Average Case Analysis of Multichannel Basis Pursuit

Holger Rauhut [(1)], Yonina C. Eldar [(2)]

(1) Hausdorff Center for Mathematics, and Institute for Numerical Simulation, University of Bonn
Endenicher Allee 62, 53115 Bonn, Germany.
(2) Department of Electrical Engineering, Technion, Israel Institute of Technology, Haifa, Israel 32000.
rauhut@hcm.uni-bonn.de, yonina@ee.technion.ac.il

Abstract:

We consider the recovery of jointly sparse multichannel signals from incomplete measurements using convex relaxation methods. Worst case analysis is not able to provide insights into why joint sparse recovery is superior to applying standard sparse reconstruction methods to each channel individually. Therefore, we analyze an average case by imposing a probability model on the measured signals. We show that under a very mild condition on the sparsity and on the dictionary characteristics, measured for example by the coherence, the probability of recovery failure decays exponentially in the number of channels. This demonstrates that most of the time, multichannel sparse recovery is indeed superior to single channel methods.

1. Introduction

Recovery of sparse signals from a small number of measurements is a fundamental problem in many different signal processing tasks such as image denoising [3], analog-to-digital conversion [21, 11], radar, compression, inpainting, and many more. The recent framework of compressed sensing (CS), founded in the works of Donoho [8] and Candes [3], studies acquisition methods as well as efficient computational algorithms that allow reconstruction of a sparse vector x from linear measurements $y = Ax$, where A is referred to as the measurement matrix. The key observation is that y can be relatively short, and still contain enough information to recover x.

Determining the sparsest vector x consistent with the data $y = Ax$ is generally an NP-hard problem [7]. To determine x in practice, a multitude of efficient algorithms have been proposed. The most extensively studied recovery method by now is the ℓ_1-minimization approach (Basis Pursuit). Greedy methods, such as simple thresholding [23] or orthogonal matching pursuit (OMP) [26], are faster in practice, but BP provides significantly better recovery guarantees [10, 22].

The BP principle as well as greedy approaches have been extended to the multichannel setup where the signal consists of several channels [29, 30, 15, 6, 5, 20, 12, 13, 18]. Here one assumes that each channel is sparse and in addition that the channels have a small common support set. In this situation the signals are called jointly sparse. A variety of theoretical recovery results have been established already in this setting. In [5] a recovery result was derived for a mixed ℓ_p/ℓ_1 program (multichannel BP) in which the objective is to minimize the sum of the ℓ_p-norms of the rows of the estimated matrix whose columns are the unknown vectors.

Recovery results for the more general problem of block-sparsity were developed in [13] based on the restricted isometry property (RIP), and in [12] based on mutual coherence. In practice, multichannel reconstruction techniques perform much better than recovering each channel individually. However, the theoretical equivalence results predict no performance gain. The reason is that these recovery results apply to all possible input signals, and are therefore worst-case measurements. Clearly, if we input the same signal to each channel, then no additional information on the joint support is provided from multiple measurements. Therefore, in this worst-case scenario there is no advantage for multiple channels.

In order to capture more closely the true underlying behavior of existing algorithms and observe a performance gain when using several channels, we consider an average analysis. In this setting, the inputs are considered to be random variables so that the case of identical inputs in all channels has zero probability. The idea is to develop conditions on the measurement matrix A such that the inputs can be recovered with high probability given a certain input distribution. Most existing recovery results focus on worst-case analysis. Recently, there have been several papers that consider random ensembles. In [25] random sub-dictionaries of A are considered and analyzed. This allows to obtain results for BP with a single input channel. In [23], average-case performance of single channel thresholding was studied. These ideas were then extended to two multichannel recovery algorithms: thresholding and simultaneous OMP (SOMP) [18, 17]. Under a mild condition on the sparsity and on the matrix A, it was shown that the probability of reconstruction failure decays exponentially with the number of channels. In the present paper we contribute to this line of research by adding an average-case analysis of multichannel BP, that is mixed ℓ_2/ℓ_1-minimization [30, 15, 13, 12].

