Distributed Sensing of Noisy Signals by Thresholding of Redundant Expansions

Karin Schnass, Pierre Vandergheynst and Pascal Frossard
Ecole Polytechnique Fédérale de Lausanne (EPFL)
Signal Processing Institute - ITS
CH - 1015 Lausanne, Switzerland
Email: {karin.schnass,pierre.vandergheynst,pascal.frossard}@epfl.ch

***Abstract*— This paper addresses the problem of sensing or recovering a signal s, captured by distributed low-complexity sensors. Each sensor observes a noisy version of the signal of interest, and independently forms an approximant of its observation. This approximant is sent to a central decoder that tries to recover the input signal by combining the multiple sensor outputs. We propose to use redundant dictionaries, and thresholding in the sensor nodes, in order to form sparse approximants of the noisy observations, with low computational complexity. We first show that the noise can actually be beneficial in the recovery of the correct components of the signal s, since it can advantageously perturb the naive thresholding scheme. Then we illustrate the benefit of multiple observations with uncorrelated noise. By careful reconstruction with a Projection Onto Convex Sets (POCS) strategy, each additional measurement actually helps to recover more and more components of the original signal, since it tends to isolate the common part in all observations. Experimental results demonstrate the interesting recovery performance of our distributed sensing system. They show that a few observations, represented by a small number of components, are able to provide a good approximation of the signal, even in very noisy conditions.**

I. Introduction

The increasing availability of low-power micro-sensors or embedded processors has lead to the development of new applications of distributed sensing in numerous fields, from physical sensing to monitoring or medial applications. Due to power and bandwidth limitations in such scenarios, it becomes primordial to develop efficient signal processing and sensing algorithms, which allow the system to exploit the redundancy of sensors to compensate for the ad-hoc deployment of such cheap systems.

In this framework, we here address the problem of distributed sensing in noisy environments. Equivalently, we are interested in recovering a signal of interest, or at least its most important features, from multiple noisy observations recorded by distributed low-complexity sensors. To benefit from increased design flexibility in forming a short approximant of its observation, each sensor implements a simple thresholding of the signal over a redundant dictionary. The multiple approximants are then combined at the central decoder with a Projection Onto Convex Sets (POCS) strategy [1], to generate an improved approximation of the input signal.

We analyze the recovery conditions in such a distributed sensing scenario, and we show that additive noise might become beneficial in naive thresholding, since it can help to recover the correct components of the original signal. Similarly, from multiple observations with uncorrelated noise, the central decoder is able to efficiently refine the approximation of the input signal. The experimental results show that simple thresholding leads to a very interesting recovery performance, even when the sensors generate very short approximants in quite noisy environments. Moreover, the proposed method does not need any a priori model of the signal, nor any assumption on the nature of the noise, which may prove to be quite interesting compared to typical detection scenarios based on statistical frameworks.

II. Distributed recovery with overcomplete representations

A. Problem Formulation

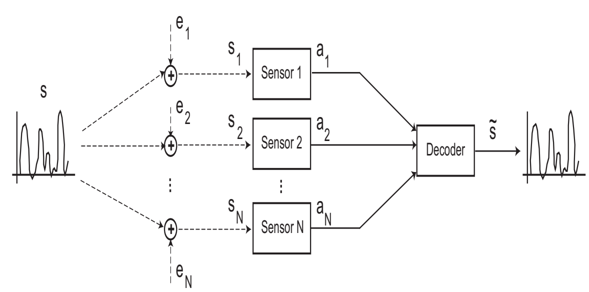

Fig. 1. Distributed sensing framework.

We consider the distributed sensing framework illustrated in Figure 1. A set of N sensors observes a signal s, which is a vector in the d-dimensional real vector space $\mathbb{R}^d$. Each sensor captures a signal s_i that is a possibly noisy version of s, i.e., $s_i = s + e_i$. It is important to note that, at this stage, we make no assumption on the statistics of the noise, its nature, or the correlation between observations s_i. Each sensor independently sends an approximation a_i of the observation s_i to a central node. The central node then reconstructs an approximation $\hat{s}$ of the signal s, from the multiple descriptions sent by the sensors. We are interested in minimizing the approximation error $\delta = |s - \hat{s}|_2$, or equivalently in maximizing the probability of recovering the signal s, based on the distributed observations s_i.

