

1-1-2011

# Spatial Whitening Framework for Distributed Estimation

Swarnendu Kar

*Syracuse University*, [swkar@syr.edu](mailto:swkar@syr.edu)

Pramod K. Varshney

*Syracuse University*, [varshney@syr.edu](mailto:varshney@syr.edu)

Follow this and additional works at: <http://surface.syr.edu/eecs>

 Part of the [Electrical and Computer Engineering Commons](#)

---

## Recommended Citation

S. Kar, P. K. Varshney, and C. Hao, "Spatial whitening framework for distributed estimation," in 2011 4th IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), 13-16 Dec. 2011, Piscataway, NJ, USA, 2011, pp. 293-6.

This Article is brought to you for free and open access by the L.C. Smith College of Engineering and Computer Science at SURFACE. It has been accepted for inclusion in Electrical Engineering and Computer Science by an authorized administrator of SURFACE. For more information, please contact [surface@syr.edu](mailto:surface@syr.edu).

# Spatial Whitening Framework for Distributed Estimation

Swarnendu Kar and Pramod K. Varshney  
Dept. of Electrical Engineering and Computer Science  
Syracuse University  
Syracuse, NY, 13244, USA  
Email: {swkar,varshney}@syr.edu

Hao Chen  
College of Engineering  
Boise State University  
Boise, ID 83725  
Email: haochen@boisestate.edu

**Abstract**—Designing resource allocation strategies for power constrained sensor network in the presence of correlated data often gives rise to intractable problem formulations. In such situations, applying well-known strategies derived from conditional-independence assumption may turn out to be fairly suboptimal. In this paper, we address this issue by proposing an adjacency-based spatial whitening scheme, where each sensor exchanges its observation with their neighbors prior to encoding their own private information and transmitting it to the fusion center. We comment on the computational limitations for obtaining the optimal whitening transformation, and propose an iterative optimization scheme to achieve the same for large networks. We demonstrate the efficacy of the whitening framework by considering the example of bit-allocation for distributed estimation.

## I. INTRODUCTION

Wireless sensor networks consist of spatially distributed noisy sensors that cooperatively monitor environmental conditions. Since the individual sensor nodes are characterized by limited energy, bandwidth and computational capability, the task of the fusion center (FC) is to make accurate inference about the phenomenon by requesting as little information from the sensor nodes as possible [1]. Depending on the particular application and set of constraints, the FC often has to adopt smart strategies to collect and process data [2]. While the design of optimum strategies in some cases is relatively easy under the assumption of *conditional independence*<sup>1</sup> across sensors, it is well known that the design gets harder and sometimes the optimum strategy is intractable when correlation has to be taken into account [3]. In particular, when the sensors are geographically close, they are expected to possess significant correlation among themselves and the optimum strategies derived for the independent case will no longer be optimal. In this paper, we introduce a framework called *spatial whitening* (to be formalized later) to deal with this problem.

Our framework stems from this idea: If two sensors in a network are highly correlated, they are also likely to be spatially close, which means that they should be able to communicate and exchange information among themselves in a relatively inexpensive manner (avoiding routing overheads and long distance

<sup>1</sup>Here, ‘independence’ refers to the statistical independence of sensor data *conditioned* on the parameter of interest. For additive Gaussian observation noise, this is equivalent to the covariance matrix of noise being diagonal. The observations are still marginally dependent, since they are observing the same parameter.

This research was partially supported by the National Science Foundation under Grant No. 0925854 and the Air Force Office of Scientific Research under Grant No. FA-9550-10-C-0179.

This is the postprint of a paper presented at the 2011 4th IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), 13-16 Dec. 2011, San Juan, Puerto Rico, USA.

Publisher’s website: <http://dx.doi.org/10.1109/CAMSAP.2011.6136007>

communications). Each sensor in the network can now use the information from neighboring nodes to achieve a *local whitening transformation*. If each of such local transformations can be coordinated, one can aim to achieve *global whitening*, and the transformed observations can then be transmitted to the FC using optimum encoding strategies (for inference, resource allocation, etc) that were derived for conditionally independent scenarios. Hence, this two-stage (whitening followed by encoding) framework potentially enables the use of several earlier known results in the presence of correlated noise.

