NONCONVEX COMPRESSED SENSING AND ERROR CORRECTION

Rick Chartrand

Los Alamos National Laboratory,
rickc@lanl.gov

ABSTRACT

The theory of compressed sensing has shown that sparse signals can be reconstructed exactly from remarkably few measurements. In this paper we consider a nonconvex extension, where the ℓ^1 norm of the basis pursuit algorithm is replaced with the ℓ^p norm, for $p < 1$. In the context of sparse error correction, we perform numerical experiments that show that for a fixed number of measurements, errors of larger support can be corrected in the nonconvex case. We also provide a theoretical justification for why this should be so.

Index Terms— Signal reconstruction, error correction, minimization methods, linear codes, random codes.

1. INTRODUCTION

Recent papers [1, 2] have introduced the concept of *compressed sensing*. The basic principle is that sparse or compressible signals can be reconstructed from a limited (or compressed) number of random projections. A few of the many potential applications are medical image reconstruction [3], image acquisition [4], and sensor networks [5].

The first algorithm presented in this context is known as basis pursuit [6]. Let Φ be an $M \times N$ measurement matrix, and Φf the vector of M measurements of an N-dimensional signal f. The reconstructed signal u^* is the minimizer of the ℓ^1 norm, subject to the data:

$$\min_u \|u\|_1, \quad \text{subject to } \Phi u = \Phi f. \tag{1}$$

A remarkable result from [7] is that if the rows of Φ are randomly chosen, standard-normally distributed vectors, there is a constant C such that if the support of f has size K and $M \geq CK \log(K/N)$, then the solution to (1) will be exactly $u^* = f$ with overwhelming probability. The required C depends on the desired probability of success, which in any case tends to one as $N \to \infty$.

Variants of this result include Φ being a random Fourier submatrix, or having values $\pm 1/\sqrt{N}$ with equal probability. Also, f can be sparse with respect to any basis, with u replaced with Ψu for suitable unitary Ψ.

A family of iterative greedy algorithms [8, 9, 10] have been shown to enjoy a similar exact reconstruction property, generally with less computational complexity. However, these algorithms require more measurements for exact reconstruction than the basis pursuit method.

In this paper, we take the opposite approach, and show that a nonconvex variant of basis pursuit will produce exact reconstruction with fewer measurements. Specifically, we replace the ℓ^1 norm with the ℓ^p norm, where $0 < p < 1$ (in which case $\|\cdot\|_p$ isn't actually a norm, though $d(x, y) = \|x - y\|_p^p$ is a metric):

$$\min_u \|u\|_p^p, \quad \text{subject to } \Phi u = \Phi f. \tag{2}$$

That fewer measurements are required for exact reconstruction than when $p = 1$ was demonstrated by numerical experiments in [11], with random and nonrandom Fourier measurements. A similar approach was used by Rao and Kreutz-Delgado [12] for basis selection. In this paper, we consider the context of error correction, and our measurements will be random Gaussian projections. In Section 2 we provide a theoretical result (based on one from [13]) justifying the increased likelihood of exact reconstruction. In Section 3, numerical experiments will show that using $p < 1$ allows perfect recovery from the corruption of a greater number of entries.

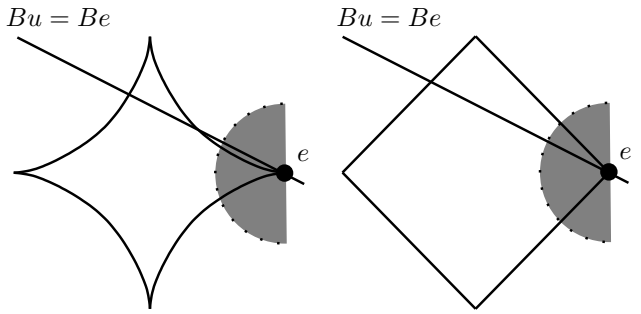
2. ERROR CORRECTION

We consider the abstract encryption framework described in [13]. Let A be an $m \times n$ matrix, with $m > n$. If A has full rank, we can regard it as a linear block cipher, with a plaintext $f \in \mathbb{R}^n$ encrypted as Af. We suppose the ciphertext Af is corrupted by an error vector $e \in \mathbb{R}^m$, with the property that the support of e is at most r: $\|e\|_0 \leq r$. Given the corrupted ciphertext $y = Af + e$, under what circumstances can we recover Af (and hence f) exactly?

