

DISTRIBUTED SENSING OF SIGNALS LINKED BY SPARSE FILTERING

Olivier Roy, Ali Hormati, Yue M. Lu, Martin Vetterli

School of Computer and Communication Sciences
Ecole Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland
email: {olivier.roy, ali.hormati, yue.lu, martin.vetterli}@epfl.ch

ABSTRACT

We consider the task of recovering correlated vectors at a central decoder based on fixed linear measurements obtained by distributed sensors. A general formulation of the problem is proposed, under both a universal and an almost sure reconstruction requirement. We then study a specific correlation model which involves a filter that is sparse in the time domain. While this sparsity assumption does not allow reducing the description cost in the universal case, we show that large gains can be achieved in the almost sure scenario by means of a novel distributed scheme based on annihilating filters. The robustness of the proposed method is also investigated.

Index Terms— Annihilating Filter, Compressive Sampling, Distributed Sensing, Sparse Reconstruction

1. INTRODUCTION

Consider two signals that are linked by an unknown filtering operation, where the filter is sparse in the time domain. Each signal is independently observed by a different sensor, which sends certain linear measurements of that signal to a central decoder. In a similar spirit to the Slepian-Wolf problem in distributed source coding [1], we study how the correlation induced by the above model can be exploited to reduce the number of measurements needed for perfect reconstruction, *without* any inter-sensor communication.

The present work is related to the distributed compressed sensing framework introduced by Baron *et al.* [2], where the authors studied the distributed sampling of signals correlated through certain joint sparsity models. The first contribution of this paper is a novel correlation model which relates two distributed signals through a sparse filtering operation, without requiring the signals themselves to be sparse in any domain. Such a model can be useful in describing the signal correlation in several practical scenarios, including, e.g., that between the transmitted and received signals in an unknown multipath environment. The second contribution is a concrete distributed sampling and reconstruction scheme that exploits the existing correlation, and recovers the original signals in an efficient and robust way.

Our goal in this paper is twofold. First, from a theoretical point of view, our interest is to determine the minimum number of measurements required at each sensor to allow for a perfect reconstruction of the original data. Second, from a practical perspective, we seek distributed sampling and reconstruction schemes that accomplish this task in a computationally efficient manner.

M. Vetterli is also with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720, USA. This research was supported by the Swiss National Science Foundation under grant numbers 5005-67322 (NCCR-MICS) and 200020-103729.

After a description of the considered model, we state in Section 2 a general formulation of the distributed sensing problem. In Section 3, we demonstrate a somewhat surprising result, namely, that if one requires *all* possible vectors to be reconstructed perfectly, the aforementioned correlation between the observed vectors cannot be exploited. In this case, the simple strategy of sending all the coefficients is optimal. However, if one only considers the perfect recovery of *almost all* vectors, then substantial gains can be achieved by means of a computationally efficient distributed algorithm based on annihilating filters [3, 4]. This scheme is presented in Section 4. Moreover, it is shown how the proposed method can be made robust to model mismatch using an iterative procedure due to Cadzow [5]. Finally, we present, in Section 5, some simulation results to illustrate the robustness of our scheme with respect to additive noise. Conclusions are given in Section 6.

2. SIGNAL MODEL AND PROBLEM STATEMENT

2.1. Proposed Correlation Model

Consider two signals $x_1(t)$ and $x_2(t)$, where $x_2(t)$ can be obtained as a filtered version of $x_1(t)$. In particular, we assume that

$$x_2(t) = (x_1 * h)(t), \quad (1)$$

where $h(t) = \sum_{k=1}^K c_k \delta(t - t_k)$ is a stream of K Diracs with *unknown* delays t_k and coefficients c_k . The above model can be used to characterize the correlation between a pair of signals of interest in various practical applications. Examples include the correlation between transmitted and received signals under multipath propagation with unknown channel condition, or the spatial correlation between signals recorded by two closely-spaced microphones in a simple acoustic environment composed of a single source.

In this work, we study the following finite-dimensional version of the model given in (1).