We introduce the log-determinant divergence based formulation of spatial whitening in Section II. To illustrate the potential usage of this framework, we employ the problem of distributed parameter estimation [1], where several sensor nodes quantize their individual observations before sending them to FC. The goal is to minimize the expected distortion of the estimated parameter subject to a constraint on the total number of bits transmitted to the FC. We demonstrate that an optimal strategy for bit allocation (derived for independent scenario [3]) delivers increasingly better performance with increasing degree of whitening.

The whitening transformation described in this paper requires local message passing which is certainly not without cost. However, in this paper, we assign no cost to whitening, acknowledging fully that any actual implementation of a system would have to consider the tradeoff between the benefits of whitening and the cost of it. Investigations on this tradeoff is a worthy topic for future research.

The concept of whitening, in general, has mostly been addressed in a global framework till now. It is well known that the Karhunen-Loève Transform (KLT) [4] (also referred to as Principal Components Analysis, PCA) of a random vector with covariance matrix  $\Sigma = U\Lambda U^T$  provides the unique whitening transformation ( $U^T$ ) that is also orthogonal. However, PCA is ill suited for our problem, since those whitening transformations are not local, while the orthogonality property serves no additional purpose. The Cholesky decomposition  $\Sigma = LL^T$ , which provides the unique lower triangular whitening transformation ( $L^{-1}$ ), also requires non-local transformations. Moreover, the lower-triangular property imposes a tree-type dependence structure while in fact there is no natural ordering of spatially correlated data [5]. Other sparsity-inducing decompositions like Sparse-PCA [6] and vector Sparse-PCA [7] are *exploratory*<sup>2</sup> in nature, which means that the resulting transformations are not guaranteed to be local. In [8], a hardware-friendly technique was

<sup>2</sup>The placeholders for non-zero coefficients are not known/specified beforehand.

proposed to achieve generic spatial whitening transformations that were also global in scope. In distributed-KLT [4], individual nodes observe non-overlapping portions of a random vector and perform dimensionality-reduction (without collaboration with neighbors) for optimum reconstruction at the FC.

Local communication among sensors has mostly been used to address *in-network* inference problems till now. Distributed consensus problems [9] aim at designing iterative message passing schemes in order to compute (some linearly weighted) average at all the nodes. Graphical model based problems involve the selection of structured inverse-covariance matrices (example 2.5 in [10]) and the subsequent design of message passing schemes for (posterior) belief computation [11] at all nodes. In this paper, we address the fixed FC based problem - where the inference is performed at the FC rather than inside the network.

Our primary contribution in this paper is the formulation of a whitening framework that harnesses (some minimal) local communication among sensors for efficient resource allocation in fixed FC applications.

## II. PROBLEM STATEMENT

We consider  $N$  sensors in a network that is observing an unknown, deterministic, scalar parameter of interest  $\theta$  in the presence of zero-mean, correlated Gaussian noise with covariance  $\Sigma$ . Hence the sensor observations  $\mathbf{x} = [x_1, x_2, \dots, x_N]$  follow

$$\mathbf{x} \sim \mathcal{N}(\mathbf{1}\theta, \Sigma), \quad \mathbf{1} \triangleq [1 \dots 1]^T \in \mathbb{R}^N. \quad (1)$$

Note that the sensor observations  $x_k$ -s are conditionally-independent when  $\Sigma$  is a diagonal matrix. Let the neighborhood structure among the various nodes be represented by the  $N \times N$  adjacency matrix  $\mathbf{A}$ ,  $A_{ij} \in \{0, 1\}$ , which is expected to be sparsely populated. Entries  $A_{ij} = 1$  signify that node  $i$  is a neighbor of node  $j$ . A low-cost link for local communication is assumed to be available between two neighboring links. Since each node is trivially connected to itself,  $A_{ii} = 1$ . We denote the set of all  $\mathbf{A}$ -sparse matrices as

$$\mathcal{S}_A \triangleq \{\mathbf{W} \in \mathbb{R}^{N \times N} : W_{ij} = 0 \text{ if } A_{ij} = 0\}. \quad (2)$$