This problem can be recast into the form of (2) by the use of a matrix B whose kernel is the range of A. Then $By = B(Af + e) = Be$. We attempt to reconstruct e from the measurement vector $By(= Be)$, by solving (2):

$$\min_d \|d\|_p^p, \quad \text{subject to } Bd = Be. \tag{3}$$

If the unique minimizer is $d = e$, then we will have the error vector e, from which we can recover the plaintext from $Af = y - e$.



(a) $p < 1$ (b) $p = 1$

Fig. 1. Exact reconstruction occurs if the constraint plane meets the ℓ^p sphere containing e only at e. For sparse e, this condition is more likely for $p < 1$ if one only need consider points near e.

Still following [13], the substitution $d = y - A\tilde{f}$ yields the unconstrained problem

$$\min_{\tilde{f}} \|y - A\tilde{f}\|_p^p, \tag{4}$$

as $B(y - A\tilde{f}) = By = Be$ for all $\tilde{f} \in \mathbb{R}^n$. Our numerical experiments in Section 3 will consist of solving (4) and comparing the minimizer with f.

The geometry of (3) is depicted in Figure 1. Exact reconstruction corresponds to e being the only point of intersection of the affine space $Bd = Be$ and the ℓ^p-sphere containing e. If e is sparse, this will be true for many B, seemingly to the same degree whether $p = 1$ or $p < 1$. This changes, however, in higher dimensions, or if a solution of (3) must also be close to e. The smaller p is, and the closer a solution to (3) must be to e, the more likely that a given choice of B will yield exact reconstruction. And it is the sparsity of e that will contribute to the requirement that a minimizer be close to e.

This brings us to the concept of an approximate S-restricted isometry, as introduced in [14]. For a $k \times m$ matrix B and $T \subset \{1, \ldots, m\}$, let B_T be the matrix consisting of the columns b_j of B for $j \in T$. (We will use similar notation for vectors, with $u_T(t) = u(t)$ if $t \in T$ and 0 otherwise.) For each number S, define the S-restricted isometry constant of B to be the smallest $\delta_S \geq 0$ such that for all subsets T with $|T| \leq S$ and all $c \in \mathbb{R}^{|T|}$,

$$(1 - \delta_S)\|c\|_2^2 \leq \|B_T c\|_2^2 \leq (1 + \delta_S)\|c\|_2^2. \tag{5}$$

Thus if T_0 is the support of e, $Bd = Be$, and d is supported on T_0, we will have $\|d - e\|_2^2 \leq \|B(d - e)\|_2^2/(1 - \delta_r) = 0$, provided $\delta_r < 1$. However, there is no guarantee that a minimizer of (3) will be supported on T_0, or even be sparse.

Working in tandem with (5) will be the following observation, essentially from [15]. Let d be a solution of (3), and let $h = d - e$. By the triangle inequality for $\|\cdot\|_p^p$, we have

$$|\|e\|_p^p - \| - h_{T_0}\|_p^p| \leq \|e + h_{T_0}\|_p^p. \tag{6}$$

Since $T_0 \cap T_0^c = \emptyset$, we have

$$\begin{aligned} \|e\|_p^p - \|h_{T_0}\|_p^p + \|h_{T_0^c}\|_p^p &\leq \|e + h_{T_0} + h_{T_0^c}\|_p^p \\ &= \|e + h\|_p^p = \|d\|_p^p \leq \|e\|_p^p, \end{aligned} \tag{7}$$

the last inequality holding because d solves (3). The result is that

$$\|h_{T_0^c}\|_p^p \leq \|h_{T_0}\|_p^p. \tag{8}$$

In other words, although d need not be sparse, a bound exists on the portion of d outside the support of e (note that $d_{T_0^c} = h_{T_0^c}$). The more sparse e is, the stronger (8) is.

The final piece of this picture is the following result. It quantifies the restricted isometry condition necessary for exact reconstruction, and generalizes and improves for $p < 1$ the corresponding result of [13].

Theorem 2.1. *Let the block cipher A be an $m \times n$ matrix with $m > n$. Let $f \in \mathbb{R}^n$ be a plaintext, $e \in \mathbb{R}^m$ an error vector, and let $r = \|e\|_0$ be the size of the support of e. Let B be a matrix whose kernel is the range of A. Let $p \in (0, 1]$, $a = 3^{p/(2-p)}$. Suppose that B satisfies*

$$\delta_{ar} + 3\delta_{(a+1)r} < 2. \tag{9}$$

Then the unique minimizer of (3) *is exactly e, and we can recover the plaintext f exactly from the corrupted ciphertext $y = Af + e$ as the unique minimizer of* (4)*.*

For $p = 1$, this is exactly as appears in [13]. For a given B, the restricted isometry condition (9) will hold for larger values of r when $p < 1$. We thus can expect to be able to correct errors of larger support in this case.

