DISTRIBUTED ADAPTIVE ESTIMATION OF CORRELATED NODE-SPECIFIC SIGNALS IN A FULLY CONNECTED SENSOR NETWORK

Alexander Bertrand, Marc Moonen*

Katholieke Universiteit Leuven - Dept. ESAT
Kasteelpark Arenberg 10, B-3001 Leuven, Belgium
E-mail: alexander.bertrand@esat.kuleuven.be; marc.moonen@esat.kuleuven.be

ABSTRACT

We introduce a distributed adaptive estimation algorithm operating in an ideal fully connected sensor network. The algorithm estimates node-specific signals at each node based on reduced-dimensionality sensor measurements of other nodes in the network. If the node-specific signals to be estimated are linearly dependent on a common latent process with a low dimension compared to the dimension of the sensor measurements, the algorithm can significantly reduce the required communication bandwidth and still provide the optimal linear estimator at each node as if all sensor measurements were available in every node. Because of its adaptive nature and fast convergence properties, the algorithm is suited for real-time applications in dynamic environments, such as speech enhancement in acoustic sensor networks.

***Index Terms*—** Distributed estimation, wireless sensor networks (WSNs), adaptive estimation, distributed compression

1. INTRODUCTION

In a sensor network [1] a general objective is to utilize all information available in the entire network to perform a certain task, such as the estimation of a parameter or signal. In many multi-node estimation frameworks the measurement data is fused, possibly through a fusion center, to estimate a common parameter or signal assumed to be the same for each node (e.g. [2–6]). This can be viewed as a special case of the more general problem where each node in the network estimates a different node-specific signal. In this paper, we introduce a distributed adaptive node-specific signal estimation algorithm (DANSE), operating in an ideal fully connected network. The algorithm is based on reduced-dimensionality sensor observations to reduce the required communication bandwidth.

We will not make any assumptions on the data measured by the sensors. All node-specific desired signals, i.e. the signals to be estimated, are assumed linearly dependent on a common latent random process. If this process has a low dimensionality in comparison to the dimension of the sensor observations, the DANSE algorithm can exploit this to significantly compress the data to be broadcast by each node. Assuming the communication links are ideal, the algorithm will converge to the exact minimum mean squared error (MMSE) estimate at each node as if all sensor measurements were available in every node. Unlike other compression schemes for multi-dimensional sensor data (e.g. [4–6]), the algorithm does not need prior knowledge of the intra- and inter-sensor cross-correlation structure of the network. Nodes estimate and re-estimate all necessary statistics on the compressed data during operation.

Because of its adaptive nature and fast convergence properties, the algorithm is particularly relevant in dynamic environments, such as real-time speech enhancement. A pruned version of the DANSE algorithm, referred to as distributed multi-channel Wiener filtering (db-MWF), has partly been addressed in [7] and was used for microphone-array based noise reduction in binaural hearing aids (i.e. a network with 2 nodes). Optimality and convergence was proven for the case of a single desired speech source. The general DANSE algorithm introduced in this paper generalizes this to a scheme with multiple desired sources and more than 2 nodes, where convergence to an optimal estimator is still guaranteed.

This paper is organized as follows. The problem formulation and notation are presented in section 2. In section 3, we first address the simple case in which the node-specific desired signals are scaled versions of a common single dimension latent random variable. This is generalized to the case in which the node-specific desired signals are linear combinations of a Q-dimensional latent random variable in section 4. In section 5, we introduce a modification to the scheme that yields convergence when nodes update simultaneously, which permits parallel computation and uncoordinated updating. Conclusions are given in section 6.

