-m,k-m-1}$. We can rearrange (19) as follows:

$$\begin{aligned} r_{k,k-m} &= \frac{1}{2} (r_{k,k-m-1} + \rho_m \sigma_0^2) \\ &\quad - \frac{P_{k-m-1} + \sigma_c^2 - \sigma_0^2}{2G'(k-m)} (r_{k,k-m-1} - \rho_m \sigma_0^2) \\ &= \frac{1}{2} [(1 - F(k-m)) r_{k,k-m-1} \\ &\quad + (1 + F(k-m)) \rho_m \sigma_0^2] \end{aligned}$$

where $F(k-m) \triangleq (P_{k-m-1} + \sigma_c^2 - \sigma_0^2)/G'(k-m)$. Back substitution of $r_{k,k-m}$ for $m = 2, 3, \dots, k$ in $r_{k,k-1}$ and using the fact that $r_{k,0} = \mathbb{E}\{v_k \tilde{\theta}_0\} = \rho_k \sigma_0^2$ leads to

$$\begin{aligned} r_{k,k-1} &= \frac{\rho_k \sigma_0^2}{2^{k-1}} \prod_{m=1}^{k-1} (1 - F(m)) + \sum_{j=1}^{k-1} \frac{\rho_j \sigma_0^2}{2^j} (1 + F(k-j)) \\ &\quad \times \prod_{m=1}^{j-1} (1 - F(k-m)) \quad (20) \\ &= \frac{\rho_k \sigma_0^2}{2^{k-1}} \prod_{m=1}^{k-1} (1 - F(m)) + \frac{\rho_1 \sigma_0^2}{2} (1 + F(k-1)) \\ &\quad + \sum_{j=1}^{k-2} \frac{\rho_{j+1} \sigma_0^2}{2^{j+1}} (1 + F(k-1-j)) \end{aligned}$$

$$\times \prod_{m=1}^j (1 - F(k-m)). \quad (21)$$

Now using the spatial correlation model $\rho_j = \rho^j$ for all j , and introducing a change of index $m \rightarrow m+1$, (21) can be manipulated into

$$\begin{aligned} r_{k,k-1} &= \frac{(1 - F(k-1)) \rho}{2} \frac{\rho^{k-1} \sigma_0^2}{2^{k-2}} \prod_{m=1}^{k-2} (1 - F(m)) \\ &\quad + \frac{\rho \sigma_0^2}{2} (1 + F(k-1)) \\ &\quad + \sum_{j=1}^{k-2} \frac{\rho^{j+1} \sigma_0^2}{2^{j+1}} (1 + F(k-1-j)) \\ &\quad \times \prod_{m=0}^{j-1} (1 - F(k-1-m)) \\ &= \frac{\rho}{2} (1 + F(k-1)) \sigma_0^2 + \frac{\rho(1 - F(k-1))}{2} \\ &\quad \times \left(\frac{\rho^{k-1} \sigma_0^2}{2^{k-2}} \prod_{m=1}^{k-2} (1 - F(m)) \right. \\ &\quad \left. + \sum_{j=1}^{k-2} \frac{\rho^j \sigma_0^2}{2^j} (1 + F(k-1-j)) \right. \\ &\quad \left. \times \prod_{m=1}^{j-1} (1 - F(k-1-m)) \right) \\ &= \frac{\rho}{2} [(1 + F(k-1)) \sigma_0^2 + (1 - F(k-1)) r_{k-1,k-2}] \end{aligned}$$

where the last step follows by comparison with (20).

REFERENCES

- [1] R. S. Blum, S. A. Kassam, and H. V. Poor, "Distributed detection with multiple sensors: Part II—Advanced topics," *Proc. IEEE*, vol. 85, no. 1, pp. 64–79, Jan. 1997.
- [2] P. K. Varshney, *Distributed Detection and Data Fusion*. New York: Springer-Verlag, 1996.
- [3] R. Viswanathan and P. K. Varshney, "Distributed detection with multiple sensors: Part I—Fundamentals," *Proc. IEEE*, vol. 85, no. 1, pp. 54–63, Jan. 1997.
- [4] V. V. Veeravalli, "Comments on decentralized sequential detection," *IEEE Trans. Inf. Theory*, vol. 38, no. 7, pp. 1428–1429, Jul. 1992.
- [5] V. V. Veeravalli, T. Basar, and H. V. Poor, "Decentralized sequential detection with a fusion center performing the sequential test," *IEEE Trans. Inf. Theory*, vol. 39, no. 2, pp. 433–442, Mar. 1993.
- [6] J. J. Xiao, S. Cui, Z. Q. Luo, and A. J. Goldsmith, "Joint estimation in sensor networks under energy constraints," in *Proc. IEEE 1st Conf. Sensor and Ad Hoc Communications and Networks*, Santa Clara, CA, Oct. 2004.
- [7] J. Xiao, A. Ribeiro, Z. Luo, and G. Giannakis, "Distributed compression-estimation using wireless sensor networks," *IEEE Signal Process. Mag.*, vol. 23, no. 4, pp. 27–41, Jul. 2006.
- [8] O. Roy and M. Vetterli, "Dimensionality reduction for distributed estimation in the infinite dimensional regime," *IEEE Trans. Inf. Theory*, vol. 54, no. 4, pp. 1655–1669, Apr. 2008.
- [9] A. H. Sayed and C. Lopes, "Distributed recursive least-squares strategies over adaptive networks," in *Proc. 40th Annu. Asilomar Conf. Signals, Systems and Computers*, Pacific Grove, CA, Nov. 2006, pp. 233–237.
- [10] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ: Prentice-Hall, 1993.