It is also shown in [13] that in the case of random, Gaussian ciphers, the condition of Theorem 2.1 holds (for $p = 1$, *a fortiori* for $p < 1$) with overwhelming probability, provided $r < \rho m$ for some constant ρ. The value of ρ given is very far from sharp, however.

Proof of Theorem 2.1. The proof generally follows the lines of [13], but with a simplification. (Specifically, equation (2.2) therein is not required.) As above, we consider a solution d of (3) (that such exists is geometrically obvious). Let $h = d - e$; we wish to show that $h = 0$. Let T_0 be the support of e. Let $M = ar$. Arrange the elements of T_0^c in order of decreasing magnitude of $|h|$ and partition into $T_0^c = T_1 \cup T_2 \cup \cdots \cup T_L$, where each T_j has M elements (except possibly T_L). We do this because the restricted isometry condition gives us control over the action of B on small sets. Denote $T_{01} = T_0 \cup T_1$.

We decompose Bh:

$$\begin{aligned}
0 = \|Bd - Be\|_2 = \|Bh\|_2 &= \left\| B_{T_{01}} h_{T_{01}} + \sum_{j=2}^{L} B_{T_j} h_{T_j} \right\|_2 \\
&\geq \|B_{T_{01}} h_{T_{01}}\|_2 - \left\| \sum_{j=2}^{L} B_{T_j} h_{T_j} \right\|_2 \\
&\geq \|B_{T_{01}} h_{T_{01}}\|_2 - \sum_{j=2}^{L} \|B_{T_j} h_{T_j}\|_2 \\
&\geq \sqrt{1-\delta_{M+r}} \|h_{T_{01}}\|_2 - \sqrt{1+\delta_M} \sum_{j=2}^{L} \|h_{T_j}\|_2.
\end{aligned} \tag{10}$$

Now we need to control the size of the $\|h_{T_j}\|_2$. We aim to use (8), for which we must estimate the ℓ^2 norm in terms of the ℓ^p norm. For each $t \in T_j$ and $s \in T_{j-1}$, $|h(t)| \leq |h(s)|$, so that

$$|h(t)|^p \leq \|h_{T_{j-1}}\|_p^p / M. \tag{11}$$

Then

$$\|h_{T_j}\|_2^2 \leq M \|h_{T_{j-1}}\|_p^2 / M^{2/p}, \tag{12}$$

so that

$$\begin{aligned}
\sum_{j=2}^{L} \|h_{T_j}\|_2 &\leq \left(\sum_{j=1}^{L} \|h_{T_j}\|_p \right) / M^{1/p-1/2} \\
&\leq \|h_{T_0^c}\|_p / M^{1/p-1/2},
\end{aligned} \tag{13}$$

where we have used the reverse triangle inequality property of the ℓ^p norm for $p \leq 1$. Now we may use (8), and then convert back from ℓ^p to ℓ^2 by means of Hölder's inequality:

$$\begin{aligned}
\|h_{T_0}\|_p^p = \sum_{t \in T_0} |h(t)|^p \cdot 1 &\leq \left(\sum_{T_0} |h(t)|^2 \right)^{\frac{p}{2}} \left(\sum_{T_0} 1 \right)^{1-\frac{p}{2}} \\
&= \|h_{T_0}\|_2^p |T_0|^{1-p/2}.
\end{aligned} \tag{14}$$

Combining, we obtain

$$\begin{aligned}
\sum_{j=2}^{L} \|h_{T_j}\|_2 \leq \|h_{T_0}\|_p / M^{1/p-1/2} &\leq \|h_{T_0}\|_2 \left(\frac{|T_0|}{M} \right)^{\frac{1}{p}-\frac{1}{2}} \\
&= \|h_{T_0}\|_2 / \sqrt{3}.
\end{aligned} \tag{15}$$

Putting together with (10), we have

$$\begin{aligned}
0 &\geq \sqrt{1-\delta_{M+r}} \|h_{T_{01}}\|_2 - \sqrt{1+\delta_M} \|h_{T_0}\|_2 / \sqrt{3} \\
&\geq \left(\sqrt{1-\delta_{M+r}} - \sqrt{1+\delta_M} / \sqrt{3} \right) \|h_{T_{01}}\|_2.
\end{aligned} \tag{16}$$