We denote by A_S the submatrix of A consisting of the columns indexed by $S \subset 1, \dots, N$, while X^S is the submatrix of X consisting of the rows of X indexed by S. The ℓth column of A is denoted by a_ℓ or A_ℓ. The ℓ_p-norm is denoted by $\|\cdot\|_p$ while $\|\cdot\|_F$ is the Frobenius norm.

2. Multichannel ℓ_1-minimization

We consider multichannel signal recovery where our goal is to recover a jointly-sparse matrix $X \in \mathbb{C}^{N\times L}$ from n linear measurements per channel. Here N denotes the signal length and L the number of channels, i.e., the number of signals. We assume that X is jointly k-sparse, meaning that there are at most k rows in the matrix X that are not identically zero. More formally, we define the support of the matrix X as $\operatorname{supp} X = \bigcup_{\ell=1}^{L} \operatorname{supp} X_\ell$, where the support of the ℓth column is $\operatorname{supp} X_\ell = \{j, X_{j\ell} \neq 0\}$. Our assumption is that $\|X\|_0 = k$ where $\|X\|_0$ is equal to the size of the support. The measurements are given by

$$Y = AX, \quad Y \in \mathbb{C}^{n\times L}, \tag{1}$$

where $A \in \mathbb{C}^{n\times N}$ is a given measurement matrix. Each measurement vector Y_ℓ corresponds to a measurement of the corresponding signal X_ℓ.
The natural approach to determine X given Y is to solve the problem

$$\min_X \|X\|_0 \quad \text{such that} \quad AX = Y. \tag{2}$$

However, (2) is NP hard in general [7]. Therefore, we consider instead the convex relaxation [30, 15, 13] defined by

$$\min \|X\|_{2,1} = \sum_{j=1}^{N} \|X^j\|_2, \qquad \text{subject to } AX = Y, \tag{3}$$

which promotes joint sparsity, as argued for instance in [15]. In the single channel case $L = 1$ this is the usual BP principle.

3. Worst Case Recovery Results

Recovery results for the program (3) were considered in [5, 13, 12]. In particular, the lemma below is derived in [5] and follows also from [12].

Proposition 3..1 *Let $S \subset 1,\ldots,N$ and suppose that*

$$\|A_S^\dagger a_\ell\|_1 < 1 \quad \textit{for all } \ell \notin S, \tag{4}$$

with $A_S^\dagger = (A_S^* A_S)^{-1} A_S^*$ denoting the pseudo-inverse of A_S. Then (3) *recovers all $X \in \mathbb{C}^{N\times L}$ with $\operatorname{supp} X = S$ from $Y = AX$.*

Assuming the columns of A are normalized, $\|a_\ell\|_2 = 1$, we can guarantee that (4) holds as long as the coherence μ of A is small enough, where [9]

$$\mu = \max_{j\neq \ell} |\langle a_j, a_\ell\rangle|. \tag{5}$$

The following result follows from Proposition 3..1 or from [12] by noting that the block coherence in this setting is equal to μ/d.

Proposition 3..2 *Assume that*

$$(2k-1)\mu < 1. \tag{6}$$

Then (3) *recovers all X with $\|X\|_0 \le k$ from $Y = AX$.*

Note that in both of the cited results the conditions do not depend on the number of channels. Indeed, under the same conditions as in Propositions 3..1 and 3..2, it is shown in [26] that BP will recover a single k-sparse vector. Therefore, if (4) holds, then instead of solving (3) we may as well use BP on each of the columns of Y.
The coherence is lower bounded by $\mu \geq \sqrt{\frac{N-n}{n(N-1)}}$ [24]. The lower bound behaves like $1/\sqrt{n}$ for large N, which limits the Proposition 3..2 to maximal sparsities $k = \mathcal{O}(\sqrt{n})$. To improve on this we can generalize existing recovery results [3, 2] based on RIP to the multichannel setup. The next proposition follows from [13]:

Proposition 3..3 *Assume $X \in \mathbb{C}^{n\times N}$ with $\delta_{2k} < \sqrt{2}-1$, where δ_{2k} is the smallest constant δ such that*

$$(1-\delta)\|x\|_2^2 \le \|Ax\|_2^2 \le (1+\delta)\|x\|_2^2,$$

for all vectors x that are $2k$-sparse. Let $X \in \mathbb{C}^{N\times L}$, $Y = AX$, and let $\overline{X}$ be the minimizer of (3). *Then*

$$\|X - \overline{X}\|_F \le Ck^{-1/2}\|X - \hat{X}^{(k)}\|_{1,2}$$

where C is a constant, $\|X\|_F = \sqrt{\operatorname{Tr}(X^*X)}$ is the Frobenius norm of X, $\|X\|_{1,2} = \sum_{j=1}^{N} \|X^j\|_2$, and $\hat{X}^{(k)}$ denotes the best k-term approximation of X, i.e., $\operatorname{supp} \hat{X}^{(k)}$ consists of the indices corresponding to the k largest row norms $\|X^\ell\|_2$. In particular, recovery is exact if $|\operatorname{supp} X| \le k$.

It is well known that Gaussian and Bernoulli random matrices $A \in \mathbb{R}^{n\times N}$ satisfy $\delta_{2k} \le \sqrt{2}-1$ with high probability as long as [1, 4]

$$n \ge Ck\log(N/k). \tag{7}$$

Therefore, Proposition 3..3 allows for a smaller number of measurements. However, there is still no dependency on the number of channels. Indeed, under the same RIP condition BP will recover a single k-sparse vector and therefore, as before, BP may as well be applied to each of the columns of Y individually.

4. Average Case Analysis

Intuitively, we would expect multichannel sparse recovery to perform better than single channel recovery. However, in the worst case setting this is not true as already suggested by the results cited above. The reason is very simple. If each channel carries the same signal, $X_\ell = x$ for $\ell = 1,\ldots,L$, then also the components of $Y = AX$ are all the same and we do not have more information on the support of X than provided by a single component Y_ℓ. This can indeed be proven rigorously.

Proposition 4..1 *Suppose there exists a k-sparse vector $x \in \mathbb{R}^N$ that ℓ_1-minimization is not able to recover from $y = Ax$. Then there exists a k-sparse multichannel signal $X \in \mathbb{R}^{N\times L}$ for which mixed ℓ_2/ℓ_1-minimization fails on $Y = AX$.*

For the simple proof we refer to the journal version [14]. Realizing that (3) is not more powerful than usual BP in the worst case, we seek an average-case analysis. This means that we impose a probability model on the k-sparse X. In particular, as in [18], we will assume that on the k-sparse support set S the coefficients of X are independent and follow a normal distribution,

$$X^S = \Sigma\Phi \tag{8}$$

where $\Sigma = \operatorname{diag}(\sigma_j, j \in S) \in \mathbb{R}^{k\times k}$ is an arbitrary diagonal matrix with non-zero diagonal elements σ_j, while $\Phi \in \mathbb{R}^{k\times L}$ is a Gaussian random matrix, i.e., all entries are independent standard normal random variables. Note that taking Σ to be the identity matrix results in a standard Gaussian random matrix, while taking arbitrary non-zero σ_j's on the diagonal of Σ allows for different variances.

The following recovery condition is instrumental in proving average case recovery results for multichannel BP. It generalizes results of [27, 16] for the monochannel case. In order to introduce we need to introduce the sign $\operatorname{sgn}(X)$ of a signal matrix,

$$\operatorname{sgn}(X)_{\ell j} = \begin{cases} \frac{X_{\ell j}}{\|X^\ell\|_2}, & \|X^\ell\|_2 \neq 0; \\ 0, & \|X^\ell\|_2 = 0. \end{cases}$$

Proposition 4..2 *Let $X \in \mathbb{C}^{N\times L}$ with $\operatorname{supp} X = S$ and assume A_S to be non-singular. If*

$$\|\operatorname{sgn}(X^S)^* A_S^\dagger a_\ell\|_2 < 1 \quad \textit{for all } \ell \notin S \tag{9}$$

then X is the unique minimizer of (3).

Combining the above proposition with a concentration inequality for sums of independent random variables that are uniformly distributed on the sphere [19], we arrive at the following average case recovery result for multichannel BP.