Note that because  $\mathbf{A}$  is the adjacency matrix, all linear transformations of the form

$$\tilde{\mathbf{x}} = \mathbf{W}\mathbf{x} \sim \mathcal{N}(\mathbf{W}\mathbf{1}\theta, \mathbf{W}\Sigma\mathbf{W}^T), \quad \mathbf{W} \in \mathcal{S}_A, \quad (3)$$

can be realized relatively inexpensively through local data transmissions, i.e., node  $k$  realizes the transformation  $\tilde{x}_k = \sum_{j \in \mathcal{N}_k} W_{kj}x_j$  by collaborating with its set of neighbors  $\mathcal{N}_k \triangleq \{j_1, j_2, \dots, j_{|\mathcal{N}_k|}\}$ , i.e., all the columns-indices of  $\mathbf{A}$  such that  $A_{k,j_i} = 1$ .

The goal is to find the optimal mean-preserving, whitening transformation, i.e., one for which  $\mathbf{W}\mathbf{1} = \mathbf{1}$ , and  $\mathbf{W}\Sigma\mathbf{W}^T$  is as *near* to some diagonal matrix as possible. The mean-preserving condition ensures that the problem framework is preserved, i.e., any resource allocation algorithm previously designed for the observation domain  $\mathbf{x}$  is applicable to new transformed domain  $\tilde{\mathbf{x}}$ . The whitening condition helps induce conditional independence across sensors (in some optimal sense). We chose the log-determinant divergence [12] as our metric for matrix-*nearness*, a point that we will elaborate later. The idea is that the nodes can use (optimally) whitened observations  $\tilde{x}_k$  (instead of original correlated observations  $x_k$ ) as the information to be encoded and relayed to the FC. This way an encoding strategy that was derived using conditional independence assumption

across sensors can be used to enhance the performance of the system. We will consider the application of optimal encoding for distributed estimation in Section IV and show the resulting improvement in performance due to the two-stage processing. But before that we describe our approach towards finding the optimum whitening transformation and comment on the computational aspects.

In the domain of symmetric positive-definite  $N \times N$  matrices, the log-determinant divergence of  $\mathbf{P}$  from  $\mathbf{Q}$  is defined [12] as

$$\mathcal{L}(\mathbf{P}; \mathbf{Q}) \triangleq \text{Tr } \mathbf{Q}^{-1}\mathbf{P} - \log \det \mathbf{P} - N + \log \det \mathbf{Q}. \quad (4)$$

It is well known that  $\mathcal{L}(\mathbf{P}; \mathbf{Q})$  is a Bregman-divergence [12] and hence convex in  $\mathbf{P}$  for any fixed  $\mathbf{Q}$ . Also  $\mathcal{L}(\mathbf{P}; \mathbf{Q}) \geq 0$  for all  $\mathbf{P}$  and  $\mathbf{Q}$  with equality if and only if  $\mathbf{P} = \mathbf{Q}$ . We formulate the spatial whitening problem as finding an  $\mathbf{A}$ -sparse, mean-preserving transformation  $\mathbf{W}$  and a diagonal matrix (with positive entries)  $\mathbf{D}$  such that the divergence  $\mathcal{L}(\mathbf{W}\Sigma\mathbf{W}^T; \mathbf{D})$  is minimized,

$$\min_{\mathbf{W}, \mathbf{D}} \mathcal{L}(\mathbf{W}\Sigma\mathbf{W}^T; \mathbf{D}) \quad \text{s.t. } \mathbf{W} \in \mathcal{S}_A, \mathbf{W}\mathbf{1} = \mathbf{1}. \quad (5)$$

We note from definition (4) that

$$\mathcal{L}(\mathbf{W}\Sigma\mathbf{W}^T; \mathbf{D}) = \mathcal{L}(\mathbf{D}^{-\frac{1}{2}}\mathbf{W}\Sigma\mathbf{W}^T\mathbf{D}^{-\frac{1}{2}}; \mathbf{I}), \quad (6)$$

where  $\mathbf{I}$  is the identity matrix. Using (6), we obtain an equivalent formulation of (5),

$$\min_{\mathbf{Z}} \mathcal{L}(\mathbf{Z}\Sigma\mathbf{Z}^T; \mathbf{I}) \quad \text{s.t. } \mathbf{Z} \in \mathcal{S}_A, \quad (7)$$