Definition 1 (Correlation Model) *The signals of interest are two vectors $\mathbf{x}_1 = (x_1[0], \dots, x_1[N-1])^T$ and $\mathbf{x}_2 = (x_2[0], \dots, x_2[N-1])^T$, linked to each other through a circular convolution*

$$x_2[n] = (x_1 \otimes h)[n] \quad \text{for } n = 0, 1, \dots, N-1, \quad (2)$$

where $\mathbf{h} = (h[0], \dots, h[N-1])^T \in \mathbb{R}^N$ is an unknown K -sparse vector, that is, $\|\mathbf{h}\|_0 = K$.

The model in (2) can be seen as an approximation of the original continuous model given in (1). In practice, the finite-dimensional signals \mathbf{x}_1 and \mathbf{x}_2 can be obtained from $x_1(t)$ and $x_2(t)$ through sampling and windowing, and the filtering operation in (1) may then be

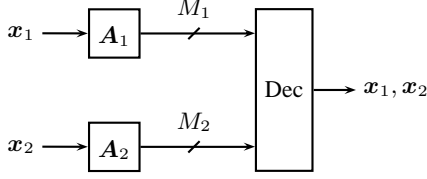


Fig. 1. Distributed sensing setup. Sensor i provides a M_i -dimensional version of its observation x_i by means of a linear transform A_i ($i = 1, 2$). The central decoder reconstructs the original vectors based on the received measurements.

approximated by the circular convolution through appropriate zero-padding. A popular and computationally efficient approach is to use a discrete Fourier transform (DFT) filter bank for this purpose.

In what follows, it is often convenient to use the notation of a stacked vector $\mathbf{x}^T \stackrel{\text{def}}{=} (x_1^T, x_2^T) \in \mathbb{R}^{2N}$. We denote by \mathcal{X} the set of all stacked vectors such that its components x_1 and x_2 satisfy (2) for some K -sparse vector \mathbf{h} .

2.2. Distributed Sensing and Problem Statement

We consider the problem of sensing $\mathbf{x}^T = (x_1^T, x_2^T)$ in a *distributed* fashion, by two independent sensors taking linear measurements of x_1 and x_2 , respectively. As depicted in Figure 1, suppose that the i th sensor ($i = 1, 2$) takes M_i linear measurements of x_i . We can write

$$\mathbf{y}_i = \mathbf{A}_i \mathbf{x}_i,$$

where $\mathbf{y}_i \in \mathbb{R}^{M_i}$ represents the vector of samples taken by the i th sensor, and \mathbf{A}_i is the corresponding sampling matrix. Considering the stacked vector $\mathbf{y}^T \stackrel{\text{def}}{=} (\mathbf{y}_1^T, \mathbf{y}_2^T)$, we have $\mathbf{y} = \mathbf{A} \mathbf{x}$, where \mathbf{A} is obtained as

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{0}_{M_1 \times N} \\ \mathbf{0}_{M_2 \times N} & \mathbf{A}_2 \end{bmatrix}. \quad (3)$$

Note that the block-diagonal structure of \mathbf{A} is due to the fact that x_1 and x_2 are processed separately. This is in contrast to the *centralized* scenario, where x_1 and x_2 can be processed jointly, in which case the matrix \mathbf{A} can be arbitrary.

The measurements \mathbf{y}_1 and \mathbf{y}_2 are transmitted to a central decoder, which attempts to reconstruct the vector \mathbf{x} through some (possibly nonlinear) mapping $\psi: \mathbb{R}^{M_1+M_2} \mapsto \mathcal{X}$ as

$$\hat{\mathbf{x}} = \psi(\mathbf{y}).$$

By analogy to the Slepian-Wolf problem in distributed source coding [1], the natural questions to pursue in the above sampling setup are the following:

1. What choices of sampling pairs (M_1, M_2) will allow us to reconstruct signals $\mathbf{x} \in \mathcal{X}$ from their samples?
2. What is the loss incurred by the distributed infrastructure over the centralized scenario in terms of the total number of measurements $M_1 + M_2$?
3. How can we reconstruct the original signals from their samples in a computationally efficient way?