Theorem 4..3 *Let $S \subset \{1,\ldots,N\}$ be a set of cardinality k and let $X \in \mathbb{R}^{N\times L}$ with $\operatorname{supp} X \subset \{1,\ldots,N\}$ such that the coefficients on S are given by* (8) *with some diagonal matrix $\Sigma \in \mathbb{R}^{k\times k}$. If*

$$\|A_S^\dagger a_\ell\|_2 \leq \alpha < 1 \quad \textit{for all } \ell \notin S, \tag{10}$$

then with probability at least

$$1 - N\exp\left(-\frac{L}{2}(\alpha^{-2} - \log(\alpha^{-2}) - 1)\right) \tag{11}$$

(3) *recovers X from $Y = AX$.*

The proof of the theorem will appear in the journal version [14]. For $\alpha < 1$ we are guaranteed that the exponent has a negative argument, and therefore the error decays exponentially in L. We note that for the monochannel case $L = 1$, Theorem 4..3 is contained implicitly in [28, Theorem 13]. The appearance of the 2-norm in (10) instead of the 1-norm as in (4) makes the condition of the theorem weaker than worst-case estimates.

Let us finally state conditions on the matrix A and the sparsity level k ensuring that $\|A_S^\dagger a_\ell\|_2$ is small, which is needed in order to apply Theorem 4..3.

Proposition 4..4 *Suppose A has restricted isometry constant $\delta_{k+1} \leq \delta < 1/2$. If $S \subset \{1,\ldots,N\}$ has cardinality k then*

$$\|A_S^\dagger a_\ell\|_2 \leq \frac{\delta}{1-\delta} < 1 \quad \textit{for all } \ell \notin S.$$

Note that in contrast to the worst case result in Proposition 3..3 where a condition on δ_{2k} is needed, we only require that δ_{k+1} is small, which is clearly weaker. For random matrices A we have the following bound on $\|A_S^\dagger a_\ell\|_2$.

Proposition 4..5 *Let $S \subset \{1,\ldots,N\}$ be a set of cardinality k and suppose that $A \in \mathbb{R}^{n\times N}$ is drawn at random according to a Gaussian or Bernoulli distribution. Then*

$$\|A_S^\dagger a_\ell\|_2 \leq \delta \quad \textit{for all } \ell \notin S$$

with probability at least $1-\epsilon$ provided that

$$n \geq C\delta^{-2}[(k+1)\ln(1+12/\delta) + \ln(2N/\epsilon)]. \tag{12}$$

The constant C is no larger than $162/7 \approx 23.1$.

Note that the log-factor in (12) enters only as an additive term, while in (7) it appears as multiplicative factor.

5. Conclusion

Our main result is that under mild conditions on the sparsity and measurement matrix, the probability of failure of multichannel BP (3) decays exponentially with the number of channels. To develop this result we assumed a probability model on the non-zero coefficients of a jointly sparse signal. This shows that multichannel BP outperforms single channel BP applied to each channel individually, on average. Proofs of our theorems, together with improved results for simple thresholding and numerical experiments will appear in [14].

6. Acknowledgements

The work of YE was supported in part by the Israel Science Foundation under Grant no. 1081/07 and by the European Commission in the framework of the FP7 Network of Excellence in Wireless COMmunications NEWCOM++ (contract no. 216715). HR acknowledges funding by the Hausdorff Center for Mathematics, University of Bonn and the WWTF project SPORTS (MA 07-004).

References:

[1] R. G. Baraniuk, M. Davenport, R. A. DeVore, and M. Wakin. A simple proof of the restricted isometry property for random matrices. *Constr. Approx.*, 28(3):253–263, 2008.

[2] E. J. Candès. The restricted isometry property and its implications for compressed sensing. *Compte Rendus de l'Academie des Sciences, Paris, Serie I*, 346:589–592, 2008.

[3] E. J. Candès, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.*, 59(8):1207–1223, 2006.

[4] E. J. Candès and T. Tao. Near optimal signal recovery from random projections: universal encoding strategies? *IEEE Trans. Inform. Theory*, 52(12):5406–5425, 2006.

[5] J. Chen and X. Huo. Theoretical results on sparse representations of multiple-measurement vectors. *IEEE Trans. Signal Processing*, 54(12):4634–4643, Dec. 2006.

