DENSE ERROR CORRECTION VIA L1-MINIMIZATION

John Wright and Yi Ma

Department of Electrical and Computer Engineering
University of Illinois at Urbana-Champaign
1308 West Main Street, Urbana, Illinois 61801, USA

ABSTRACT

We study the problem of recovering a non-negative sparse signal \( x \in \mathbb{R}^n \) from highly corrupted linear measurements \( y = Ax + e \), where \( e \) is an unknown (and unbounded) error. Motivated by an observation from computer vision, we prove that for highly correlated dictionaries \( A \), any non-negative, sufficiently sparse signal \( x \) can be recovered by solving an \( \ell^1 \)-minimization problem:

\[
\min \| x \|_1 + \| e \|_1 \quad \text{subject to} \quad y = Ax + e.
\]

If the fraction \( \rho \) of errors is bounded away from one and the support of \( x \) grows sublinearly in the dimension \( m \) of the observation, for large \( m \), the above \( \ell^1 \)-minimization recovers all sparse signals \( x \) from almost all sign-and-support patterns of \( e \). This suggests that accurate and efficient recovery of sparse signals is possible even with nearly 100% of the observations corrupted.

Index Terms— Error correction, Signal representation, Signal reconstruction

1. INTRODUCTION

Recovery of high-dimensional sparse signals or errors has been one of the fastest growing research areas in signal processing in the past few years. A lot of excitement has been generated by remarkable successes in application areas such as image and speech processing, bioinformatics, communications, as well as computer vision and pattern recognition.\(^1\)

One notable, and somewhat non-traditional, application of sparse representation is in automatic face recognition [3]. For each person, a set of training images are taken under different illuminations. Stack the images as columns of a matrix \( A \in \mathbb{R}^{m \times n} \), where \( m \) is the number of pixels in an image and \( n \) is the total number of images for all the subjects of interest. We can try to represent a new query image, stacked as a vector \( y \in \mathbb{R}^m \) as a linear combination of all the images, i.e., \( y = Ax \) for some \( x \in \mathbb{R}^n \). Since in practice \( n \) can potentially be larger than \( m \), the equations can be underdetermined and the solution \( x \) may not be unique. In this context, it is natural to seek a sparse solution for \( x \) whose large non-zero coefficients provide information about the subject’s true identity. This can be done by solving an \( \ell^1 \)-minimization problem:

\[
\min_x \| x \|_1 \quad \text{subject to} \quad y = Ax.
\]

The problem becomes more interesting if the query image \( y \) is severely occluded or corrupted, as in Figure 1 (inset). In this case, one needs to solve a corrupted set of linear equations \( y = Ax + e \), where \( e \in \mathbb{R}^m \) is an unknown (and possibly unbounded) error vector. For sparse errors \( e \) and tall matrices \( A (m > n) \), Candes and Tao [4] proposed to multiply the equation \( y = Ax + e \) with a matrix \( B \) such that \( BA = 0 \), and then use \( \ell^1 \)-minimization to recover the error vector \( e \) from the underdetermined linear equation \( By = Be \).

In face recognition (and many other applications), \( n \) can be larger than \( m \) and \( A \) can be full rank. One cannot directly apply the above technique even if the error \( e \) is known to be very sparse. To resolve this difficulty, in [3], the authors proposed to instead seek \( [x, e] \) together as the sparsest solution to the extended equation \( y = [A \ I] w \) with \( w = [x, e] \in \mathbb{R}^{m+n} \), by solving the extended \( \ell^1 \)-minimization problem:

\[
\min_w \| w \|_1 \quad \text{subject to} \quad y = [A \ I] w.
\]

---

\(^1\)This work is partially supported by grants NSF CRS-EHS-0509151, NSF CCF-TF-0514955, ONR YIP N00014-05-1-0633, and NSF IIS 07-03756. John Wright is also supported by a Microsoft Fellowship (sponsored by Microsoft Live Labs, Redmond).

\(^2\)For a more thorough survey of this rapidly expanding literature, see [1, 2].
the dimension $m$ increases (and the sample size $n$ grows in proportion), the percentage of errors that the $\ell^1$-minimization (2) can correct seems to approach 100%! This may seem surprising, but this paper explains why this should be expected.

