Average case analysis of multichannel sparse approximations using p-thresholding

Karin Schnass, Pierre Vandergheynst[a], Rémi Gribonval, Boris Mailhe[b] and Holger Rauhut[c]

[a]Ecole Polytechnique Fédérale de Lausanne (EPFL), Signal Processing Institute, Station 11, CH- 1015 Lausanne, Switzerland;
[b]IRISA, Campus de Beaulieu, F-35042, Rennes cedex, France;
[c]NuHAG, Faculty of Mathematics, University of Vienna, Nordbergstrasse 15, A-1090 Wien, Austria.

ABSTRACT

This paper introduces p-thresholding, an algorithm to compute simultaneous sparse approximations of multichannel signals over redundant dictionaries. We work out both worst case and average case recovery analyses of this algorithm and show that the latter results in much weaker conditions on the dictionary. Numerical simulations confirm our theoretical findings and show that p-thresholding is an interesting low complexity alternative to simultaneous greedy or convex relaxation algorithms for processing sparse multichannel signals with balanced coefficients.

1. OUR PROBLEM AND AN ALGORITHM TO SOLVE IT

Suppose we are to design a network of N sensors monitoring a common phenomenon. Each of our sensors observes a d-dimensional signal $y_n \in \mathbb{R}^d$, $n = 1, ..., N$, but our set of signals obey a strong sparsity hypothesis : we will assume that each y_n admits a sparse approximation over a single dictionary $\mathbf{\Phi}$:

$$y_n = \mathbf{\Phi} x_n + e_n,\ n = 1, ..., N.$$

In order to model correlations between signals, we will refine this model by imposing that all signals share a common sparse support, i.e.

$$y_n = \mathbf{\Phi}_\Lambda x_n + e_n,$$

where $\mathbf{\Phi}_\Lambda$ is the restriction of the synthesis matrix $\mathbf{\Phi}$ to the columns listed in the set Λ. This model is inspired by a recent series of papers on distributed sensing, see[1] and references therein. It describes a network of sensors monitoring a signal with a strong global component that appears at each node. Localized effects are modeled by letting synthesis coefficients $x_n \in \mathbb{R}^S$, $S := |\Lambda|$, vary across nodes and through the noise e_n. In order to obtain a sufficiently general model, we will assume that the components $x_n(k)$ of the random vector x_n are independent Gaussian variables of variance α_k. This model is fairly general to accommodate various practical problems: the Gaussian assumption is one of the most widely used in signal processing, while incorporating different variances allows us to shape the synthesis coefficients, imposing statistical decay for example on the $x_n(k)$.

In order to simplify our analysis we will adopt a global matrix notation. We will collect all signals y_n on the columns of the $d \times N$ matrix Y and the synthesis coefficients x_n on the columns of the $S \times N$ matrix X. Let U be a $S \times N$ random matrix with independent standard Gaussian entries and let D be a $S \times S$ diagonal matrix whose entries are positive real numbers α_k. Our model can then be written in compact form

$$Y = \mathbf{\Phi}_\Lambda X + E = \mathbf{\Phi}_\Lambda D U + E, \tag{1}$$

where E is a $d \times N$ matrix collecting noise signals e_n on its columns. The problem we will face in this paper is to recover the joint support Λ by sensing the set of signals in a very simple way.

Let us now describe more precisely that sensing algorithm. The observed signals y_n are sent to a central processing unit that tries to recover the common sparse support Λ. The problem thus boils down to estimating the joint sparse support of a set of signals generated from a redundant dictionary $\mathbf{\Phi}$. A number of algorithms have been proposed lately to jointly process sparse signals, most of them based on multichannel generalizations of greedy algorithms[2] or convex relaxation algorithms. A common weakness to all these techniques is a high computational complexity. To overcome this problem, we would like to resort here to one of the simplest possible algorithms: thresholding. More precisely, our algorithm computes the p-norm of the correlation of the multichannel signal Y with the atoms ψ_k of a sensing dictionary $\mathbf{\Psi}$:

$$\|\psi_k^\star Y\|_p^p := \sum_{n=1}^{N} |\langle \psi_k, y_n \rangle|^p.$$

The sensing dictionary $\mathbf{\Psi}$ has the same cardinality as $\mathbf{\Phi}$, so the atoms in both dictionaries are in a one-to-one relationship. We could set $\mathbf{\Psi} \equiv \mathbf{\Phi}$, but we voluntarily keep the possibility of optimizing both dictionaries in the spirit of.[3]