[6] S. F. Cotter, B. D. Rao, K. Engan, and K. Kreutz-Delgado. Sparse solutions to linear inverse problems with multiple measurement vectors. *IEEE Trans. Signal Processing*, 53(7):2477–2488, July 2005.

[7] G. Davis, S. Mallat, and M. Avellaneda. Adaptive greedy approximations. *Constr. Approx.*, 13(1):57–98, 1997.

[8] D. L. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4):1289–1306, 2006.

[9] D. L. Donoho and X. Huo. Uncertainty principles and ideal atomic decomposition. *IEEE Transactions Info. Theory*, 47(7):2845–2862, 2001.

[10] David L. Donoho. For most large underdetermined systems of linear equations the minimal l^1 solution is also the sparsest solution. *Commun. Pure Appl. Anal.*, 59(6):797–829, 2006.

[11] Y. C. Eldar. Compressed sensing of analog signals. submitted to *IEEE Trans. Signal Processing*.

[12] Y. C. Eldar and H. Bölcskei. Block-sparsity: Coherence and efficient recovery. to appear in *ICASSP09*.

[13] Y. C. Eldar and M. Mishali. Robust recovery of signals from a union of subspaces. submitted to *IEEE Trans. Inf. Theory*.

[14] Y. C. Eldar and H. Rauhut. Average case analysis for multichannel sparse recovery using convex relaxation. *preprint*, 2009.

[15] M. Fornasier and H. Rauhut. Recovery algorithms for vector valued data with joint sparsity constraints. *SIAM J. Numer. Anal.*, 46(2):577–613, 2008.

[16] J. J. Fuchs. On sparse representations in arbitrary redundant bases. *IEEE Trans. Inform. Theory*, 50(6):1341–1344, 2004.

[17] R. Gribonval, B. Mailhe, H. Rauhut, K. Schnass, and P. Vandergheynst. Average case analysis of multichannel thresholding. In *Proc. IEEE Intl. Conf. Acoust. Speech Signal Process.*, 2007.

[18] 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. *J. Fourier Anal. Appl.*, 14(5):655–687, 2008.

[19] H. König and S. Kwapień. Best Khintchine type inequalities for sums of independent, rotationally invariant random vectors. *Positivity*, 5(2):115–152, 2001.

[20] M. Mishali and Y. C. Eldar. Reduce and boost: Recovering arbitrary sets of jointly sparse vectors. *IEEE Trans. Signal Process.*, 56(10):4692–4702, Oct. 2008.

[21] M. Mishali and Y. C. Eldar. Blind multi-band signal reconstruction: Compressed sensing for analog signals. *IEEE Trans. Signal Process.*, 57(3):993–1009, Mar. 2009.

[22] H. Rauhut. On the impossibility of uniform sparse reconstruction using greedy methods. *Sampl. Theory Signal Image Process.*, 7(2):197–215, 2008.

[23] K. Schnass and P. Vandergheynst. Average performance analysis for thresholding. *IEEE Signal Processing Letters*, 14(11):828–831, Nov. 2007.

[24] T. Strohmer and R. W. Heath. Grassmannian frames with applications to coding and communication. *Appl. Comput. Harmon. Anal.*, 14(3):257–275, 2003.

[25] G. Teschke. Multi-frame representations in linear inverse problems with mixed multi-constraints. *Appl. Comput. Harmon. Anal.*, 22(1):43–60, 2007.

[26] J. A. Tropp. Greed is good: Algorithmic results for sparse approximation. *IEEE Trans. Inform. Theory*, 50(10):2231–2242, 2004.

[27] J. A. Tropp. Recovery of short, complex linear combinations via l_1 minimization. *IEEE Trans. Inform. Theory*, 51(4):1568–1570, 2005.

[28] J. A. Tropp. On the conditioning of random subdictionaries. *Appl. Comput. Harmon. Anal.*, to appear.

[29] J. A. Tropp, A. C. Gilbert, and M. J. Strauss. Algorithms for simultaneous sparse approximation: part I: Greedy pursuit. *Signal Processing*, 86(3):572 – 588, 2006.

[30] J. A. Tropp, A. C. Gilbert, and M. J. Strauss. Algorithms for simultaneous sparse approximation: part II: Convex relaxation. *Signal Processing*, 86(3):589 – 602, 2006.