In what follows, we first answer the above questions in the case of *universal reconstruction* (Section 3), that is, when *all* signals must be perfectly recovered. We prove that the sparsity assumption on \mathbf{h} does not allow reducing the minimum number of measurements. We then consider *almost sure reconstruction* (Section 4), when only

almost all signals must perfectly reconstructed. In this case, we demonstrate that the correlation between x_1 and x_2 can indeed be exploited by means of a computationally efficient algorithm. We further show how the proposed method can be made robust to noise and model mismatch. Finally, some simulation results are presented (Section 5).

3. UNIVERSAL RECOVERY

Let \mathbf{A}_1 and \mathbf{A}_2 be the sampling matrices used by the two sensors, and \mathbf{A} be the block-diagonal matrix defined in (3). We first focus on finding those \mathbf{A}_1 and \mathbf{A}_2 such that every $\mathbf{x} \in \mathcal{X}$ is uniquely determined by its sampling data $\mathbf{A} \mathbf{x}$.

Definition 2 (Universal Achievability) We say a sampling pair (M_1, M_2) is *achievable for universal reconstruction* if there exists fixed measurement matrices $\mathbf{A}_1 \in \mathbb{R}^{M_1 \times N}$ and $\mathbf{A}_2 \in \mathbb{R}^{M_2 \times N}$ such that the set

$$\mathcal{B}(\mathbf{A}_1, \mathbf{A}_2) \stackrel{\text{def}}{=} \{ \mathbf{x} \in \mathcal{X} : \exists \mathbf{x}' \in \mathcal{X} \text{ with } \mathbf{x} \neq \mathbf{x}' \text{ but } \mathbf{A} \mathbf{x} = \mathbf{A} \mathbf{x}' \} \quad (4)$$

is empty.

Intuition suggests that, due to the correlation between the vectors x_1 and x_2 , the minimum number of samples needed to perfectly describe all possible vectors \mathbf{x} can be made smaller than the total number of coefficients $2N$. The following proposition shows that, surprisingly, this is not the case.

Proposition 1 A sampling pair (M_1, M_2) is *achievable for universal reconstruction* if and only if $M_1 \geq N$ and $M_2 \geq N$.

Proof: Let us consider two stacked vectors $\mathbf{x}^T = (x_1^T, x_2^T)$ and $\mathbf{x}'^T = (x_1'^T, x_2'^T)$, each following the correlation model (2). They can be written under the form

$$\mathbf{x} = \begin{bmatrix} \mathbf{I}_N \\ \mathbf{C} \end{bmatrix} \mathbf{x}_1 \quad \text{and} \quad \mathbf{x}' = \begin{bmatrix} \mathbf{I}_N \\ \mathbf{C}' \end{bmatrix} \mathbf{x}'_1,$$

where \mathbf{C} and \mathbf{C}' are circulant matrices with vectors \mathbf{h} and \mathbf{h}' as the first column, respectively. It holds that

$$\mathbf{x} - \mathbf{x}' = \begin{bmatrix} \mathbf{I}_N & -\mathbf{I}_N \\ \mathbf{C} & -\mathbf{C}' \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}'_1 \end{bmatrix}.$$

Moreover, we have that

$$\text{rank} \begin{bmatrix} \mathbf{I}_N & -\mathbf{I}_N \\ \mathbf{C} & -\mathbf{C}' \end{bmatrix} = N + \text{rank}(\mathbf{C} - \mathbf{C}').$$

When $\mathbf{C} - \mathbf{C}'$ is of full rank, the above matrix is of rank $2N$. This happens, for example, when $K = 1$ with $\mathbf{C} = 2\mathbf{I}_N$ and $\mathbf{C}' = \mathbf{I}_N$. In this case, $\mathbf{x} - \mathbf{x}'$ can take any possible values in \mathbb{R}^{2N} . Hence, a necessary (and sufficient) condition for the set (4) to be empty is that the block-diagonal matrix \mathbf{A} is a $M \times 2N$ -dimensional matrix of full rank, with $M \geq 2N$. In particular, \mathbf{A}_1 and \mathbf{A}_2 must be full rank matrices of size $M_1 \times N$ and $M_2 \times N$, respectively, with $M_1, M_2 \geq N$. Note that, in the centralized scenario, the full rank condition would still require to take at least $2N$ measurements. ■

As a direct consequence of the above result, each sensor can process its vector independently without any loss of optimality. In particular, the simple strategy of sending all observed coefficients is optimal. Moreover, it is seen in the proof of Proposition 1 that there is no penalty associated with the distributed nature of the setup. In other words, the total number of measurements cannot be made smaller than $2N$ if the vectors x_1 and x_2 are processed jointly. The region of achievable sampling pairs is depicted in Figure 2.