$$\text{where } \mathbf{W} = \delta^{-1}(\mathbf{Z}\mathbf{1})\mathbf{Z}, \quad \mathbf{D} = \delta^{-2}(\mathbf{Z}\mathbf{1}), \quad (8)$$

where  $\delta(\cdot)$  is the diagonalization<sup>3</sup> operator. We note that (7) is a significantly simplified re-formulation of (5). Using (4), we define the cost function w.r.t.  $\mathbf{Z}$  as

$$l(\mathbf{Z}) \triangleq \mathcal{L}(\mathbf{Z}\Sigma\mathbf{Z}^T; \mathbf{I}) = \text{Tr } \mathbf{Z}\Sigma\mathbf{Z}^T - \log \det \mathbf{Z}\mathbf{Z}^T + c_0, \quad (9)$$

where  $c_0 \triangleq -N - \log \det \Sigma$  is a constant. We refer to (7) as the *log-determinant divergence based spatial whitening* problem. If the cardinality of non-zero elements of  $\mathbf{A}$  is  $\text{nz}(\mathbf{A}) \leq N^2$ , then (7) is an optimization problem in  $\mathbb{R}^{\text{nz}(\mathbf{A})}$ .

Since  $\mathbf{Z}$  is not restricted to the set of symmetric positive-definite matrices (denoted by  $\mathcal{S}^{++}$ ), our objective function (9) does not inherit the convexity property of well known *max-det* problems [10]. Neither does the first-order gradient condition, written in matrix-derivative notations [13],

$$\frac{dl(\mathbf{Z})}{d\mathbf{Z}} = 2(\mathbf{Z}\Sigma - \mathbf{Z}^{-T}) \circ \mathbf{A} = 0, \quad (10)$$

where  $\circ$  denotes the element-wise (or Hadamard) product, lend itself to any known closed-form solution except in the trivial situation when  $\mathbf{A}$  is the all-1 matrix (in which case,  $\mathbf{Z}\Sigma\mathbf{Z}^T = \mathbf{I}$ , and any orthogonal multiple of the Cholesky factor  $\mathbf{L}^{-1}$  is a solution for  $\mathbf{Z}$ ). In the next section, we provide an iterative algorithm that finds (locally) optimal solutions to problem (7). Multiple runs using *good*<sup>4</sup> starting points must be used to mitigate the local-maxima problem and obtain a satisfactory solution. It may be noted here that in most of existing literature, matrix factorization problems of this nature (involving sparsity/structure) are inherently non-convex and can

<sup>3</sup>Function  $\mathbf{X} = \delta(\mathbf{x})$  is defined as  $\delta : \mathbb{R}^N \rightarrow \mathbb{R}^{N \times N}$  such that  $\mathbf{x}$  corresponds to the diagonal elements of  $\mathbf{X}$ , other elements being zero.

<sup>4</sup> $l(\mathbf{Z})$  is convex in the smaller subset  $\mathcal{S}_A \cap \mathcal{S}^{++}$ , the minima within which can be efficiently computed and considered a *good* starting point.

only guarantee locally optimal solutions [4], [6], [7].

### III. ITERATIVE ALGORITHM FOR SPATIAL WHITENING

In our iterative approach to solving problem (7), we update each row of elements in  $\mathbf{Z}$  to achieve the optimum decrement in divergence, while keeping the rest of the matrix unchanged. This process is repeated until convergence. Each such iteration is a convex optimization problem and we obtain closed form expressions for the updates. Some of the details in this section is skipped for the sake of brevity and relegated to [14].

Optimizing (7) with respect to the row-vector  $\mathbf{z}_k \triangleq \mathbf{Z}_{k, \mathcal{N}_k} \in \mathbb{R}^{|\mathcal{N}_k|}$  while keeping all the other elements of  $\mathbf{Z}$  constant is equivalent [14] to minimizing

$$g(\mathbf{z}_k) = \frac{1}{2} \mathbf{z}_k^T \boldsymbol{\Sigma}_k \mathbf{z}_k - \log(\mathbf{z}_k^T \mathbf{c}_k), \quad (11)$$

$$\boldsymbol{\Sigma}_k \in \mathbb{R}^{|\mathcal{N}_k| \times |\mathcal{N}_k|}, \mathbf{c}_k \in \mathbb{R}^{|\mathcal{N}_k|},$$