Define Λ_S, the set of indices k with the S largest p-norms. This algorithm is successful if for $S = \sharp\Lambda$ we have $\Lambda_S = \Lambda$. Since $\mathbf{\Psi}^\star Y = \mathbf{\Psi}^\star \mathbf{\Phi}_\Lambda X + \mathbf{\Psi}^\star E$, the strongest p-norm of projections on the set $\overline{\Lambda}$ of bad atoms is

$$\|\mathbf{\Psi}_{\overline{\Lambda}}^\star Y\|_{p,\infty} \leq \|\mathbf{\Psi}_{\overline{\Lambda}}^\star \mathbf{\Phi}_\Lambda X\|_{p,\infty} + \|\mathbf{\Psi}_{\overline{\Lambda}}^\star E\|_{p,\infty},$$

where the (p,∞)-norm of a matrix $\|\mathbf{M}\|_{p,\infty}$ is defined as the maximum of the p-norms of its rows. Conversely, the smallest p-norm of projections on the set of good atoms reads

$$\min_{i\in\Lambda} \|\psi_i^\star Y\|_p \geq \min_{i\in\Lambda} \|\psi_i^\star \mathbf{\Phi}_\Lambda X\|_p - \|\mathbf{\Psi}_\Lambda^\star E\|_{p,\infty}.$$

and the algorithm will thus succeed as soon as

$$\begin{aligned} \min_{i\in\Lambda} \|\psi_i^\star \mathbf{\Phi}_\Lambda X\|_p - \|\mathbf{\Psi}_{\overline{\Lambda}}^\star \mathbf{\Phi}_\Lambda X\|_{p,\infty} >& \|\mathbf{\Psi}_\Lambda^\star E\|_{p,\infty} \\ &+ \|\mathbf{\Psi}_{\overline{\Lambda}}^\star E\|_{p,\infty}. \end{aligned} \tag{2}$$

2. WORST CASE BEHAVIOUR OF P-THRESHOLDING

The recovery condition (2) can be checked based on simple characteristics of the multichannel signals and the dictionaries. To capture the requirements on the dictionary we need to define $\beta := \min_{i\in\Lambda} |\langle \psi_i, \varphi_i \rangle|$ the minimum correlation between sensing and synthesis atoms, and to adapt the definition of the standard cumulative coherence:[4]

$$\mu_q(\mathbf{\Psi}, \mathbf{\Phi}, \Lambda) := \sup_{l\notin\Lambda} \|\mathbf{\Phi}_\Lambda^\star \psi_l\|_q = \sup_{l\notin\Lambda} \left(\sum_{i\in\Lambda} |\langle \psi_l, \varphi_i \rangle|^q \right)^{1/q}. \tag{3}$$

As for properties of the signal we need to define the p-Peak SNR and the dynamic range R_p:

$$\begin{aligned} \mathrm{PSNR}_p &:= \frac{\|\mathbf{\Psi}_{\overline{\Lambda}}^\star E\|_{p,\infty} + \|\mathbf{\Psi}_\Lambda^\star E\|_{p,\infty}}{\|X\|_{p,\infty}}, \\ R_p &:= \frac{\min_{i\in\Lambda} \|X(i)\|_p}{\|X\|_{p,\infty}}, \end{aligned}$$

where we denote $\|X(i)\|_p = (\sum_{n=1}^{N} |x_n(i)|^p)^{1/p}$ the p-norm of the i-th row of X. Following the analysis in,[5] it is easy to check that the following condition implies (2):

$$\begin{aligned} &\mu_1(\mathbf{\Psi}, \mathbf{\Phi}, \Lambda) + \sup_{i\in\Lambda} \mu_1(\mathbf{\Psi}_\Lambda, \mathbf{\Phi}_\Lambda, \Lambda/\{i\}) \\ &\qquad < \beta \cdot R_p - \mathrm{PSNR}_p. \end{aligned} \tag{4}$$