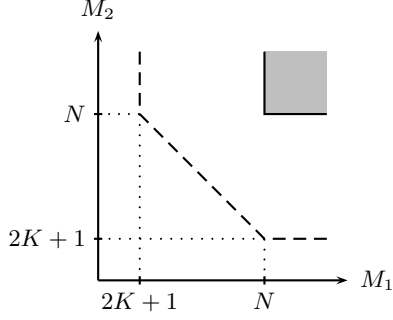


Fig. 2. Achievable sampling region for universal reconstruction (shaded area), and sampling pairs achieved by the proposed distributed sensing scheme for almost sure reconstruction (dashed line).

4. ALMOST SURE RECOVERY

As shown in Proposition 1, universal recovery is a rather strong requirement to satisfy since we have to take at least N samples at each sensor, without being able to exploit the existing correlation. In many situations, however, it is sufficient to consider a weaker requirement, which aims to find measurement matrices that permit the perfect recovery of *almost all* signals from \mathcal{X} . In this case, the set \mathcal{X} is endowed with a probability distribution.

Definition 3 (Almost Sure Achievability) *We say a sampling pair (M_1, M_2) is achievable for almost sure reconstruction if there exist fixed measurement matrices $\mathbf{A}_1 \in \mathbb{R}^{M_1 \times N}$ and $\mathbf{A}_2 \in \mathbb{R}^{M_2 \times N}$ such that the set $\mathcal{B}(\mathbf{A}_1, \mathbf{A}_2)$, as defined in (4), is of probability zero.*

In contrast to the universal scenario, we now demonstrate, by means of a novel distributed sensing algorithm, that the correlation between the vectors \mathbf{x}_1 and \mathbf{x}_2 can be exploited to reduce the total number of measurements needed.

4.1. Distributed Sensing Algorithm

The proposed distributed sensing scheme is based on a frequency-domain representation of the input signals. Let us denote by $\mathbf{X}_1 \in \mathbb{C}^N$ and $\mathbf{X}_2 \in \mathbb{C}^N$ the DFTs of the vectors \mathbf{x}_1 and \mathbf{x}_2 , respectively. The circular convolution (2) can be expressed as

$$\mathbf{X}_2 = \mathbf{H} \odot \mathbf{X}_1, \quad (5)$$

where $\mathbf{H} \in \mathbb{C}^N$ is the DFT of the filter \mathbf{h} and \odot denotes the element-wise product.

We first show how a decoder can almost surely recover the unknown filter \mathbf{h} using only the first $K+1$ DFT coefficients of \mathbf{x}_1 and \mathbf{x}_2 . This is achieved using an annihilating filter approach as follows. The DFT coefficients of the filter \mathbf{h} are given by

$$H[m] = \sum_{k=1}^K c_k e^{-j \frac{2\pi}{N} n_k m} \quad \text{for } m = 0, 1, \dots, N-1. \quad (6)$$

The sequence $H[m]$ is the sum of K complex exponentials, whose frequencies are determined by the positions n_k of the non-zero coefficients of the filter. It can be shown [6] that $H[m]$ can be annihilated by a filter $A[m]$ of degree K whose roots are of the form $e^{j2\pi n_k/N}$

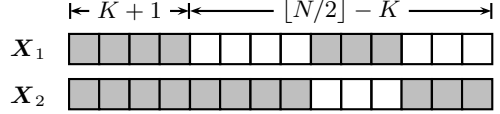


Fig. 3. Sensors 1 and 2 both send the first $K+1$ DFT coefficients of their observation, but only complementary subsets of the remaining frequency components (up to the Nyquist frequency).

