4-1-2010

Mixed Operators in Compressed Sensing

Matthew A. Herman
University of California - Los Angeles

Deanna Needell
Claremont McKenna College

Recommended Citation

Herman, M., Needell, D., "Mixed operators in compressed sensing", CISS 2010 (44th Annual Conference on Information Sciences and Systems), doi: 10.1109/CISS.2010.5464909

MIXED OPERATORS IN COMPRESSED SENSING

MATTHEW A. HERMAN AND DEANNA NEEDELL

Abstract. Applications of compressed sensing motivate the possibility of using different operators to encode and decode a signal of interest. Since it is clear that the operators cannot be too different, we can view the discrepancy between the two matrices as a perturbation. The stability of ℓ_1-minimization and greedy algorithms to recover the signal in the presence of *additive noise* is by now well-known. Recently however, work has been done to analyze these methods with noise in the measurement matrix, which generates a *multiplicative noise* term. This new framework of generalized perturbations (i.e., both additive and multiplicative noise) extends the prior work on stable signal recovery from incomplete and inaccurate measurements of Candès, Romberg and Tao using Basis Pursuit (BP), and of Needell and Tropp using Compressive Sampling Matching Pursuit (CoSaMP). We show, under reasonable assumptions, that the stability of the reconstructed signal by both BP and CoSaMP is limited by the noise level in the observation. Our analysis extends easily to arbitrary greedy methods.

1. Introduction

Compressed sensing refers to the problem of realizing a sparse, or nearly sparse, signal from a small set of linear measurements. There are many applications of compressed sensing in engineering and the the sciences. Examples include biomedical imaging, x-ray crystallography, audio source separation, seismic exploration, radar and remote sensing, telecommunications, distributed and multi-sensor networks, machine learning, robotics and control, astronomy, surface metrology, coded aperture imaging, biosensing of DNA, and many more. See [5] for an extensive list of the latest literature.

To precisely formulate the problem, we define an s-sparse signal $x \in \mathbb{C}^d$ to be one with s or fewer non-zero components,

$$\|x\|_0 \stackrel{\text{def}}{=} |\operatorname{supp}(x)| \leq s \ll d.$$

We apply a matrix $A \in \mathbb{C}^{m \times d}$ to the signal and acquire measurements $b = Ax$. Often, we encounter additive noise so that the measurements become $y = b + e = Ax + e$, where e is an error or noise term usually assumed to have bounded energy $\|e\|_2 \leq \epsilon$. The field of compressed sensing has provided many recovery algorithms for sparse and nearly sparse signals, most with strong theoretical and numerical results.

One major approach to sparse recovery is ℓ_1-minimization or Basis Pursuit [7, 4]. This method simply solves an optimization problem to recover the signal x,

$$\min_z \|z\|_1 \quad \text{such that} \quad \|Az - y\|_2 \leq \epsilon. \tag{1.1}$$

1991 *Mathematics Subject Classification.* 68W20, 65T50, 41A46.

D.N. is with the Dept. of Statistics, Stanford University, 390 Serra Mall, Stanford CA 94305, USA. e-mail: dneedell@stanford.edu.

M.H. is with the Dept. of Mathematics, University of California, Los Angeles, 520 Portola Plaza, Los Angeles, CA 90095, USA. e-mail: mattyh@math.ucla.edu.

M.H. is partially supported by NSF Grant No. DMS-0811169, NSF VIGRE Grant No. DMS-0636297, and a grant from the DoD at UCLA. D.N. is partially supported by the NSF DMS EMSW21-VIGRE grant.

This problem can be solved using convex optimization techniques and is thus computationally feasible. Candès and Tao show in [4] that if the signal x is sparse and the measurement matrix A satisfies a certain quantitative property, then (1.1) recovers the signal x exactly.

Definition 1.1. A measurement matrix A satisfies the *restricted isometry property* (RIP) with parameters (s, δ) if for every s-sparse vector x, we have

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

The parameter δ is also referred to as the *restricted isometry constant* (RIC) of matrix A.

It is now well known that many $m \times d$ matrices (e.g., random Gaussian, Bernoulli, and partial Fourier) satisfy the RIP with parameters (s, δ) when $m = \mathrm{O}(s \log d)$, see [14, 17] for details. It has been shown in [4, 3] that if A satisfies the RIP with parameters $(3s, 0.2)$, then (1.1) recovers a signal $x^\star$ that satisfies

$$\|x^\star - x\|_2 \le C_0 \frac{\|x - x_s\|_1}{\sqrt{s}} + C_1\epsilon \tag{1.2}$$

where x_s denotes the vector consisting of the s largest components of x in magnitude.