The success of p-thresholding is thus governed by the condition that the dynamic range of the signal should be bigger than the noise level and the sum of correlations among atoms on the support and between the support and the remaining of $\mathbf{\Phi}$. We note that μ_1 can be very big even for reasonably small Λ. For example, when

$\mathbf{\Psi} = \mathbf{\Phi}$, the quantity $\mu_1(\mathbf{\Psi}, \mathbf{\Phi}, \Lambda) + \mu_1(\mathbf{\Psi}_\Lambda, \mathbf{\Phi}_\Lambda, \Lambda/\{i\})$ is often replaced by its upper estimate $(2S-1)\mu$. The r.h.s in (4) is at most one, so the resulting condition can only be satisfied when $S < (1+\mu^{-1})/2$. In the next sections, we develop an average case analysis of p-thresholding and show that the *typical* recovery conditions are much less restrictive.

The central contribution of this paper is to show that the simple p-thresholding algorithm will succeed in recovering the correct support Λ with high probability. As we will see below, the sparsity constrain is expressed in terms of the 2-cumulative coherence μ_2 and is thus much weaker than worst case conditions that are usually expressed in terms of μ_1. Moreover, the recovery probability scales exponentially with the number of channels.

3. AVERAGE CASE ANALYSIS OF P-THRESHOLDING

To state our central theoretical result for the average case we need to define a probabilistic PSNR and dynamic range, remember we had $Y = \mathbf{\Phi}_\Lambda DU + E$ where $D = \mathbf{diag}(\alpha_i)$,

$$\begin{aligned} \overline{\mathrm{PSNR}}_p &:= \frac{\|\mathbf{\Psi}_{\overline{\Lambda}}^\star E\|_{p,\infty} + \|\mathbf{\Psi}_\Lambda^\star E\|_{p,\infty}}{\max_{i\in\Lambda}|\alpha_i|}, \\ \overline{R} &:= \frac{\min_{i\in\Lambda}|\alpha_i|}{\max_{i\in\Lambda}|\alpha_i|}. \end{aligned}$$

We also need the following generalization of the cumulative coherence of a dictionary :

$$\mu_q(\mathbf{\Psi}, \mathbf{\Phi}, \Lambda) := \sup_{l\notin\Lambda} \|\mathbf{\Phi}_\Lambda^\star \psi_l\|_q = \sup_{l\notin\Lambda} \left(\sum_{i\in\Lambda} |\langle \psi_l, \varphi_i\rangle|^q \right)^{1/q}. \quad (5)$$

Theorem 3.1. *Assume that the noise level and the dynamic range are sufficiently small (respectively large), that is to say*

$$\mu_2(\mathbf{\Phi}, \mathbf{\Psi}, \Lambda) < \min_{i\in\Lambda} \|\mathbf{\Phi}_\Lambda^\star \psi_i\|_2 \cdot \overline{R} - \overline{\mathrm{PSNR}}_p / C_p(N). \quad (6)$$

where $C_p(N)$ is a constant depending only on p and the number of channels N. Then, under signal model (1), the probability that p-thresholding fails to recover the indices of the atoms in Λ does not exceed

$$\mathbb{P}(p - \textit{thresholding fails}) \le K \cdot \exp\left(-AN\gamma^2\right)$$

with

$$\gamma = \frac{\overline{R} \cdot \min_{i\in\Lambda} \|\mathbf{\Phi}_\Lambda^\star \psi_i\|_2 - \overline{\mathrm{PSNR}}_p/C_p(N) - \mu_2(\mathbf{\Phi}, \mathbf{\Psi}, \Lambda)}{\overline{R} \cdot \min_{i\in\Lambda} \|\mathbf{\Phi}_\Lambda^\star \psi_i\|_2 + \mu_2(\mathbf{\Phi}, \mathbf{\Psi}, \Lambda)}.$$