In order to build short meaningful representations at each sensor, and for increased flexibility, we propose to use redundant dictionaries for signal expansion. The terminology dictionary refers to a large collection of unit vectors $\varphi_k \in \mathbb{R}^d$ sometimes called atoms. It is convenient to arrange these atoms along the columns of a large $d \times N$ matrix $\mathbf{\Phi}$ called the *synthesis matrix* of the dictionary $\mathcal{D}$. The dictionary is redundant when the cardinality N is much larger than the signal dimension : $N \gg d$.

We assume that the signals can be represented as very sparse expansions over the dictionary $\mathcal{D}$. In other words, the signal s is built on a subset of atoms I, whose cardinality is small compared to d : $|I| \ll d$. Thus, we write any such s using a submatrix $\mathbf{\Phi}_I$ of $\mathbf{\Phi}$, applied to a vector of coefficients c :

$$s = \mathbf{\Phi}_I c\,.$$

Similarly, the noisy observations of s, captured by the N independent sensors, can be written as :

$$s_j = \mathbf{\Phi}_I c + e_j, \text{ with } j = 1, ..., N. \tag{1}$$

B. Recovery with distributed thresholding

The problem is now to form good sparse approximations a_k of the observations s_k, such that the signal of interest s, can be efficiently recovered at the central node. Since sensors are generally low power devices, we restrict ourselves to distributed processing methods that can be implemented with simple algorithms, and we shift the computational complexity to the decoder end. One of the simplest algorithms that can form an approximant a_k consists of a thresholding of the observation s_k. Given s_k, the sensor computes all projections with the dictionary $v = \mathbf{\Phi}^\star s_k$, where v is a N-dimensional vector, and $\mathbf{\Phi}^\star$ represents the transpose matrix of $\mathbf{\Phi}$. Each sensor is set to keep a fixed number L of components which simply correspond to the L largest entries in v, in absolute value. Let I_k denote the set of components selected in that way. An approximant a_k is then built by computing the orthogonal projection of s_k on I_k,

$$a_k = \mathbf{\Phi}_{I_k}^\dagger \mathbf{\Phi}_{I_k}^\star s_k\,, \tag{2}$$

where $\mathbf{\Phi}_{I_k}^\dagger$ is the pseudo-inverse of $\mathbf{\Phi}_{I_k}$ [2]. The current sensor then sends the estimated components I_k and their relative contributions c_k (or equivalently, the approximant a_k) to the receiver. This one finally builds an approximation of s using the POCS (Projection Onto Convex Sets) algorithm, i.e., by successively enforcing constraints given by the different observations of s.

III. Recovery conditions

Thresholding is a simple but very naive algorithm that does not take into account the correlations among elements of the dictionary. Indeed, in the noiseless case $e_k = 0$, if Φ is an orthogonal matrix (i.e., the dictionary is an orthogonal basis), I_k would simply list the strongest components of s and a_k would be an optimal M-term approximation of the signal. This is not the case anymore for a redundant dictionary : the algorithm can actually select wrong components, even in this simple case! There are thus two important points to be investigated towards the design of efficient distributed recovery :

1) Given the correct components I and the noise e_k, when does thresholding correctly recover components of I in I_k ?
2) Knowing that we have multiple noisy observations of s, under which conditions do they combine to improve the performance of the algorithm ?

The first point has already been addressed in [3] and we will briefly review this result. Given a noisy sparse signal of the form,

$$s = \mathbf{\Phi}_I x + e\,,$$

the behaviour of thresholding is influenced by two sources of confusion. The first one is the amount of correlation between the elements of I and the dictionary and it can be characterized by the dictionary's Inter Symbol Interference [3]:

$$\text{ISI}(\mathbf{\Phi}, I) := \mu_1\big(\mathbf{\Phi}, I\big) + \sup_{l \in I} \mu_1\big(\mathbf{\Phi}_I, I/\{l\}\big). \tag{3}$$

where

$$\mu_1(\mathbf{\Phi}, I) := \sup_{k \notin I} \sum_{i \in I} |\langle \varphi_k, \varphi_i \rangle|. \tag{4}$$

is the so called *setwise Babel function* defined in [4]. In the following it will be convenient to use the notation μ_{out} and μ_{in} for both terms on the rhs of (3). The second source of confusion is the noise e, which in general can influence both the correct support I and its complementary set $\overline{I}$. It was proved in [3] that a necessary condition for recovering a component φ_m in (1) is that its coefficient x_m satisfies :

$$\frac{|x_m|}{\|x\|_\infty} > \frac{\|\mathbf{\Phi}_{\overline{I}} e\|_\infty + \|\mathbf{\Phi}_I e\|_\infty}{\|x\|_\infty} + \text{ISI}(\mathbf{\Phi}, I). \tag{5}$$