2. PROBLEM FORMULATION AND NOTATION

Assume an ideal fully-connected network with J sensor nodes, i.e. a broadcast by any node can be captured by all other nodes in the network through an ideal link. Each node k has access to observations of an M_k-dimensional random complex measurement variable or signal $\mathbf{y}_k$. Denote $\mathbf{y}$ as the M dimensional random vector in which all $\mathbf{y}_k$ are stacked, where $M = \sum_{j=1}^{J} M_j$. In what follows, we will use the term 'single-channel/multi-channel signal' to refer to one-dimensional/multi-dimensional random processes. The objective for each node k is to estimate a complex desired signal d_k that is correlated to $\mathbf{y}$. For the sake of an easy exposition, we assume d_k to be a single-channel signal. In section 4, we will generalize this to multi-channel signals. We use a linear estimator $\hat{d}_k = \mathbf{w}_k^H \mathbf{y}$ for node k with $\mathbf{w}_k$ a complex M dimensional vector and superscript H denoting the conjugate transpose operator. Unlike [5, 6], we do not restrict ourselves to any data model for $\mathbf{y}$ nor do we make any assumptions on the statistics of the desired signals and the sen-

*Alexander Bertrand is a Research Assistant with the I.W.T. (Flemish Institute for Scientific and Technological Research in Industry). This research work was carried out at the ESAT laboratory of the Katholieke Universiteit Leuven, in the frame of the Belgian Programme on Interuniversity Attraction Poles, initiated by the Belgian Federal Science Policy Office IUAP P6/04 (DYSCO, 'Dynamical systems, control and optimization', 2007-2011), the Concerted Research Action GOA-AMBioRICS, and Research Project FWO nr. G.0600.08 ('Signal processing and network design for wireless acoustic sensor networks'). The scientific responsibility is assumed by its authors.

sor measurements, except for an implicit assumption on short-term stationarity. We will use a minimum mean squared error (MMSE) criterion for the node-specific estimator, i.e.

$$\mathbf{w}_k = \arg\min_{\mathbf{w}_k} E\{|d_k - \mathbf{w}_k^H \mathbf{y}|^2\} , \qquad (1)$$

where $E\{.\}$ denotes the expected value operator. We define a partitioning of the vector $\mathbf{w}_k$ as $\mathbf{w}_k = [\mathbf{w}_{k1}^T \ \ldots \ \mathbf{w}_{kJ}^T]^T$ where $\mathbf{w}_{kq}$ is the part of $\mathbf{w}_k$ that corresponds to $\mathbf{y}_q$. The equivalent of (1) is then

$$\mathbf{w}_k = \begin{bmatrix} \mathbf{w}_{k1} \\ \mathbf{w}_{k2} \\ \vdots \\ \mathbf{w}_{kJ} \end{bmatrix} = \arg\min_{\{\mathbf{w}_{k1},\ldots,\mathbf{w}_{kJ}\}} E\{|d_k - \sum_{l=1}^{J} \mathbf{w}_{kl}^H \mathbf{y}_l|^2\} . \quad (2)$$

The objective is to solve all J different MMSE problems, i.e. one for each node. Each node k only has access to $\mathbf{y}_k$ which is a subset of the full data vector $\mathbf{y}$. Notice that this approach differs from [2, 3], where the objective was to fit a linear model with coefficients $\mathbf{w}$, which are assumed to be equal for all nodes in the network, and where each node has access to different outcomes of the full data vector $\mathbf{y}$ and the joint desired signal d. In that case, only the estimation parameters must be transmitted, allowing for e.g. incremental strategies.

Assuming that the correlation matrix $\mathbf{R}_{yy} = E\{\mathbf{y}\mathbf{y}^H\}$ has full rank, the solution of (1) is

$$\hat{\mathbf{w}}_k = \mathbf{R}_{yy}^{-1} \mathbf{r}_k \qquad (3)$$

with $\mathbf{r}_k = E\{\mathbf{y} d_k^*\}$, where d_k^* denotes the complex conjugate of d_k. $\mathbf{r}_k$ can be estimated by using training sequences, or by exploiting on-off behavior of the desired signal, e.g. in a speech-plus-noise model, as in [7].

To find the optimal MMSE solution (3), each node k has to broadcast its M_k-channel signal $\mathbf{y}_k$ to all other nodes in the network, which requires a large communication bandwidth. One possibility to reduce the bandwidth is to broadcast only a few linear combinations of the M_k signals in $\mathbf{y}_k$. In general this will not lead to the optimal solution (3). In many practical cases however, the d_k signals are correlated through a common latent random process. The most simple case is when all $d_k = d$, i.e. the signal to be estimated is the same for all nodes. We will first handle the more general case where all d_k are scaled versions of a common latent random variable d. For this scenario, we will introduce an adaptive algorithm, in which the amount of data to be transmitted by each node k is compressed by a factor M_k. Despite this compression, the algorithm converges to the optimal node-specific solution (3) at every node as if each node has access to the full M-channel signal $\mathbf{y}$.