Algorithm 1

- 1: Sensors 1 and 2 send the first $K+1$ DFT coefficients of \mathbf{x}_1 and \mathbf{x}_2 , respectively. They also send complementary subsets of the remaining DFT coefficients (up to the Nyquist frequency).
 - 2: The decoder computes $2K$ consecutive DFT coefficients of \mathbf{h} from (8).
 - 3: The decoder retrieves the filter \mathbf{h} using the annihilating filter method.
 - 4: The decoder reconstructs \mathbf{x}_1 and \mathbf{x}_2 from (5) using \mathbf{h} and the remaining DFT coefficients of \mathbf{x}_1 and \mathbf{x}_2 .
-

for $k = 1, 2, \dots, K$. The coefficients of this filter satisfy

$$A[m] * H[m] = \sum_{i=0}^K A[i] H[m-i] = 0,$$

or in matrix form,

$$\begin{bmatrix} H[0] & H[-1] & \cdots & H[-K] \\ H[1] & H[0] & \cdots & H[-K+1] \\ \vdots & \vdots & \vdots & \vdots \\ H[K-1] & H[K-2] & \cdots & H[-1] \end{bmatrix} \begin{bmatrix} A[0] \\ A[1] \\ \vdots \\ A[K] \end{bmatrix} = 0. \quad (7)$$

The above matrix is of size $K \times (K+1)$ and is built from $2K$ consecutive DFT coefficients. Moreover, it can be shown to be of rank K so that its null space is of dimension one. Therefore, we can find a solution to the above system by means of its singular value decomposition (SVD). Note that, due to the conjugate symmetry property, the coefficients of the matrix in (7) can be computed as

$$H[m] = \frac{X_2[m]}{X_1[m]} \quad \text{and} \quad H[-m] = H^*[m] \quad (8)$$

provided that $X_1[m]$ is non-zero for $m = 0, 1, \dots, K$. This happens almost surely if its distribution is, for example, continuous. Once the coefficients of the annihilating filter have been obtained, it is simply a matter of computing its roots to retrieve the unknown positions n_k . The filter weights c_k can then be recovered by means of the linear system of equations (6).

Based on the above considerations, our distributed sensing scheme can be described as follows. Both sensors send the first $K+1$ DFT coefficients of their signal to the decoder ($2K+1$ real values each). They also transmit complementary subsets of frequency indices up to the Nyquist frequency ($N-2K-1$ real values in total). This is illustrated in Figure 3. The first $K+1$ DFT coefficients allow to almost surely reconstruct the filter \mathbf{h} . The missing frequency components of \mathbf{x}_1 (resp. \mathbf{x}_2) are then recovered from the available DFT coefficients of \mathbf{x}_2 (resp. \mathbf{x}_1) using the relation (5). The method is summarized in Algorithm 1. In terms of achievability, we have thus shown the following result.

Proposition 2 A sampling pair (M_1, M_2) is achievable for almost sure reconstruction if

$$\begin{aligned} M_1 &\geq \min \{2K + 1, N\}, \\ M_2 &\geq \min \{2K + 1, N\}, \\ \text{and } M_1 + M_2 &\geq \min \{N + 2K + 1, 2N\}. \end{aligned}$$

In contrast to the universal reconstruction, the total number of measurements can be reduced from $2N$ to $N + 2K + 1$. Typically, $K \ll N$, such that a large gain is achieved by the proposed distributed sensing scheme. This is depicted in Figure 2.

It should be emphasized that Proposition 2 only provides an achievability result, that is, its conditions are sufficient but not necessary. In fact, it can be shown that, if N and $K + 1$ are not both even, a combinatorial approach allows reducing the factor $2K$ in the above lower bounds to K . Note also that the proposed correlation model can be extended to filters which admit a sparse representation in an arbitrary basis. In this case, one can use sampling techniques which involve taking random frequency measurements, and reconstruction methods based on ℓ_1 minimization, as presented in [7].