Condition (5) is a little too pessimistic as it considers any influence of the noise as harmful. However a finer inspection of the noise shows that it may sometimes have positive effects. We will assume from now on that none of the components (except for the largest) of an exactly sparse signal can be identified because their coefficients fail the noiseless recovery condition, i.e.,

$$\frac{|x_m|}{\|x\|_\infty} \leq \text{ISI}(\mathbf{\Phi}, I).$$

Then adding noise is the only hope to correctly identify (parts of) the support as it can 'boost' normally undetectable components to a detectable level. Without loss of generality, we shall split the noise into a component on the optimal support I and its orthogonal complement : $e = \mathbf{\Phi}_I b + e_r$ where $\langle e_r, \mathbf{\Phi}_I b \rangle = 0$. The following two propositions characterize the kind of noise that can have a positive influence.

Proposition 1: Let $y = \mathbf{\Phi}_I x + e$ be a noisy sparse expansion. If a component φ_m fails the noiseless recovery condition by a margin δ, i.e.,

$$\frac{|x_m|}{\|x\|_\infty} \geq \mu_{out} + \mu_{in} - \delta$$

then any noise satisfying

$$b_m \cdot x_m > 0$$
$$|b_m| > \delta\|x\|_\infty + \mu_{in}\|b\|_\infty + \|\mathbf{\Phi}^\star_{\overline{I}} e\|_\infty$$

will guarantee the selection of this component. Conversely if a component φ_m satisfies the noiseless thresholding condition with a margin δ, i.e.,

$$\frac{|x_m|}{\|x\|_\infty} - \delta \geq \mu_{out} + \mu_{in}$$

then any noise satisfying

$$b_m \cdot x_m < 0$$
$$|b_m| < \delta\|x\|_\infty - \mu_{in}\|b\|_\infty - \|\mathbf{\Phi}_{\overline{I}} e\|_\infty$$

will not disturb the recovery of φ_m.

Proof: By definition of the noise we have:

$$\begin{aligned}
|\langle y, \varphi_m\rangle| &\geq |x_m + b_m| - \|(\mathbf{\Phi}^\star_{\overline{I}}\mathbf{\Phi}_I - \mathbf{Id})(x+b)\|_\infty \\
&\geq |x_m| + |b_m| - \mu_{in}\|b\|_\infty + \mu_{in}\|x\|_\infty \\
&\geq |x_m| + \delta\|x\|_\infty + \|\mathbf{\Phi}^\star_{\overline{I}} e\|_\infty + \mu_{in}\|x\|_\infty \\
&\geq \|x\|_\infty \mu_{out} + \|\mathbf{\Phi}^\star_{\overline{I}} e\|_\infty \\
&\geq \|\mathbf{\Phi}^\star_{\overline{I}}\mathbf{\Phi}_I x\|_\infty + \|\mathbf{\Phi}^\star_{\overline{I}} e\|_\infty \\
&\geq \|\mathbf{\Phi}^\star_{\overline{I}}\mathbf{\Phi}_I x + \mathbf{\Phi}^\star_{\overline{I}} e\|_\infty = \max_{k\in\overline{I}} |\langle y, \varphi_k\rangle|
\end{aligned}$$

Analogously,

$$\begin{aligned}
|\langle y, \varphi_m\rangle| &\geq |x_m| - |b_m| - \mu_{in}\|b\|_\infty + \mu_{in}\|x\|_\infty \\
&\geq |x_m| - \delta\|x\|_\infty + \|\mathbf{\Phi}^\star_{\overline{I}} e\|_\infty + \mu_{in}\|x\|_\infty \\
&\geq \|x\|_\infty \mu_{out} + \|\mathbf{\Phi}^\star_{\overline{I}} e\|_\infty \\
&\geq \|\mathbf{\Phi}^\star_{\overline{I}}\mathbf{\Phi}_I x + \mathbf{\Phi}^\star_{\overline{I}} e\|_\infty = \max_{k\in\overline{I}} |\langle y, \varphi_k\rangle|
\end{aligned}$$

■

This first kind of noise helps because it increases the coefficient x_m of a component. However even noise that does not increase any coefficient but simply leaves some of them uninfluenced may be beneficial. We then have the following proposition.