This scenario can then be extended to a more general case where all desired signals d_k are linear combinations of a Q-dimensional random process or signal. If each node is able to capture the Q-dimensional signal subspace generating the d_k's, then the amount of data to be transmitted by each node k can be compressed by a factor $\frac{M_k}{Q}$, and still the optimal node-specific solutions (3) are obtained at all nodes. This means that each node k only needs to broadcast Q linear combinations of the signals in $\mathbf{y}_k$.

3. DANSE IN A SINGLE-DIMENSIONAL SIGNAL SPACE (Q=1)

The algorithm introduced in this paper is an iterative scheme referred to as distributed adaptive node-specific signal estimation (DANSE), since its objective is to estimate a node-specific signal at each node in a distributed fashion. In the general scheme, each node k broadcasts a multi-channel signal with $\min\{K, M_k\}$ channels. We will refer to this with DANSE$_K$, where the subscript denotes the number of channels of the broadcast signal. For the sake of an easy exposition, we first introduce the DANSE$_1$ algorithm for the simple case where $K = 1$. In section 4 we will generalize these results to the more general DANSE$_K$ algorithm.

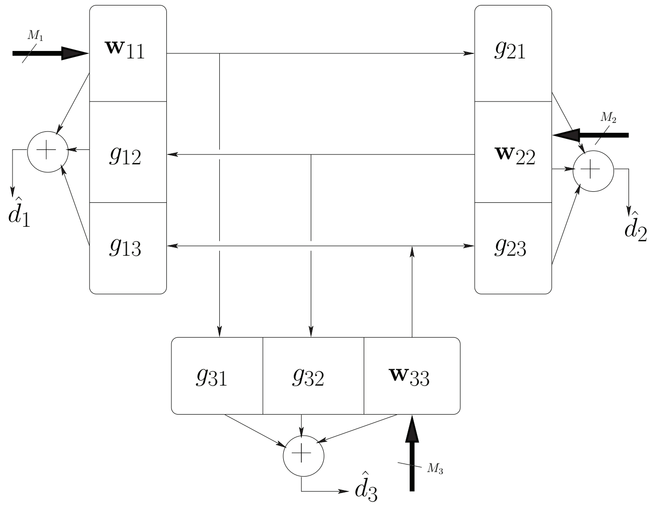

Fig. 1. The DANSE$_1$ scheme with 3 nodes ($J = 3$). Each node k estimates a signal d_k using its own M_k-channel signal, and 2 single-channel signals broadcast by the other two nodes.

The algorithm is described in batch mode. The iterative characteristic of the algorithm may therefore suggest that the same data must be broadcast multiple times, i.e. once after every iteration. However, in practical applications, iterations are spread over time, which means that subsequent iterations are performed on different signal segments. By exploiting the implicit assumption on short-term stationarity of the signals, every data segment only needs to be broadcast once, yet the convergence of DANSE and the optimality of the resulting estimators, as described infra, remains valid.

3.1. The DANSE$_1$ algorithm

The goal for each node is to estimate the signal d_k via the linear estimator $\hat{d}_k = \mathbf{w}_k^H \mathbf{y}$. We aim to find the MMSE solution (3) in an iterative way, without the need for each node to broadcast all channels of the M_k-channel signal $\mathbf{y}_k$. Instead, each node k will broadcast the signal $z_k^i = \mathbf{w}_{kk}^{i\,H} \mathbf{y}_k$, with superscript i denoting the iteration index and $\mathbf{w}_{kk}^i$ the estimate of $\mathbf{w}_{kk}$ as defined in (2) at iteration i. This reduces the data to be broadcast by a factor M_k. This means that each node k only has access to $\mathbf{y}_k$, and $J - 1$ linear combinations of the other channels in $\mathbf{y}$, generated by $\mathbf{w}_{qq}^{i\,H} \mathbf{y}_q$ with $q \in \{1, \ldots, J\} \backslash \{k\}$.