where  $\boldsymbol{\Sigma}_k$  denotes the  $\mathcal{N}_k$ -cliq covariance matrix extracted from  $\boldsymbol{\Sigma}$ , and the elements of  $\mathbf{c}_k$  are defined by

$$(\mathbf{c}_k)_i \triangleq (-1)^{k+j_i} \det(\mathbf{Z}_{-k, j_i}), \quad i = 1, 2, \dots, |\mathcal{N}_k|, \quad (12)$$

with  $\mathbf{Z}_{-k, j_i}$  denoting the matrix obtained after truncating the  $k^{\text{th}}$  row and  $j_i^{\text{th}}$  column of  $\mathbf{Z}$ . The first-order gradient condition of (11) implies  $(\mathbf{z}_k^T \mathbf{c}_k) \boldsymbol{\Sigma}_k \mathbf{z}_k = \mathbf{c}_k$ , solving which one obtains the unique extremum of (11),

$$\mathbf{z}_k^* = \frac{\boldsymbol{\Sigma}_k^{-1} \mathbf{c}_k}{\sqrt{\mathbf{c}_k^T \boldsymbol{\Sigma}_k^{-1} \mathbf{c}_k}}. \quad (13)$$

That  $\mathbf{z}_k^*$  is the minimizer follows from the convexity of (11) (the Hessian is  $(\boldsymbol{\Sigma}_k + (\mathbf{z}_k^T \mathbf{c}_k)^{-2} \mathbf{c}_k \mathbf{c}_k^T)$ , which is positive definite).

Each rank-one update of the form (13) can be efficiently computed using the well-known Woodbury-formula, details of which are relegated to [14]. Since the overall divergence of (7) decreases at each of the iterations of (11), and the minimum divergence is lower bounded (see equation (10)) by

$$\sup_{\mathbf{Z} \in \mathcal{S}_A} l(\mathbf{Z}) \geq \sup_{\mathbf{Z} \in \mathbb{R}^{N \times N}} l(\mathbf{Z}) = l(\mathbf{L}^{-1}) = 0, \quad (14)$$

this iterative algorithm is guaranteed to converge. It may be noted that these kind of iterative techniques are sometimes called block-coordinate-descent or terminal-by-terminal optimization [4].

In the remainder of this paper, we will focus on the application of spatial whitening to distributed estimation.

### IV. EXAMPLE: BIT-ALLOCATION FOR DISTRIBUTED ESTIMATION

We consider the practical parameter-estimation problem where individual sensors in a network are required to quantize their real-valued local measurements to an appropriate length and send the resulting discrete message to the FC, while the latter combines all the received messages to produce a final estimate [1]. The critical resource that needs to be conserved is the bandwidth or equivalently, the rate of transmission. Assume that the network consisting of  $N$  nodes is allowed to transmit only  $B$  bits in totality for a one-shot estimation problem. The question then is how to judiciously allocate the  $B$  bits among the various sensors such that the resulting distortion of estimate is minimized at the FC [3], [15]. For the sake of simplicity, we assume that each sensor incurs an equal per-bit cost for transmission.

We would use the quantization and bit allocation framework outlined in [3]. All observations  $x_k$ -s are assumed to be bounded to a finite interval  $[-U, U]$  and a uniform probabilistic quantization is performed. An observation is quantized with  $b_k$ -bits as follows. The quantization points  $a_j^{(k)} \in [-U, U], j = 1, \dots, 2^{b_k}$  are uniformly spaced such that  $a_{j+1}^{(k)} - a_j^{(k)} = 2U/(2^{b_k} - 1) \triangleq \Delta_k$ . Suppose that  $x_k \in [a_j^{(k)}, a_{j+1}^{(k)})$ . Then  $x_k$  is quantized to either  $a_{j+1}^{(k)}$  or  $a_j^{(k)}$  according to

$$P(m_k = a_j^{(k)}) = q, \quad P(m_k = a_{j+1}^{(k)}) = 1 - q, \quad (15)$$

where  $m_k$  is the resulting message and  $q = (a_{j+1}^{(k)} - x_k)/\Delta_k$ .