2. PROBLEM SETTING AND MAIN RESULT

Motivated by the face recognition example introduced above, we consider the problem of recovering a non-negative sparse signal $x_0 \in \mathbb{R}^n$ from highly corrupted observations $y \in \mathbb{R}^m$:

$$y = Ax_0 + e_0,$$

where $e_0 \in \mathbb{R}^m$ is a sparse vector of errors of arbitrary magnitude. The model for $A \in \mathbb{R}^{m \times n}$ should capture the idea that it consists of small deviations about a mean, hence a “bouquet.” In this paper, we consider the case where the columns of $A$ are iid samples from a Gaussian distribution:

$$A = [a_1 \ldots a_n] \in \mathbb{R}^{m \times n}, \quad a_i \sim iid N \left(\mu, \frac{\sigma^2}{m} I_m\right),$$

Together, the two assumptions on the mean force it to remain incoherent with the standard basis (or “cross”) as $m \to \infty$.

We study the behavior of the solution to the $\ell^1$-minimization (2) for this model, in the following asymptotic scenario:

**Assumption 1 (Weak Proportional Growth).** A sequence of signal-error problems exhibits weak proportional growth with parameters $\delta > 0, \rho \in (0, 1), C_0 > 0, \eta_0 > 0$, denoted WPG $0, \rho, C_0, \eta_0$, if as $m \to \infty$,

$$\frac{n}{m} \to \delta, \quad \frac{\|e_0\|_0}{m} \to \rho, \quad \|x_0\|_0 \leq C_0 m^{1-\eta_0}. \quad (4)$$

This should be contrasted with the “total proportional growth” (TPG) setting of, e.g., [6], in which the number of nonzero entries $k_1$ in the signal $x_0$ also grows as a fixed fraction of the dimension. In that setting, one might expect a sharp phase transition in the combined sparsity of $(x_0, e_0)$ that can be recovered by $\ell^1$-minimization. In WPG, on the other hand, we observe a striking phenomenon not seen in TPG: the correction of arbitrary fractions of errors. This comes at the expense of the stronger assumption that $k_1 \approx \|x_0\|_0$ is sublinear, an assumption that is valid in some real applications such as the face recognition example above.

In the following, we say the cross-and-bouquet model is $\ell^1$-recoverable at $(I, J, \sigma)$ if for all $x_0 \geq 0$ with support $I$ and $e_0$ with support $J$ and signs $\sigma$,

$$(x_0, e_0) = \arg \min_x \|x\|_1 + \|e\|_1$$

subject to $Ax + e = Ax_0 + e_0$, (5)

and the minimizer is uniquely defined. From the geometry of $\ell^1$-minimization, if (5) does not hold for some pair $(x_0, e_0)$,
then it does not hold for any \((x, e)\) with the same signs and support as \((x_0, e_0)\) [7]. Understanding \(\ell^1\)-recoverability at each \((I, J, \sigma)\) completely characterizes which solutions to \(y = Ax + e\) can be correctly recovered. In this language, our main result can be stated more precisely as:

**Theorem 1** (Error Correction with the Cross-and-Bouquet). For any \(\delta > 0, \exists \nu_0(\delta) > 0\) such that if \(\nu < \nu_0\) and \(\rho < 1\), in WPGs, with \(A \in \mathcal{D}(\rho, \nu)\), with \(A\) distributed according to (3), if the error support \(J\) and signs \(\sigma\) are chosen uniformly at random, then as \(m \to \infty\),

\[
P_{A,I,\sigma}\left[\ell^1\text{-recoverability at } (I, J, \sigma) \; \forall \; I \in \binom{[n]}{k_1}\right] \to 1.
\]

In other words, as long as the bouquet is sufficiently tight, asymptotically \(\ell^1\)-minimization recovers any non-negative sparse signal from almost any error with support size less than 100%. The proof of the above result relies on a careful characterization of the faces of the polytope spanned by the cross and bouquet. While it requires only standard ideas from geometry, linear algebra and measure concentration, the details are far beyond the scope of this paper. We refer the interested reader to [2].

### 3. SIMULATIONS AND EXPERIMENTS

**a) Comparison with alternative approaches.** We first compare the performance of the extended \(\ell^1\)-minimization (2) to two alternative approaches. The first is the error correction approach of [4], which multiplies by a full rank matrix \(B\) such that \(BA = 0\), solves \(\min \|e\|_1\) subject to \(Be = By\), and then subsequently recovers \(x\) from the clean system of equations \(Ax = y - e\). The second is the Regularized Orthogonal Matching Pursuit (ROMP) algorithm [8], a state-of-the-art greedy method for recovering sparse signals.