4.2. Robustness to Model Mismatch

Noise or, more generally, model mismatch makes the solution discussed previously rarely directly applicable in practice. Instead, robustness must be included by sending additional measurements. This redundancy allows solving (7) using a total least square approach and then finding the mean square solution to (6). To further improve robustness, we resort to an iterative method devised by Cadzow [5]. In our context, it can be summarized as follows. Sensor i transmits the first $L + 1$ DFT coefficients of x_i ($i = 1, 2$) with $L \geq K$. A matrix of dimension $L \times (L + 1)$ of the form (7) is then built from these measurements. In the noiseless case, this matrix has two key properties: (i) it is of rank K and (ii) it is of Toeplitz form. In the noisy case, these two properties can be enforced by alternatively performing the two following steps:

- (i) Enforce rank K by setting the $L - K$ smallest singular values to zero.
- (ii) Enforce Toeplitz form by averaging the coefficients along the diagonals.

The above procedure is guaranteed to converge to the matrix which exhibits the desired properties and is the closest in Frobenius norm to the noisy one [5]. The denoised filter DFT coefficients are then extracted from the first row and first column, and used to solve the systems (6) and (7), as mentioned above.

5. SIMULATION RESULTS

To assess the robustness of the proposed scheme, we evaluate it with respect to additive noise. We assume that the filter is of length $N = 128$ but has only $K = 2$ non-zero coefficients. Independent white Gaussian noise is added to the filter to meet a desired signal-to-noise ratio (SNR). Figure 4 shows the mean squared error (MSE) on the reconstruction of the filter as a function of the SNR, with and without Cadzow's denoising procedure. The results have been averaged over 20'000 realizations. The different sets of curves correspond to $L = 3, 6, 14$ (top to bottom). We observe that the MSE can be significantly reduced by sending just a few more measurements than the minimum required. Moreover, the gain provided by Cadzow's procedure increases as the number of transmitted coefficients increases.

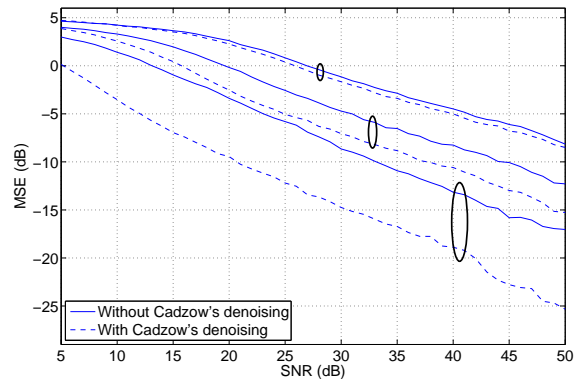


Fig. 4. Reconstruction error of the filter in additive white Gaussian noise, with and without Cadzow's denoising. The parameters are $N = 128$ and $K = 2$. The different sets of curves correspond to $L = 3, 6, 14$ (top to bottom).

6. CONCLUSIONS

A general formulation of the distributed sensing problem has been proposed. The setup has been studied in more details for a specific correlation model involving a sparse filter. In this context, both universal and almost sure reconstruction were addressed. A distributed sensing scheme was presented, together with a method to make it robust to model mismatch. Current research efforts focus on the application of the proposed method to the distributed coding of spatial audio, as well as the determination of achievable sampling regions for other correlation models of practical interest.

7. REFERENCES

- [1] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inform. Theory*, vol. 19, pp. 471–480, Jul. 1973.
- [2] D. Baron, M. B. Wakin, M. F. Duarte, S. Sarvotham, and R. G. Baraniuk, "Distributed compressed sensing," Tech. Rep. ECE-0612, Electrical and Computer Engineering Department, Rice University, Dec. 2006.
- [3] P. Stoica and R. Moses, *Introduction to Spectral Analysis*, Prentice-Hall, Englewood Cliffs, NJ, 2005.
- [4] M. Vetterli, P. Marziliano, and T. Blu, "Sampling signals with finite rate of innovation," *IEEE Trans. Signal Processing*, vol. 50, no. 6, pp. 1417–1428, Jun. 2002.
- [5] J. A. Cadzow, "Signal enhancement – A composite property mapping algorithm," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, no. 1, pp. 49–67, Jan. 1988.
- [6] T. Blu, P.L. Dragotti, M. Vetterli, P. Marziliano, and L. Coulot, "Sparse sampling of signal innovations," *IEEE Signal Processing Mag.*, vol. 25, no. 2, pp. 31–40, 2008.
- [7] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.