The condition (9) of the theorem ensures that the scalar factor is positive, so $h_{T_{01}} = 0$. In particular, $h_{T_0} = 0$; then $h = 0$ follows from (8). □

3. NUMERICAL EXPERIMENTS

We present the results of numerical experiments investigating the ability of (4) to reconstruct a plaintext from a corrupted ciphertext. We adopt the approach of [13], to facilitate direct comparison. We used $n = 128$, and both $m = 256$ and $m = 512$. For each m, we used 20 different values of r, chosen as a percentage of m. For each value of m and r, the following was repeated 100 times. The elements of the $m \times n$ cipher A and the $n \times 1$ plaintext f were randomly chosen from the standard normal distribution. The r entries to be corrupted were randomly chosen, and the corresponding values of the error vector e were chosen from the standard normal distribution. We let $y = Af + e$, and computed a local minimizer f^* of (4), for each $p \in \{0.1, 0.2, \ldots, 1\}$. The reconstruction was deemed exact if every entry of $|f^* - f|$ was less than 10^{-6}; for $p < 1$, such "exact" maximum residuals were generally less than 10^{-13}. Further iteration of the algorithm below would generally reduce $p = 1$ residuals below 10^{-13} as well.

To compute a local minimizer of (4), we used an algorithm based on the lagged-diffusivity algorithm of Vogel and Oman [16] for total-variation minimization. Consider the Euler-Lagrange equation for (4):

$$A^T |A\tilde{f} - y|^{p-2} (A\tilde{f} - y) = 0. \tag{17}$$

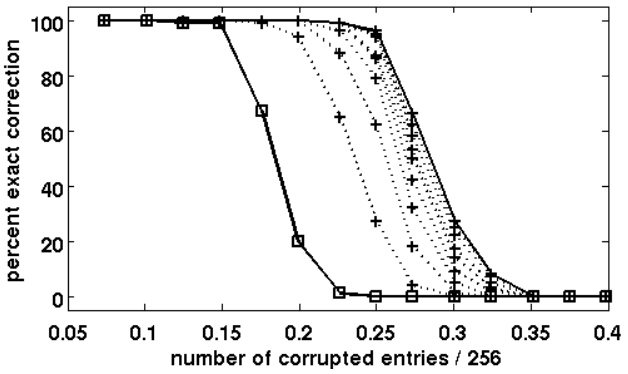
Given the nth iterate $\tilde{f}_n$, we solve for the next iterate $\tilde{f}_{n+1}$ by "lagging" the nonlinear terms in (17), resulting in a linear equation:

$$A^T |A\tilde{f}_n - y|^{p-2} A\tilde{f}_{n+1} = A^T |A\tilde{f}_n - y|^{p-2} y. \tag{18}$$

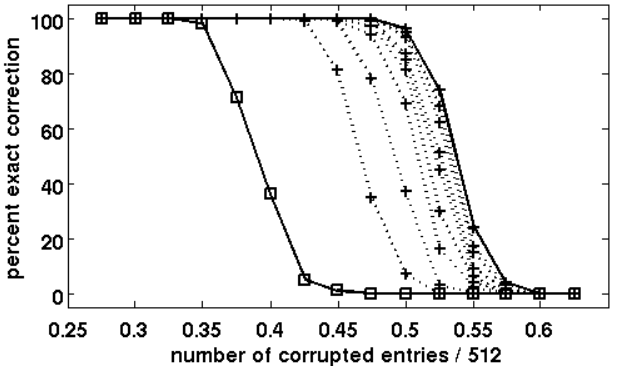
The iteration was begun with the least-squares solution (that for $p = 2$). To avoid division by zero, $|A\tilde{f} - y|$ was approximated by $\left((A\tilde{f} - y)^2 + \epsilon\right)^{1/2}$. The value of ϵ was initially set to 1, and the minimizer computed. The process was then iterated with ϵ 100 times smaller than the previous value, and with the previous minimizer used as the initial iterate, a total of 10 times. The entire process took approximately 9 seconds on a 2.8 GHz processor for $m = 512$, 3 seconds for $m = 256$.