In the DANSE$_1$ scheme, a node k can scale the signal $\mathbf{w}_{qq}^{i\,H} \mathbf{y}_q$ that it receives from node q by a scalar g_{kq}^i. The structure of $\mathbf{w}_k^i$ is therefore

$$\mathbf{w}_k^i = \begin{bmatrix} g_{k1}^i \mathbf{w}_{11}^i \\ g_{k2}^i \mathbf{w}_{22}^i \\ \vdots \\ g_{kJ}^i \mathbf{w}_{JJ}^i \end{bmatrix} \qquad (4)$$

where node k can only optimize the parameters $\mathbf{w}_{kk}^i$ and $\mathbf{g}_k^i = [g_{k1}^i \ \dots \ g_{kJ}^i]^T$. We assume that $g_{kk}^i = 1$ for any i to minimize the degrees of freedom. We denote $\mathbf{g}_{k-k}^i$ as the vector $\mathbf{g}_k^i$ with entry g_{kk}^i omitted. A schematic illustration of the DANSE$_1$ scheme is shown in figure 1.

The DANSE$_1$ algorithm consists of the following iteration steps:

1. Initialize the iteration index $i \leftarrow 0$. For every $q \in \{1, ..., J\}$: initialize $\mathbf{w}_{qq}$ and $\mathbf{g}_{q-q}$ with non-zero random vectors $\mathbf{w}_{qq}^0$ and $\mathbf{g}_{q-q}^0$ respectively. Initialize $k \leftarrow 1$, denoting the next node that will update its local parameters $\mathbf{w}_{kk}$ and $\mathbf{g}_{k-k}$.
2. Node k updates its local parameters $\mathbf{w}_{kk}$ and $\mathbf{g}_{k-k}$ to minimize the local MSE, given its inputs consisting of the signal $\mathbf{y}_k$ and the compressed signals $z_q^i = \mathbf{w}_{qq}^{i\,H}\mathbf{y}_q$ that it received from the other nodes $q \neq k$. This comes down to solving the smaller local MMSE problem:
$$\left[\frac{\mathbf{w}_{kk}^{i+1}}{\mathbf{g}_{k-k}^{i+1}}\right] = \underset{\mathbf{w}_{kk}, \mathbf{g}_{k-k}}{\arg\min}\, E\left\{\left|d_k - \left[\mathbf{w}_{kk}^H \mid \mathbf{g}_{k-k}^H\right]\left[\frac{\mathbf{y}_k}{\mathbf{z}_{-k}^i}\right]\right|^2\right\} \tag{5}$$
with $\mathbf{z}_{-k}^i = \left[z_1^i \dots z_{k-1}^i z_{k+1}^i \dots z_J^i\right]^T$. The parameters of the other nodes do not change, i.e.
$$\forall\, q \in \{1, \dots, J\} \backslash \{k\} : \ \mathbf{w}_{qq}^{i+1} = \mathbf{w}_{qq}^i,\ \mathbf{g}_{q-q}^{i+1} = \mathbf{g}_{q-q}^i\ . \tag{6}$$
3. $k \leftarrow (k \mod J) + 1$
$i \leftarrow i + 1$
4. Return to step 2

3.2. Convergence and optimality of DANSE$_1$ if $Q = 1$

Assume that all d_k are a scaled version of the same signal d, i.e. $d_k = \alpha_k d$, with α_k a non-zero complex scalar. Formula (3) shows that in this case, all $\hat{\mathbf{w}}_k$ are parallel, i.e.

$$\hat{\mathbf{w}}_k = \alpha_{kq}\hat{\mathbf{w}}_q \ \ \forall\, k, q \in \{1, ..., J\} \tag{7}$$

with $\alpha_{kq} = \alpha_k^*/\alpha_q^*$. This shows that the global-network MMSE solution (3) at each node k is in the solution space defined by the parametrization (4).