***Proposition** 2:* Assume that a component φ_m fails the noiseless recovery condition by δ. Then any noise satisfying:

$$\alpha = \|x\|_\infty - \|x+b\|_\infty \geq 0$$
$$b_m = 0$$
$$\alpha \cdot \frac{|x_m|}{\|x\|_\infty} > \delta\|x+b\|_\infty + \|\mathbf{\Phi}^\star_{\overline{I}} e\|_\infty$$

will guarantee the selection of φ_m.

Proof:

$$\begin{aligned}
|\langle y, \varphi_m\rangle| &\geq |x_m| - \|(\mathbf{\Phi}^\star_{\overline{I}}\mathbf{\Phi}_I - \mathbf{Id})(x+b)\|_\infty \\
&\geq \|x+b\|_\infty \frac{|x_m|}{\|x\|_\infty} + \alpha\frac{|x_m|}{\|x\|_\infty} - \mu_{in}\|x+b\|_\infty \\
&\geq \|x+b\|_\infty \left(\frac{|x_m|}{\|x\|_\infty} - \mu_{in}\right) + \delta\|x+b\|_\infty + \|\mathbf{\Phi}^\star_{\overline{I}} e\|_\infty \\
&\geq \|x+b\|_\infty \mu_{out} + \|\mathbf{\Phi}^\star_{\overline{I}} e\|_\infty \\
&\geq \|\mathbf{\Phi}^\star_{\overline{I}}\mathbf{\Phi}_I x + \mathbf{\Phi}^\star_{\overline{I}} e\|_\infty = \max_{k\in\overline{I}} |\langle y, \varphi_k\rangle|
\end{aligned}$$

■

Interestingly, we now show that, even if the noiseless signal cannot be estimated correctly with a single sensor, the conjunction of multiple measurements and noise helps in recovering additional components of the signal s. When the noise contaminating our observations s_k satisfies any of the situations depicted above, some signal components get a sufficient boost and become detectable. Since the noise varies from sensor to sensor, the receiver thus collects more and more correct components of the original signal as the systems keeps on recording new observations. One way of exploiting this phenomenon is to combine the approximants a_k into a single better approximation $\tilde{s}$ of the original signal s. Remember that each a_k is an orthogonal projection of the noisy signal s_k onto the estimated supports I_k, each of which contains some correct components of the original support I. Our signal is thus at the intersection of K linear constraints of the form

$$\mathbf{\Phi}^\dagger_{I_k}\mathbf{\Phi}^\star_{I_k} s = a_k,\ k = 1, ..., K.$$

and we can recover s using the POCS (Projection Onto Convex Sets) algorithm by enforcing each constraint in turn. Note that the algorithm will converge to the correct signal if and only if $\cup_k I_k = I$. Spurious components may still cause the algorithm to diverge. However, if $\cup_k I_k \subset I$, then we get a combined approximant $\tilde{s}$ that is better than any of the a_k's.

IV. Experimental results

We now describe experimental results obtained using the algorithm described in the previous section.

Our experimental protocol was designed as follows.

1) **Dictionary and signals:** For these experiments, we have constructed a random dictionary of 400 vectors uniformly spread over the unit sphere in $\mathbb{R}^{50}$. A sparse support I of length $M = 8$ has been selected at random in the dictionary and a signal is built using random coefficients c. In our experiments, we have used either zero-mean gaussian coefficients or binary coefficients with random signs.
2) **Sensor Field:** A large number K of noisy measurements s_k was built by adding gaussian noise with different standard deviations to the signal. These observations were then independently processed by thresholding, set to keep $L \leq M$ coefficients. Finally K approximants a_k were constructed.

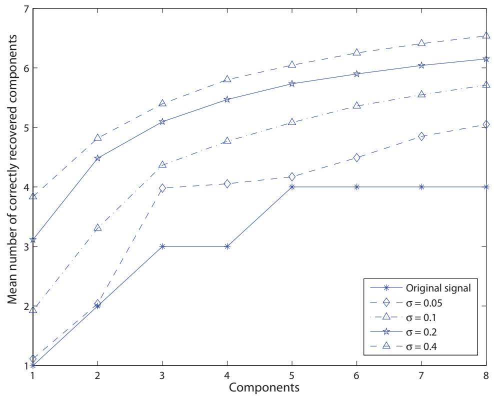

Fig. 2. Average number of correctly recovered components as a function of the number of selected components L in each observation. $K = 20$ observations and the noise level set at $\sigma = 0.05, 0.1, 0.2, 0.4$.