The proof of this result is somewhat lengthy and relies heavily on measure concentration inequalities. The interested reader will find all details in.[6] This result has unique features compared to more classical worst case analysis. First, the condition on $\mathbf{\Phi}$ is expressed in terms of the cumulative coherence of order 2 which is much smaller than that of order one. For example assuming that there is no noise and that the variances α_i are constant the r.h.s in (6) is larger than one. If additionally $\mathbf{\Psi} = \mathbf{\Phi}$, an upper estimate of $\mu_2(\mathbf{\Phi}, \mathbf{\Psi}, \Lambda)$ is $\mu\sqrt{S}$ and we see that typically thresholding can be successful even when $S \approx \mu^{-2} \gg \mu^{-1}$. Second, due to typicality, we see that the probability of failure quickly diminishes as the number of channel grows, suggesting that we should use $N \sim \log K$ channels in practice. These findings are confirmed by simulation results as we show in the next section.

4. EXPERIMENTAL RESULTS

In this section we compare our theoretical findings with simulations of the performance of 2-thresholding with $\mathbf{\Psi} = \mathbf{\Phi}$. As dictionary we chose a combination of the Dirac and Fourier basis, $\mathbf{\Phi} = (\mathbf{I}_d, \mathcal{F}_d)$, in dimension $d = 1024$, which has coherence $\mu = 1/\sqrt{d}$. For each number of channels N, varying from 1 to 128, and support size, varying from 1 to 1024 in steps of 16, we created 180 signals by choosing a support Λ uniformly at random

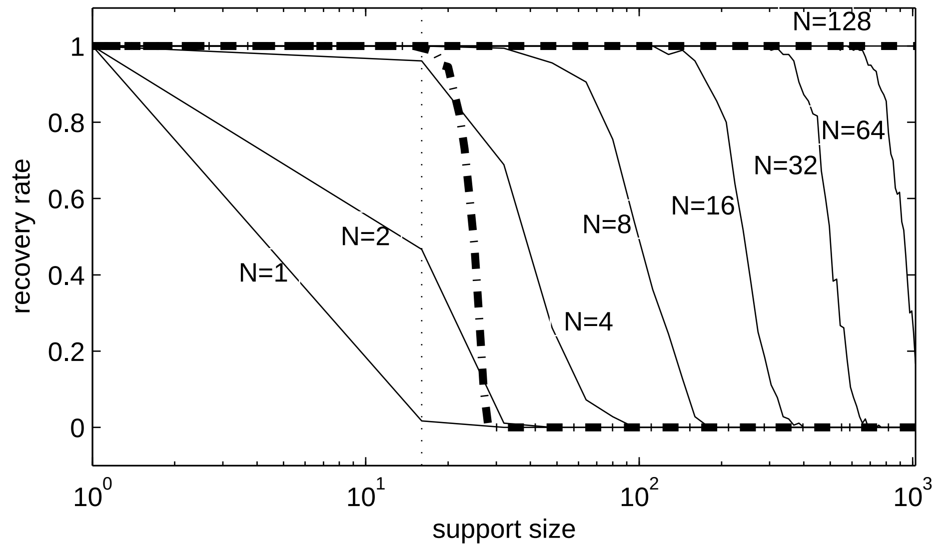

Figure 1. Comparison of Recovery Rates for Different Support Sizes and Number of Channels.

and independent Gaussian coefficients with variances $\alpha_i = 1$ and calculated the percentage of thresholding being able to recover the full support. The results can be seen in Figure 1.