When the noise is Gaussian and independent across sensors, the subsequent near-optimal strategy [3] is particularly simple and allocates

$$b_k = \text{ROUND} \left[ \log_2 \left( 1 + \frac{1}{\lambda \sigma_k^2} \right) \right] \quad (16)$$

bits to the  $k^{\text{th}}$  sensor, where  $\sigma_k^2$  is the individual variance,  $\lambda > 0$  controls the overall sum of bits  $\sum_{k=1}^N b_k = B$  and the rounding is performed to the nearest integer. The idea is that FC broadcasts a lower value of  $\lambda$  when a more precise parameter estimate is needed. However, when the noise is correlated, strategy (16) is suboptimal and this is where spatial whitening can be of help. Once we perform a spatial whitening transformation in the observation space, the idea is that we effectively de-correlate the noise without losing any information and hence a strategy like (16) applied on the modified space can still deliver near-optimal performance.

Next we state the distortion metric derived in [3] which we shall use for comparing the performance of various schemes. For a random variable  $\mathbf{y} \sim \mathcal{N}(\boldsymbol{\theta}, \mathbf{C})$  that is effectively range limited in  $[-U, U]$ , the mean-square-error (MSE) for estimating  $\hat{\boldsymbol{\theta}}$  at FC (when  $y_k$  is quantized to  $m_k$  using  $b_k$  bits) following the scheme in (15), is given by

$$\text{MSE}(\hat{\boldsymbol{\theta}}) \approx \frac{\mathbf{1}^T \mathbf{C}^{-1} (\mathbf{C} + \mathbf{Q}) \mathbf{C}^{-1} \mathbf{1}}{(\mathbf{1}^T \mathbf{C}^{-1} \mathbf{1})^2}, \quad (17)$$

where  $\mathbf{Q}$  is the diagonal matrix with elements  $Q_{kk} = (U^2)/(2^{b_k} - 1)^2$ . It is assumed that FC is using the optimally weighted fusion rule  $\hat{\boldsymbol{\theta}} = (\mathbf{1}^T \mathbf{C}^{-1} \mathbf{1})^{-1} \mathbf{1}^T \mathbf{C}^{-1} \mathbf{m}$  (see [16]) on the quantized observations.

Our simulation setup is as follows. The spatial placement and neighborhood structure is modeled as a Random Geometric Graph  $RGG(N, r)$  [17], where sensors are uniformly distributed over a unit square with communication links present only for pairwise distances of at most  $r$ . The noise is modeled as an exponentially correlated Gaussian covariance matrix  $\boldsymbol{\Sigma}$ ,

$$\mathbf{x} \sim \mathcal{N}(\mathbf{1}\boldsymbol{\theta}, \boldsymbol{\Sigma}), \quad \Sigma_{i,j} = \sigma_i \sigma_j \alpha^{d_{i,j}}, \quad (18)$$

where  $\alpha \in (0, 1)$  is indicative of the degree of spatial correlation. A smaller value of  $\alpha$  indicates lower correlation with  $\alpha \rightarrow 0$  signifying completely independent observations.

We consider  $N = 50$  nodes and the particular RGG used for our simulation is depicted in Figure 1. The individual sensor variances  $\sigma_k^2$  are generated by uniform random numbers in the range  $[0.5, 1.5]$  and the correlation parameter  $\alpha = 0.02$ . The range-limit of observations is taken as  $U = 20$ .

In Figure 2, we compare the distortion performance  $\text{MSE}(\hat{\boldsymbol{\theta}})$  (17) corresponding to the three scenarios when strategy (16) is applied to various transformations of the data. The line labeled

not whitened corresponds to the naive case of strategy (16) being directly applied to the observation space  $\mathbf{x}$ . Expectedly, the performance of this scheme is suboptimal. In *spatially whitened* cases, we use the transformed variable (see (8))

$$\tilde{\mathbf{x}} = \mathbf{W}_r \mathbf{x}, \quad \mathbf{W}_r = \delta^{-1}(\mathbf{Z}_r \mathbf{1}) \mathbf{Z}_r, \mathbf{D}_r = \delta^{-2}(\mathbf{Z}_r \mathbf{1}), \quad (19)$$

where  $\mathbf{Z}_r$  is the minimum-divergence solution (7) subject to constraints that  $[\mathbf{Z}_r]_{ij} = 0$  if  $d_{ij} > r$ . We note that