In [2] Candès sharpened this bound to work for matrices satisfying the RIP with parameters $(2s, \sqrt{2}-1)$, and later Foucart and Lai sharpened it to work for $(2s, 0.4531)$ in [12].

Although the recovery guarantees provided by ℓ_1-minimization are strong, it requires methods of convex optimization which, although often quite efficient in practice, have a polynomial runtime. For this reason, much work in compressed sensing has been done to find faster methods. Many of these algorithms are greedy, and compute the (support of the) signal iteratively (see e.g., [18, 1, 8, 11, 16, 6]). Our analysis in this work focuses on Needell and Tropp's Compressive Sampling Matching Pursuit (CoSaMP) [15]. CoSaMP provides a fast runtime while also providing strong guarantees analogous to those of ℓ_1-minimization.

The CoSaMP algorithm can be described as follows. We use the notation $w|_T$ and A_T to denote the vector w restricted to indices given by a set T, and the matrix A restricted to the the columns indexed by T, respectively.

Compressive Sampling Matching Pursuit (CoSaMP)

Input: Measurement matrix A, measurement vector y, sparsity level s
Output: s-sparse reconstructed vector $\hat{x} = a$
Procedure:

- **Initialize:** Set $a^0 = 0$, $v = y$, $k = 0$. Repeat the following steps and increment k until the halting criterion is true.
- **Signal Proxy:** Set $u = A^*v$, $\Omega = \mathrm{supp}(u_{2s})$ and merge the supports: $T = \Omega \cup \mathrm{supp}(a^{k-1})$.
- **Signal Estimation:** Using least-squares, set $w|_T = A_T^\dagger y$ and $w|_{T^c} = 0$.
- **Prune:** To obtain the next approximation, set $a^k = w_s$.
- **Sample Update:** Update the current samples: $v = y - Aa^k$.

In [15] it is shown that when the measurement matrix has a small RIC that CoSaMP approximately recovers arbitrary signals from noisy measurements. This is summarized by the following.

Theorem 1.2 (CoSaMP [15])**.** *Suppose that A is a measurement matrix with RIC $\delta_{4s} \le 0.1$. Let $y = Ax + e$ be a vector of samples of an arbitrary signal x, contaminated with arbitrary* *noise. Then the algorithm CoSaMP produces an s-sparse approximation $x^\sharp$ that satisfies*

$$\|x^\sharp - x\|_2 \le C \cdot \big(\|x - x_s\|_2 + \frac{1}{\sqrt{s}}\|x - x_s\|_1 + \|e\|_2\big).$$

2. Mixed Operators

Applying the theories of compressed sensing to real-world problems raises the following question: what happens when the operator used to *encode* the signal is *different* from the operator used to *decode* the measurements? In many of the applications mentioned in Section 1 the sensing, or measurement, matrix A actually represents a system which the signal passes through. In other scenarios A represents some other physical phenomenon. For example, in screening for genetic disorders [10], the standard deviation of error in the sensing matrix is about 3%. This error is due to human handling when pipetting the DNA samples [9]. Whatever the setting may be, it is often the case that the true nature of this system is not known exactly. When this happens the system behavior is (perhaps unknowingly) approximated, or assumed to be represented, by a different matrix Φ.

It is clear that the encoding and decoding operators, A and Φ, cannot be too different, but until recently there has been no analysis of the effect this difference has on reconstruction error. In particular, the perturbations in the sensing matrices create a multiplicative noise term of the form $(A - \Phi)x$. Herman and Strohmer first showed in [13] that a noisy measurement matrix can be successfully used to recover a signal using ℓ_1-minimization. The natural question is whether this extends to the case of greedy algorithms. In this work, we consider the case of mixed operators in CoSaMP, and our results naturally apply to other greedy algorithms.