As reference we also calculated how many out of 200 randomly chosen supports of a given size satisfy the worst case recovery condition $\mu_1(\lambda) + \sup_{i \in \Lambda} \mu_1(\Lambda/\{i\}) < 1$. This is indicated by the dash dotted line and can be seen to drop rapidly once the theoretical limit $|\Lambda| = 16$ is reached. Since $\mu = 1/\sqrt{d}$ the average recovery condition $\mu_2(\Lambda) < 1$, indicated by the dashed line, is always satisfied. We can see that as predicted by Theorem 3.1 with an increasing number of channels we get closer to the average case bound, which is actually attained once $N = 128$.

5. CONCLUSIONS

Thresholding is a computationally inexpensive algorithm for simultaneous sparse signal approximation. We have shown that, in a probabilistic multichannel setting, it shares good recovery properties with much more complex alternatives such as greedy algorithms and convex relaxation algorithms. The worst case recovery condition is reminiscent of Tropp's recovery condition, see,[4] but the typical behaviour is instead driven by a much less restrictive condition and improves with the numbers of channels. This is clearly confirmed by our simulation results.

It has to be noted that the results obtained in this paper do not *scale down* to a single channel. Indeed, our average case results rely heavily on typicality across channels. On the other hand, single channel average case results have been obtained for the simple thresholding algorithm in[7] and confirm the the 2-coherence is a characteristic performance measure.

One of the main drawbacks of thresholding is that its performance relies heavily on the assumption that the signal coefficients are well balanced, in addition to the Gaussian model. Orthogonal Matching Pursuit is a natural candidate for dealing with signals that do not have balanced coefficients. Preliminary results[8] indicate that its typical performance in a multi-channel probabilistic setup is also driven by much less restrictive conditions on the dictionary than the worst case ones. Last but not least, since the characterization of what drives the average performance of thresholding involves the mutual coherence of order 2 between a sensing dictionary and a synthesis dictionary, an interesting new perspective is the design of a sensing dictionary to optimize the recovery performance for a given signal model. Another interesting question would be to study how practical thresholding can be in the framework of Compressed Sensing.[9] It has been proved in[10] that thresholding can be used a recovery algorithm in this setting and its lower computational complexity (as compared with OMP) might be useful in particular applications.

REFERENCES

1. D. Baron, M. Duarte, S. Sarvotham, M. Wakin, and R. Baraniuk, "An information-theoretic approach to distributed compressed sensing," in *Proc. 45rd Conference on Communication, Control, and Computing*, 2005.
2. J. Tropp, A. Gilbert, and M. Strauss, "Algorithms for simultaneous sparse approximations. Part I: Greedy pursuit," *Signal Processing, special issue "Sparse approximations in signal and image processing"* **86**, pp. 572–588, 2006.
3. K. Schnass and P. Vandergheynst, "Dictionary preconditioning for greedy algorithms," *IEEE Transactions on Signal Processing, accepted* , May 2007.
4. J. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. on Information Theory* **50**, pp. 2231–2242, October 2004.
5. R. Gribonval, M. Nielsen, and P. Vandergheynst, "Towards an adaptive computational strategy for sparse signal approximation." IRISA preprint, Rennes, France, 2006.
6. R. Gribonval, H. Rauhut, K. Schnass, and P. Vandergheynst, "Atoms of all channels, unite! average case analysis of multi-channel sparse recovery using greedy algorithms," *Journal of Fourier Analysis and Applications* , 2007. Submitted to the special issue on Sparsity.
7. K. Schnass and P. Vandergheynst, "Average Performance Analysis for Thresholding," *IEEE Signal Processing Letters* .
8. R. Gribonval, H. Rauhut, K. Schnass, and P. Vandergheynst, "Average case analysis of multichannel thresholding and greedy algorithms." in preparation.
9. E. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Comm. Pure Appl. Math.* **59**, pp. 1207–1223, 2005.
10. H. Rauhut, K. Schnass, and P. Vandergheynst, "Compressed sensing and redundant dictionaries," *IEEE Transactions on Information Theory* , 2007.