Results of the experiments are plotted in Figure 2. Call the corruption rate $\rho = r/m$. For plaintext size $n = 256$ and $p = 1$, exact reconstruction occurred all 100 times for a corruption rate of $\rho \leq 10\%$, and 99 times for $\rho = 15\%$. Using $p = 0.9$ gave exact reconstruction 100 times for $\rho \leq 15\%$ and 99 times for $\rho = 17.5\%$. For $p = 0.8$ or less, exact reconstruction always occurred for $\rho \leq 20\%$.

When the plaintext size was $n = 512$, exact reconstruction occurred always for $p = 1$ when $\rho \leq 32.5\%$, 99% of the time for $\rho = 35\%$. For $p = 0.9$, we had 100% exact reconstruction for $\rho \leq 40\%$, and 99 times for $\rho = 42.5\%$. Decreasing p to 0.8 or 0.7 increased the corresponding values of ρ to 42.5% and 45%. For $p = 0.6$ and 0.5, exact reconstruction always occurred for $\rho \leq 45\%$. For $p \leq 0.4$, this happened for $\rho \leq 47.5\%$.



(a) $m = 256$



(b) $m = 512$

Fig. 2. Plots of observed frequency of exact reconstruction versus sparsity of ciphertext errors, for values of p used in ℓ^p minimization ranging from 1 (solid line, square marker) down to 0.1 (solid line, '+' marker; values from 0.9 to 0.2 are dotted lines, from left to right). Even $p = 0.9$ shows substantial improvement over $p = 1$. When m, the ciphertext size, is 256, decreasing p from 1 to 0.8 or lower allows an additional 25 entries to be corrupted and still expect exact reconstruction of the plaintext. For $m = 512$, 77 more entries can be corrupted by decreasing p to 0.4 or lower.

Considering all observed probabilities of exact reconstruction, from the plots we see that even a decrease of p from 1 to 0.9 results in a substantial improvement. Decreasing p further yields improvement, but by less and less as p gets smaller.

4. REFERENCES

[1] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, pp. 1289–1306, 2006.

[2] E. J. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, 2006.

[3] M. Lustig, J. M. Santos, J.-H. Lee, D. L. Donoho, and J. M. Pauly, "Application of compressed sensing for rapid MR imaging," in *SPARS*, (Rennes, France), 2005.

[4] D. Takhar, J. N. Laska, M. B. Wakin, M. F. Duarte, D. Baron, S. Sarvotham, K. F. Kelly, and R. G. Baraniuk, "A new compressive imaging camera architecture using optical-domain compression," in *Computational Imaging IV - Proceedings of SPIE-IS and T Electronic Imaging*, vol. 6065, 2006.

[5] M. F. Duarte, S. Sarvotham, D. Baron, M. B. Wakin, and R. G. Baraniuk, "Distributed compressed sensing of jointly sparse signals," in *39th Asilomar Conference on Signals, Systems and Computers*, pp. 1537–1541, 2005.

[6] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, pp. 33–61, 1998.

[7] E. Candes and T. Tao, "Near optimal signal recovery from random projections: universal encoding strategies?," Tech. Rep. 04-70, UCLA Group in Computational and Applied Mathematics, December 2004.

[8] J. A. Tropp and A. C. Gilbert, "Signal recovery from partial information via orthogonal matching pursuit." Preprint.

[9] M. F. Duarte, M. B. Wakin, and R. G. Baraniuk, "Fast reconstruction of piecewise smooth signals from incoherent projections," in *SPARS*, (Rennes, France), 2005.

[10] C. La and M. N. Do, "Signal reconstruction using sparse tree representations," in *Wavelets XI*, vol. 5914, SPIE, 2005.

[11] R. Chartrand, "Exact reconstruction via nonconvex minimization." Submitted, 2006.

[12] B. D. Rao and K. Kreutz-Delgado, "An affine scaling methodology for best basis selection," *IEEE Trans. Signal Process.*, vol. 47, pp. 187–200, 1999.

[13] E. Candes, M. Rudelson, T. Tao, and R. Vershynin, "Error correction via linear programming," in *46th Annual IEEE Symposium on Foundations of Computer Science, 23-25 Oct. 2005, Pittsburgh, PA, USA*, pp. 295–308, 2005.

[14] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, pp. 4203–4215, 2005.

[15] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. Inf. Theory*, vol. 47, pp. 2845–2862, 2001.

[16] C. R. Vogel and M. E. Oman, "Iterative methods for total variation denoising," *SIAM J. Sci. Comput.*, vol. 17, no. 1, pp. 227–238, 1996.