$$\tilde{\mathbf{x}} \sim \mathcal{N}(\mathbf{1}\theta, \mathbf{W}_r \Sigma \mathbf{W}_r^T), \quad (20)$$

which implies that  $\tilde{x}_k$  possess the same mean as the signal, but corrupted only with *approximately* independent Gaussian noise with variance  $\gamma_k^2 \triangleq \text{Var}(\tilde{x}_k) \approx [\mathbf{D}_r]_{k,k} = ([\mathbf{Z}_r \mathbf{1}]_k)^{-2}$ . Strategy (16) is then applied on whitened space  $\tilde{\mathbf{x}}$  with  $\sigma_k^2$  replaced by  $\gamma_k^2$  in Equation (16). We have shown the performance for  $r = 0.1$  and  $r = 0.5$  in Figure 2. As the range  $r$  increases, we have more whitening and consequently the performance increases. Thirdly, we display the results for *orthogonally whitened (or PCA)* case, where we consider the well known eigenvalue decomposition  $\Sigma = \mathbf{U} \Lambda \mathbf{U}^T$  and consequently the whitening transformation

$$\tilde{\mathbf{x}} = \mathbf{D}^{-1} \mathbf{U}^T \mathbf{x}, \quad \mathbf{D} \triangleq \delta(\mathbf{U}^T \mathbf{1}). \quad (21)$$

Since PCA fully whitens  $\Sigma$  (by definition), its performance is expected to provide a lower bound on that of other schemes. This is confirmed by Figure 2. However, since the weights in PCA are not designed to be zero for sensors that are far apart, such a transformation may be impossible to realize in a power constrained network and hence not realistic. Finally, the Cramer-Rao lower bound  $\text{CRB} = 1/(\mathbf{1}^T \Sigma^{-1} \mathbf{1})$  is also displayed, which confirms that in the asymptotic regime with sufficient quantization bits per sensor, all these schemes perform identically.

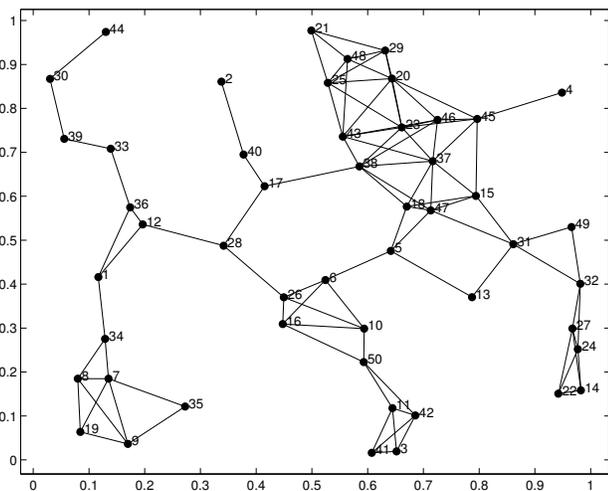


Fig. 1. Random Geometric Graph with 50 nodes, used for example in Section IV. Edges are shown of pairwise distance less than 0.18.

## V. CONCLUSION

In this paper, we have considered a two-stage framework for distributed signal processing in the presence of spatially correlated data. The first stage is designed to whiten the observation space by communicating only with neighboring sensors. In the second stage, each sensor encodes these whitened observations following well-known strategies derived using conditional independence assumption. We consider the example of bit-allocation for distributed estimation to demonstrate the potential applicability of this framework. Many research questions remain to be

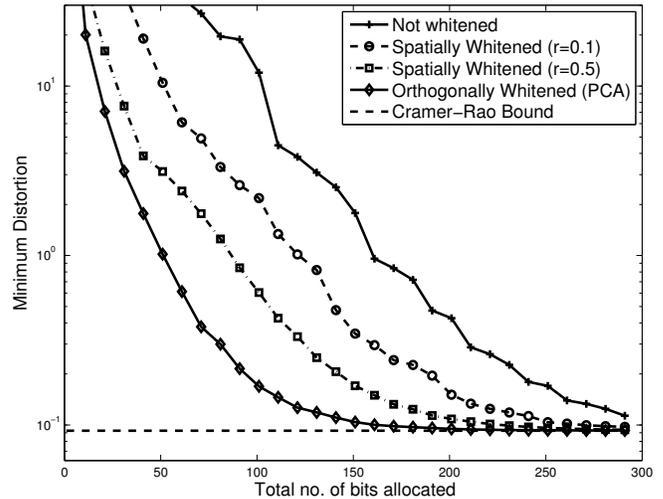


Fig. 2. Distortion reduction achieved by spatial whitening.

addressed. Some of them are efficient computation of the spatial whitening transformation, cost considerations for the whitening stage, extension of the framework to vector parameter scenarios and potential applicability in hypothesis testing problems.