In this analysis we will require examination of submatrices of certain matrices. To that end, we define $\|A\|_2^{(s)}$ to be the largest spectral norm over all s-column submatrices of A. Let

$$\varepsilon_A^{(s)} \stackrel{\text{def}}{=} \frac{\|A - \Phi\|_2^{(s)}}{\|A\|_2^{(s)}} \quad \text{and} \quad \kappa_A^{(s)} \stackrel{\text{def}}{=} \frac{\sqrt{1+\delta_s}}{\sqrt{1-\delta_s}}. \tag{2.1}$$

The first quantity is the relative perturbation of s-column submatrices of A with respect to to the spectral norm, and the second one bounds ratio of the extremal singular values of all s-column submatrices of A (see [13] for more details). We also need a measure of how "close" a signal x is to a sparse signal, and therefore define

$$\alpha_s \stackrel{\text{def}}{=} \frac{\|x - x_s\|_2}{\|x_s\|_2} \quad \text{and} \quad \beta_s \stackrel{\text{def}}{=} \frac{\|x - x_s\|_1}{\sqrt{s}\|x_s\|_2}. \tag{2.2}$$

2.1. **Mixed Operators in ℓ_1-minimization.** The work in [13] extended the previous results in ℓ_1-minimization by generalizing the error term ϵ which only accounted for additive noise. The new framework considers a *total noise term* $\varepsilon_{A,s,b}$ which allows for both multiplicative and additive noise. Theorem 2.1 below shows that the reconstruction error using ℓ_1-minimization is limited by this noise level. With regard to noise in operator A we see that the stability of the solution is a linear function of relative perturbations $\varepsilon_A, \varepsilon_A^{(s)}$.

Theorem 2.1 (Adapted from [13], Thm. 2)**.** *Let x be an arbitrary signal with measurements $b = Ax$, corrupted with noise to form $y = Ax + e$. Assume the RIC for matrix A satisfies*

$$\delta_{2s} < \frac{\sqrt{2}}{\left(1 + \varepsilon_A^{(2s)}\right)^2} - 1 \tag{2.3}$$

and that general signal x satisfies

$$\alpha_s + \beta_s \; < \; \frac{1}{\kappa_A^{(s)}}. \tag{2.4}$$

Set the total noise parameter

$$\varepsilon_{A,s,b} := \left(\frac{\varepsilon_A^{(s)} \kappa_A^{(s)} \, + \, \varepsilon_A \gamma_A \alpha_s}{1 - \kappa_A^{(s)} \big(\alpha_s + \beta_s \big)} \; + \; \varepsilon_b \right) \|b\|_2 \tag{2.5}$$

where the relative perturbations $\varepsilon_A = \frac{\|A-\Phi\|_2}{\|A\|_2}$, $\varepsilon_b = \frac{\|e\|_2}{\|b\|_2}$, *and* $\gamma_A = \frac{\|A\|_2}{\sqrt{1-\delta_s}}$. *Then the solution* $z^\star$ *to the BP problem (1.1) with* ϵ *set to* $\varepsilon_{A,s,b}$, *and using the decoding matrix* Φ *(instead of* A*) obeys*

$$\|z^\star - x\|_2 \; \le \; \frac{C_0}{\sqrt{s}} \|x - x_s\|_1 \; + \; C_1 \varepsilon_{A,s,b} \tag{2.6}$$

for some well-behaved constants C_0, C_1.

2.2. **Mixed Operators in CoSaMP.** We now turn to the case of mixed operators in greedy algorithms, and in particular CoSaMP. We will see that a result analogous to that of ℓ_1-minimization can be obtained in this case as well. Similar to condition (2.4) above, we will again need for the signal to be well approximated by a sparse signal. To that end, we require that

$$\alpha_s + \beta_s \; \le \; \frac{1}{2\,\kappa_A^{(s)}} \tag{2.7}$$

where α_s and β_s are defined in (2.2). Theorem 2.2 below shows that under this assumption, the reconstruction error in CoSaMP is again limited by the tail of the signal and the observation noise.

Theorem 2.2. *Let A be a measurement matrix with RIC*

$$\delta_{4s} \le \frac{1.1}{(1 + \varepsilon_A^{(4s)})^2} - 1. \tag{2.8}$$

Let x be an arbitrary signal with measurements $b = Ax$, corrupted with noise to form $y = Ax + e$. Let $x^\sharp$ be the reconstruction from CoSaMP using decoding matrix Φ (instead of A) on measurements y. Then if (2.7) is satisfied, the estimation satisfies

$$\|x^\sharp - x\|_2 \le C \cdot \left(\|x - x_s\|_2 + \frac{1}{\sqrt{s}} \|x - x_s\|_1 + (\varepsilon \alpha_s + \varepsilon^{(s)}) \|b\|_2 + \|e\|_2 \right)$$

where $\varepsilon = \|A - \Phi\|_2$ *and* $\varepsilon^{(s)} = \|A - \Phi\|_2^{(s)}$.

Applying Theorem 2.2 to the sparse case, we immediately have the following corollary.