Theorem 3.1. *Let $d_k = \alpha_k d$, $\forall\, k \in \{1, \dots, J\}$, with d a single-channel complex signal and $\alpha_k \in \mathbb{C}\backslash\{0\}$. Then the DANSE$_1$ algorithm converges for any initialization of the parameters to the MMSE solution (3) for any k.*

Proof. Omitted. □

4. DANSE IN A Q-DIMENSIONAL SIGNAL SPACE

4.1. The DANSE$_K$ algorithm

In the DANSE$_K$ algorithm, each node broadcasts a K-channel signal to other nodes. This compresses the data to be sent by node k by a factor of $\frac{M_k}{K}$. If the desired signals of all nodes are in the same Q-dimensional signal subspace, K should be chosen equal to Q (see section 4.2). We assume that each node k estimates a K-channel desired signal[1] $\mathbf{d}_k = [d_k(1) \ \dots \ d_k(K)]^T$. The signal(s) of interest can be a subset of this vector, in which case the other entries should be seen as auxiliary signals to capture the Q-dimensional signal space. Again, we use a linear estimator $\hat{\mathbf{d}}_k = \mathbf{W}_k^H\mathbf{y} = [\mathbf{w}_k(1) \ ... \ \mathbf{w}_k(K)]^H\mathbf{y}$. The objective for every node k is to find the solution of the MMSE problem

$$\min_{\mathbf{W_k}} E\left\{\|\mathbf{d}_k - \mathbf{W}_k^H\mathbf{y}\|^2\right\}\ . \tag{8}$$

The solution of (8) is

$$\hat{\mathbf{W}}_k = \mathbf{R}_{yy}^{-1}\mathbf{R}_k \tag{9}$$

where $\mathbf{R}_k = E\left\{\mathbf{y}\mathbf{d}_k^H\right\}$. We wish to obtain (9) without the need for each node to broadcast all channels of the M_k-channel signal $\mathbf{y}_k$. Instead each node k will broadcast the K-channel signal $\mathbf{z}_k^i = \mathbf{W}_{kk}^{i\,H}\mathbf{y}_k$ with $\mathbf{W}_{kk}^i$ the submatrix of $\mathbf{W}_k^i$ applied to the channels of $\mathbf{y}$ to which node k has access.

A node k can transform the K-channel signal that it receives from node q by a $K \times K$ transformation matrix $\mathbf{G}_{kq}^i$. The structure of $\mathbf{W}_k$ is therefore

$$\mathbf{W}_k^i = \begin{bmatrix} \mathbf{W}_{11}^i\mathbf{G}_{k1}^i \\ \mathbf{W}_{22}^i\mathbf{G}_{k2}^i \\ \vdots \\ \mathbf{W}_{JJ}^i\mathbf{G}_{kJ}^i \end{bmatrix}\ . \tag{10}$$

Node k can only optimize the parameters $\mathbf{W}_{kk}^i$ and $\mathbf{G}_k^i = [\mathbf{G}_{k1}^{i\,T} \dots \mathbf{G}_{kJ}^{i\,T}]^T$. We assume that $\mathbf{G}_{kk}^i = \mathbf{I}_K$ for any i with $\mathbf{I}_K$ denoting the $K \times K$ identity matrix.

Using this formulation, DANSE$_K$ is a straightforward generalization of the DANSE$_1$ algorithm as explained in section 3.1, where all vector-variables are replaced by their matrix equivalent. When node k updates its local variables $\mathbf{W}_{kk}$ and $\mathbf{G}_k$, it will solve the local MMSE problem defined by the generalized version of (5), i.e.

$$\left[\frac{\mathbf{W}_{kk}^{i+1}}{\mathbf{G}_{k-k}^{i+1}}\right] = \underset{\mathbf{W}_{kk}, \mathbf{G}_{k-k}}{\arg\min}\ E\left\{\|\mathbf{d}_k - \left[\mathbf{W}_{kk}^H \mid \mathbf{G}_{k-k}^H\right]\left[\frac{\mathbf{y}_k}{\mathbf{z}_{-k}^i}\right]\|^2\right\} \tag{11}$$

with $\mathbf{z}_{-k}^i = \left[\mathbf{z}_1^{i\,T} \dots \mathbf{z}_{k-1}^{i\,T}\mathbf{z}_{k+1}^{i\,T} \dots \mathbf{z}_J^{i\,T}\right]^T$.