## REFERENCES

- [1] A. Ribeiro and G. B. Giannakis, "Bandwidth-constrained distributed estimation for wireless sensor networks-Part I: Gaussian case," *Signal Processing, IEEE Transactions on*, vol. 54, no. 3, pp. 1131–1143, 2006.
- [2] P. K. Varshney, *Distributed Detection and Data Fusion*. New York: Springer, 1997.
- [3] A. Krasnoperov, J.-J. Xiao, and Z.-Q. Luo, "Minimum energy decentralized estimation in a wireless sensor network with correlated sensor noises," *EURASIP J. on Wireless Communications and Networking*, vol. 4, pp. 473–482, 2005.
- [4] M. Gastpar, P. Dragotti, and M. Vetterli, "The distributed Karhunen-Loève transform," *Information Theory, IEEE Transactions on*, vol. 52, no. 12, pp. 5177–5196, Dec. 2006.
- [5] Z. Zhu and Y. Liu, "Estimating spatial covariance using penalised likelihood with weighted  $L_1$  penalty," *Journal of Nonparametric Statistics*, vol. 21, no. 7, pp. 925–942, 2009.
- [6] H. Zou, T. Hastie, and R. Tibshirani, "Sparse principal component analysis," *Journal of Computational and Graphical Statistics*, vol. 15, no. 2, pp. 265–286, 2006.
- [7] M. Ulfarsson and V. Solo, "Vector  $l_0$  sparse variable PCA," *Signal Processing, IEEE Transactions on*, vol. 59, no. 5, pp. 1949–1958, May 2011.
- [8] S. Venkatesan, L. Mailaender, and J. Salz, "An iterative algorithm for computing a spatial whitening filter," in *Signal Processing Advances in Wireless Communications, 2004 IEEE 5th Workshop on*, July 2004.
- [9] A. Dimakis, S. Kar, J. Moura, M. Rabbat, and A. Scaglione, "Gossip algorithms for distributed signal processing," *Proceedings of the IEEE*, vol. 98, no. 11, pp. 1847–1864, Nov. 2010.
- [10] L. Vandenberghe, S. Boyd, and S. Wu, "Determinant maximization with linear matrix inequality constraints," *SIAM Journal on Matrix Analysis and Applications*, vol. 19, no. 2, pp. 499–533, 1998.
- [11] M. Cetin, L. Chen, J.W. Fisher III, A. Ihler, R. Moses, M. Wainwright, and A. Willsky, "Distributed fusion in sensor networks - A graphical models perspective," *Signal Processing Magazine, IEEE*, vol. 23, no. 4, pp. 42–55, July 2006.
- [12] I. Dhillon and J. Tropp, "Matrix nearness problems with Bregman divergences," *SIAM Journal on Matrix Analysis and Applications*, vol. 29, no. 4, pp. 1120–1146, 2007.
- [13] J. R. Magnus and H. Neudecker, *Matrix differential calculus with applications in statistics and econometrics*. New York: John Wiley, 1999.
- [14] S. Kar, H. Chen, and P. K. Varshney, "Log-determinant divergence based spatial whitening," *In preparation*.
- [15] J. Li and G. AIREgib, "Rate-constrained distributed estimation in wireless sensor networks," *Signal Processing, IEEE Transactions on*, vol. 55, no. 5, pp. 1634–1643, May 2007.
- [16] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice Hall, 1993.
- [17] N. Freris, H. Kowshik, and P. Kumar, "Fundamentals of large sensor networks: Connectivity, capacity, clocks, and computation," *Proceedings of the IEEE*, vol. 98, no. 11, pp. 1828–1846, Nov. 2010.