Corollary 2.3. *Let A be a measurement matrix with RIC $\delta_{4s} \le \frac{1.1}{(1+\varepsilon_A^{(4s)})^2} - 1$. Let x be an s-sparse signal with noisy measurements $y = b + e = Ax + e$. Let $x^\sharp$ be the reconstruction from CoSaMP using decoding matrix Φ (instead of A). Then the estimation satisfies*

$$\|x^\sharp - x\|_2 \le C \cdot \Big(\varepsilon^{(s)} \|b\|_2 + \|e\|_2 \Big)$$

where $\varepsilon^{(s)} = \|A - \Phi\|_2^{(s)}$.

We now analyze the case of mixed operators and prove our main result, Theorem 2.2. We will first utilize a result from [13] which states that matrices which are "close" to each other also have similar RICs.

Lemma 2.4 (RIP for Φ [13])**.** *For any $s = 1, 2, \ldots$, assume and fix the RIC δ_s associated with A, and the relative perturbation $\varepsilon_A^{(s)}$ associated with $A - \Phi$ as defined in (2.1). Then the RIC constant $\hat{\delta}_s$ for matrix Φ satisfies*

$$\hat{\delta}_s \;\leq\; \big(1+\delta_s\big)\Big(1+\varepsilon_A^{(s)}\Big)^2 - 1. \tag{2.9}$$

We now prove our main result, Theorem 2.2

Proof of Theorem 2.2. Lemma 2.4 applied to the case where $\delta_{4s} \leq \frac{1.1}{(1+\varepsilon_A^{(4s)})^2} - 1$ implies that the matrix Φ has an RIC that satisfies $\hat{\delta}_{4s} \leq 0.1$. We can then apply Theorem 1.2 with measurements $y = \Phi x + (A - \Phi)x + e$. This implies that the reconstruction $x^\sharp$ satisfies

$$\|x^\sharp - x\|_2 \leq C \cdot \left(\|x - x_s\|_2 + \frac{1}{\sqrt{s}}\|x - x_s\|_1 + \|(A-\Phi)x\|_2 + \|e\|_2 \right). \tag{2.10}$$

As seen in Proposition 3.5 of [15], the RIP implies that for an arbitrary signal x,

$$\|Ax\|_2 \leq \sqrt{1+\delta_s}\left(\|x\|_2 + \frac{1}{\sqrt{s}}\|x\|_1 \right).$$

As shown in [13], this and the RIP imply that

$$\|Ax\|_2 \geq \sqrt{1-\delta_s}\|x_s\|_2 - \sqrt{1+\delta_s}\big(\|x - x_s\|_2 + \frac{1}{\sqrt{s}}\|x - x_s\|_1\big).$$

We then have that

$$\|(A-\Phi)x\|_2 \leq \frac{\|A-\Phi\|_2\|x-x_s\|_2 + \|A-\Phi\|_2^{(s)}\|x_s\|_2}{\sqrt{1-\delta_s}\|x_s\|_2 - \sqrt{1+\delta_s}(\|x-x_s\|_2 + \frac{1}{\sqrt{s}}\|x-x_s\|_1)}\|Ax\|_2.$$

Condition (2.7) then gives us

$$\begin{aligned}\|(A-\Phi)x\|_2 &\leq \frac{\|A-\Phi\|_2\|x-x_s\|_2 + \|A-\Phi\|_2^{(s)}\|x_s\|_2}{\frac{1}{2}\sqrt{1-\delta_s}\|x_s\|_2}\|Ax\|_2 \\ &= \left(\frac{2\|A-\Phi\|_2}{\sqrt{1-\delta_s}} + \frac{2\|A-\Phi\|_2^{(s)}}{\sqrt{1-\delta_s}} \right)\|Ax\|_2.\end{aligned}$$

Applying the inequality $\delta_s \leq \delta_{4s} \leq 0.1$ yields

$$\|(A-\Phi)x\|_2 \leq C'\big(\|A-\Phi\|_2\alpha_s + \|A-\Phi\|_2^{(s)}\big)\|Ax\|_2.$$