4.2. Convergence and optimality of DANSE$_K$ if $Q = K$

A sufficient condition to assure that DANSE$_K$ will converge to (9), is that $\mathbf{d}_k = \mathbf{A}_k\mathbf{d}$ with $\mathbf{A}_k$ a $K \times K$ full rank matrix and $\mathbf{d}$ a K-channel complex signal. This means that all desired signals $\mathbf{d}_k$ are in the same K-dimensional signal subspace (i.e. $Q = K$). Formula (9) shows that in this case all $\hat{\mathbf{w}}_k(n)$ for any k and n are in the same K-dimensional subspace. This implies that

$$\forall\, k, q \in \{1, ..., J\} : \hat{\mathbf{W}}_k = \hat{\mathbf{W}}_q\mathbf{A}_{kq} \tag{12}$$

with $\mathbf{A}_{kq} = \mathbf{A}_q^{-H}\mathbf{A}_k^H$. Expression (12) shows that the MMSE solution (9) at each node k is in the solution space defined by the parametrization (10).

Theorem 4.1. *Let $\mathbf{d}_k = \mathbf{A}_k\mathbf{d}$, $\forall\, k \in \{1, \dots, J\}$, with $\mathbf{d}$ a complex K-channel signal and $\mathbf{A}_k$ a full rank $K \times K$ matrix. Then the DANSE$_K$ algorithm converges for any initialization of the parameters to the MMSE solution (9) for any k.*

Proof. Omitted. □

[1] The number of linearly independent signals in $\mathbf{d}_k$ should be at least K. For notational convenience, but without loss of generality, we assume that $\mathbf{d}_k$ contains exactly K signals. If the number of signals is higher than K, DANSE$_K$ selects K linearly independent signals of $\mathbf{d}_k$ which will be used for the information exchange. The remaining estimations can be handled internally by node k.

It can be proven that convergence of the DANSE$_K$ algorithm is at least as fast as the centralized equivalent that would use an alternating optimization (AO) technique (cfr. [8]) with partitioning following directly from the parameters J and M_k for each node.

5. PARALLEL COMPUTING AND UPDATING

A disadvantage of the DANSE$_K$ algorithm as described in the earlier sections is that nodes update their parameters sequentially. This implies that nodes cannot estimate their local correlation matrices and compute their inverses in parallel. Furthermore, sequential updating implies the need for a network-wide updating protocol.

In general, convergence of sequential iteration (Gauss-Seidel iteration) does not imply convergence of simultaneous updating (Jacobi iteration). Extensive simulations show that this also holds for the DANSE$_K$ algorithm: if nodes update simultaneously, the algorithm does not always converge. To achieve convergence with simultaneous updates[2], one should modify the DANSE$_K$ algorithm to a relaxed version. This means that a node will update its parameters to an interpolation point in between the newly computed parameters and the current parameters.

Consider the following update procedure that is performed for all k in parallel:

$$\mathbf{G}_k^{i+1} = \arg\min_{\mathbf{G}_k} \; E\{\|\mathbf{d}_k - \sum_{q=1}^{J} \mathbf{G}_{kq}^H \mathbf{W}_{qq}^{i\,H}\, \mathbf{y}_q\|^2\} \tag{13}$$

$$\mathbf{W}_{kk}^{i+1} = (1-\alpha^i)\mathbf{W}_{kk}^i \mathbf{G}_{kk}^{i+1} + \alpha^i F_k(\mathbf{W}^i) \tag{14}$$

with $\alpha^i \in (0,1]$, $\mathbf{W}^i = \left[\mathbf{W}_{11}^{i\,T} \ldots \mathbf{W}_{JJ}^{i\,T}\right]^T$, and F_k denoting the function that generates a new estimate for $\mathbf{W}_{kk}$ according to the DANSE$_K$ update (11). The following theorem describes a strategy for the stepsize α^i that guarantees convergence to the optimal parameter setting:

Theorem 5.1. *Assume all assumptions of theorem 4.1 are satisfied. Then the sequence $\{\mathbf{W}_k^i\}_{i\in\mathbb{N}}$ as in (10), generated by the update rules (13)-(14) with stepsizes α^i satisfying*

$$\alpha^i \in (0,1]\;, \tag{15}$$

$$\lim_{i\to\infty} \alpha^i = 0\;, \quad \sum_{i=0}^{\infty} \alpha^i = \infty\;, \tag{16}$$

converges for any initialization of the parameters to the MMSE solution (9) for any k.