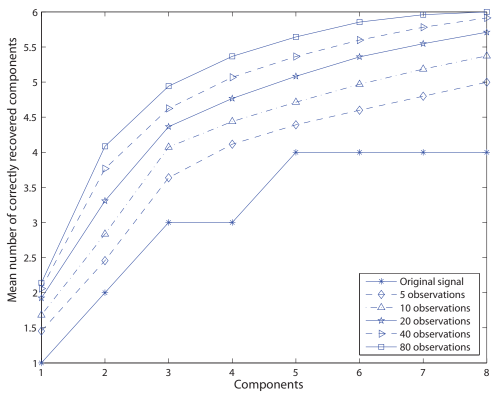

Fig. 3. Average number of correctly recovered components as a function of the number of selected components L in each observation. Noise level fixed at $\sigma = 0.2$ and the number of observations varying between $K = 5$ and $K = 80$.

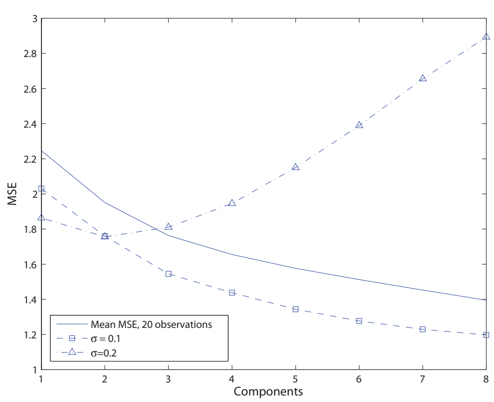

Fig. 4. MSE as a function of the number of selected components in each observation. The solid line is mean MSE over all observations, the other curves are MSE of the POCS approximant for $\sigma = 0.1$ and $\sigma = 0.2$.

3) **Receiver:** POCS is starting with the first approximant as initial condition. A fixed number of 10 iterations is performed to get the final combined approximant $\tilde{s}$.

We measured the performance of the algorithm by computing the euclidean distance between $\tilde{s}$ and s, and the number of correctly recovered components (see Figures 2 to 4). All of our results are averaged over 1000 randomly selected signals and for various values of M, K and noise standard deviation σ. As one can see on the first graph, increasing the noise allows to discover more and more correct components, though the rate of discovery clearly saturates. On the other hand one should not blindly accept this result as purely positive as the number of wrong components also increases. There is a trade-off, as can be seen from the quality of the POCS approximant on the last graph : for $\sigma = 0.1$, POCS gives a very good approximant, that is always significantly better than that of any single channel. However if we increase the noise to $\sigma = 0.2$, one sees that POCS quickly derails and the MSE actually explodes. This can be understood easily : with more noise, we get a lot of wrong components. POCS is then going to assume that the signal lives in a wrong subspace of $\mathbb{R}^d$ and will simply get lost as we keep on feeding it with wrong directions. It is also interesting to note the behaviour of the algorithm as a function of the number of observations. As we increase K, the number of correct components that are discovered discovered increases steadily, but also tends to saturate since s is a sparse signal. Note as well that in a very noisy situation, more observations also mean a higher probability of discovering wrong atoms. Here again we find a trade-off between *bias*, that is the number of wrong atoms, and *variance*, that is the proportion of wrong atoms.

V. Conclusions

This paper proposes a distributed sensing framework, where sparse noisy expansions are efficiently combined by a central decoder to yield efficient performance in recovering the signal of interest. Sensors implement a low-complexity thresholding of their observations over a redundant dictionary of atoms, where additive noise plays an interesting role in helping to recover the correct signal components. In addition, multiple observations with uncorrelated noise help to recover yet additional components of the original signal. Our low-complexity distributed recovery strategy interestingly borrows from the same perturbation principles used in dithering schemes and allows for efficient recovery performance, even with very short approximations and high noise environments. Overall, the proposed system appears quite promising for typical low-complexity sensor environments with stringent bandwidth constraints.

References

[1] P. B. Gubin L.G. and R. E.V., "The method of projections for finding the common point of convex sets," *USSR Computational Mathematics and Mathematical Physics*, vol. 7, no. 6, pp. 1–24, 1967, 643.

[2] R. Horn and C. Johnson, *Matrix Analysis*. Cambridge Univ. Press, 1985.

[3] R. Gribonval, M. Nielsen, and P. Vandergheynst, "Stable recovery of structured sparse approximations: A priori and a posteriori analysis," 2005.

[4] R. Gribonval and M. Nielsen, "Beyond sparsity: Recovering structured representations by l1 minimization and greedy algorithms." IRISA, Rennes, Publication interne 1684, 2005. [Online]. Available: http://www.irisa.fr/bibli/publi/pi/2005/1684/1684.html