Combined with (2.10), this completes the claim. □

3. Discussion

One should of course make sure that the requirements imposed by Theorems 2.1 and 2.2 are reasonable and make sense. For instance, in Theorem 2.1, to ensure that the RIC $\delta_{2s} \geq 0$ we can set the left-hand side of condition (2.3) to zero. Rearranging, this requires that $\varepsilon_A^{(2s)} < \sqrt[4]{2} - 1$, which addresses the question "how *dissimilar* can A and Φ be?" Loosely phrased, the answer is that the spectral norm of $2s$-column submatrices of Φ cannot deviate by more than about 19% of spectral norm of $2s$-column submatrices of A. The corresponding condition (2.8) in Theorem 2.2 requires that $\varepsilon_A^{(4s)} \leq \sqrt{1.1} - 1$, which translates to an approximate 5% dissimilarity between A and Φ. The second condition, (2.4), in Theorem 2.1 is discussed in [13], and essentially requires that the signal be well approximated by a sparse signal. This is, of course, a standard assumption in compressed sensing. The same argument holds for condition (2.7) in Theorem 2.2.

In conclusion, real-world applications often utilize different operators (perhaps unknowingly) to encode and decode a signal. The perturbation of the sensing matrix creates *multiplicative* noise in the system. This type of noise is *fundamentally* different than simple additive noise. For example, to overcome a poor signal-to-noise ratio (SNR) due to additive noise, one would typically increase the strength of the signal. However, if the noise is multiplicative this will not improve the situation, and in fact will actually cause the error to grow. Thus the impact on reconstruction from the error in the sensing matrices needs to be analyzed. Our Theorems 2.1 and 2.2 do just that. They show the effect of using mixed operators to recover a signal in compressed sensing: the stability of the recovered signal is a linear function of the operator perturbations defined above. This work confirms that this is the case both for ℓ_1-minimization and CoSaMP. These results can easily be extended to other greedy algorithms as well.

Acknowledgment. We would like to thank Thomas Strohmer for many thoughtful discussions and his invaluable guidance.

References

[1] T. Blumensath and M. E. Davies. Iterative hard thresholding for compressed sensing. Preprint, 2008.
[2] E. J. Candès. The restricted isometry property and its implications for compressed sensing. *C. R. Math. Acad. Sci. Paris, Serie I*, 346:589–592, 2008.
[3] E. J. Candès, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics*, 59(8):1207–1223, 2006.
[4] E. J. Candès and T. Tao. Decoding by linear programming. *IEEE Trans. Inform. Theory*, 51:4203–4215, 2005.
[5] Compressed sensing webpage. `http://www.dsp.ece.rice.edu/cs/`.
[6] W. Dai and O. Milenkovic. Subspace pursuit for compressive sensing signal reconstruction. *IEEE Trans. Info. Theory*, 55(5), 2009.
[7] D. L. Donoho and P. B. Stark. Uncertainty principles and signal recovery. *SIAM J. Appl. Math.*, 49(3):906–931, June 1989.
[8] D. L. Donoho, Y. Tsaig, I. Drori, and J.-L. Starck. Sparse solution of underdetermined linear equations by stagewise Orthogonal Matching Pursuit (StOMP). Submitted for publication, 2007.
[9] Y. Erlich. Jan. 2010. Personal communication.
[10] Y. Erlich, A. Gordon, M. Brand, G. J. Hannon, and P. P. Mitra. Compressed genotyping. *IEEE Trans. Info. Theory: Special Issue On Molecular Biology and Neuroscience*, Feb. 2010.
[11] M. Fornasier and H. Rauhut. Iterative thresholding algorithms. Preprint, 2007.
[12] S. Foucart and M.-J. Lai. Sparsest solutions of undetermined linear systems via ℓ_q-minimization for $0 < q \leq 1$. Preprint, 2008.
[13] M. A. Herman and T. Strohmer. General Deviants: An analysis of perturbations in compressed sensing. *IEEE Journal of Selected Topics in Sig. Proc.: Special Issue on Compressive Sensing*, 4(2), Apr. 2010.
[14] S. Mendelson, A. Pajor, and N. Tomczak-Jaegermann. Uniform uncertainty principle for Bernoulli and subgaussian ensembles. *Constr. Approx.*, 28(3):277–289, 2008.
[15] D. Needell and J. A. Tropp. CoSaMP: Iterative signal recovery from noisy samples. *Appl. Comput. Harmon. Anal.*, 26(3):301–321, 2008.
[16] D. Needell and R. Vershynin. Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit. *Found. Comput. Math.*, 9(3):317–334, 2007.
[17] M. Rudelson and R. Vershynin. On sparse reconstruction from Fourier and Gaussian measurements. *Comm. Pure Appl. Math.*, 61:1025–1045, 2008.
[18] J. A. Tropp and A. C. Gilbert. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Trans. Info. Theory*, 53(12):4655–4666, 2007.