Proof. Omitted. □

The update rule (13) increases the computational load at every sensor, since it solves an MMSE problem in addition to the implicit MSE minimization in $F_k(\mathbf{W}^i)$. However, extensive simulations indicate that this is not necessary. The $\mathbf{G}_{k-k}^{i+1}$ in (11), that are generated as a by-product in the evaluation of $F_k(\mathbf{W}^i)$, also yield convergence if the relaxed update (14) is applied, with $\mathbf{G}_{kk}^i = \mathbf{I}_K \;\forall\, i \in \mathbb{N}$.

The conditions (16) are quite conservative and may result in slow convergence. Extensive simulations indicate that in many cases, the parallel procedure converges without relaxation, i.e. $\alpha^i = 1, \forall\, i \in \mathbb{N}$. If this is not the case, a constant value $\alpha^i = \alpha^0$, $\forall\, i \in \mathbb{N}$, is observed to always yield convergence to (9), if α^0 is chosen small enough.

[2] In the rest of this section, we only consider simultaneous updates. As long as every node updates an infinite number of times, all results remain valid in an asynchronous updating scheme, where each node decides independently when and how often it updates its parameters. This removes the need for an updating protocol.

6. CONCLUSIONS

In this paper, we have introduced a distributed adaptive estimation algorithm for node-specific desired signals operating in a fully connected network in which nodes exchange reduced-dimension sensor measurements. If the signals to be estimated are all in the same low-dimensional signal subspace, the algorithm converges to an optimal estimator for each signal. The required statistics can be estimated and re-estimated during operation on the compressed sensor observations, rendering the algorithm suitable for application in dynamic environments. We introduced a relaxed version of the algorithm that also yields convergence when nodes compute and update simultaneously or asynchronously.

7. REFERENCES

[1] D. Estrin, L. Girod, G. Pottie, and M. Srivastava, “Instrumenting the world with wireless sensor networks,” *Acoustics, Speech, and Signal Processing, 2001. Proceedings. (ICASSP ’01). 2001 IEEE International Conference on*, vol. 4, pp. 2033–2036 vol.4, 2001.

[2] C. G. Lopes and A. H. Sayed, “Incremental adaptive strategies over distributed networks,” *Signal Processing, IEEE Transactions on*, vol. 55, no. 8, pp. 4064–4077, Aug. 2007.

[3] C. G. Lopes and A. H. Sayed, “Diffusion least-mean squares over adaptive networks: Formulation and performance analysis,” *Signal Processing, IEEE Transactions on*, vol. 56, no. 7, pp. 3122–3136, July 2008.

[4] I.D. Schizas, G.B. Giannakis, and Zhi-Quan Luo, “Distributed estimation using reduced-dimensionality sensor observations,” *Signal Processing, IEEE Transactions on*, vol. 55, no. 8, pp. 4284–4299, Aug. 2007.

[5] Zhi-Quan Luo, G.B. Giannakis, and Shuzhong Zhang, “Optimal linear decentralized estimation in a bandwidth constrained sensor network,” *Information Theory, 2005. ISIT 2005. Proceedings. International Symposium on*, pp. 1441–1445, Sept. 2005.

[6] Yunmin Zhu, Enbin Song, Jie Zhou, and Zhisheng You, “Optimal dimensionality reduction of sensor data in multisensor estimation fusion,” *Signal Processing, IEEE Transactions on*, vol. 53, no. 5, pp. 1631–1639, May 2005.

[7] Simon Doclo, Tim van den Bogaert, Marc Moonen, and Jan Wouters, “Reduced-bandwidth and distributed mwf-based noise reduction algorithms for binaural hearing aids,” *IEEE Trans. Audio, Speech and Language Processing*, vol. in press, 2008.

[8] James C. Bezdek and Richard J. Hathaway, “Some notes on alternating optimization,” in *Advances in Soft Computing*, pp. 187–195. Springer Berlin / Heidelberg, 2002.