For this experiment, the ambient dimension is \(m = 500\); the parameters of the CAB model are \(\nu = 0.05, \delta = 0.25\). We fix the signal support \(k_1 = 15\), and vary the fraction of errors from 0 to 0.95. For each error fraction, we generate 500 independent problems. Figure 3 plots the fraction of successes for each of the three algorithms, as a function of error density \(\rho\). The extended \(\ell^1\)-minimization is denoted \(L^1 - \lfloor A \rfloor\) (red curve), while the alternative approach of [4] is denoted \(L^1 - \perp\) comp (blue curve). Whereas both competitors break down around 40% corruption, the extended \(\ell^1\)-minimization continues to succeed with high probability even beyond 60% corruption.

**b) Error correction capacity.** While the previous experiment demonstrates the advantages of the extended \(\ell^1\)-minimization (2) for the CAB model, Theorem 1 suggests that more is true: As the dimension increases, the fraction of errors that the extended \(\ell^1\)-minimization can correct should approach one. We generate problem instances with \(\delta = 0.25, \nu = 0.05\), for varying \(m = 100\), \(200\), \(400\), \(800\), \(1600\). We again plot the fraction of correct recoveries as a function of \(\rho\) in Figure 4 (a) and (b). In Figure 4(a), we fix \(k_1 = 1\), while in (b), \(k_1 = m^{1/2}\). In both cases, as \(m\) increases, the fraction of errors that can be corrected approaches 1.

**c) Phase Transition in Total Proportional Growth.** Theorem 1 does not provide any explicit information about the behavior of \(\ell^1\)-minimization when the signal support \(k_1\) grows proportionally to \(m\): \(k_1/m \to \rho_1 \in (0,1)\). Based on intuition from more homogeneous polytopes (especially [9]), we might expect that when \(k_1\) also exhibits proportional growth, an asymptotically sharp phase transition between guaranteed recovery and guaranteed failure will occur at some critical error fraction \(\rho^* \in (0,1)\). We investigate this empirically.

![Figure 3](image-url)  
*Fig. 3. Comparison with alternative approaches.* Fraction of correct successes, as a function of the corruption level, \(\rho\). Here, \(\nu = 0.05, \delta = 0.25\). The extended \(\ell^1\)-minimization \("L^1 - \lfloor A \rfloor\)" outperforms the approach of [4] \("L^1 - \perp\) comp") and ROMP [8].

![Figure 4](image-url)  
*Fig. 4. Error correction in weak proportional growth.* (a), (b): Simulated examples with \(\delta = 0.25, \nu = 0.05\). Fraction of successful recoveries as a function of error density \(\rho\), for varying \(m\). In (a), \(\|x_0\|_0 = 1\), while in (b), \(\|x_0\|_0 = m^{1/2}\). In both cases, as \(m\) increases, the fraction of errors that can be corrected approaches 1.
here by again setting $\delta = 0.25$, $\nu = 0.05$, but this time allowing $k_1 = 0.05m$. Figure 5 plots the fraction of correct recovery for varying error fractions $\rho$, as $m$ grows: $m = 100, 200, 400, 600, 1600$. In this proportional growth setting, we see an increasingly sharp phase transition, near $\rho = 0.6$.

d) Error correction with real face images. Finally, we return to the motivating example of face recognition under varying illumination and random corruption. We use the Extended Yale B face database [10], which tests illumination sensitivity of face recognition algorithms. We form the matrix $A$ from images in Subsets 1 and 2, which contain mild-to-moderate illumination variations. Each column of the matrix $A$ is a $w \times h$ face image, stacked as a vector in $\mathbb{R}^m$ ($m = w \times h$). Here, the weak proportional growth setting corresponds to the case when the total number of image pixels grows proportionally to the number $n$ of face images. Since the number of images per subject is fixed, this is the same as the total image resolution growing proportionally to the number of subjects. We vary the image resolutions through the range $34 \times 30, 48 \times 42, 68 \times 60, 96 \times 84$. The matrix $A$ is formed from images of 4, 9, 19, 38 subjects, respectively, corresponding to $\delta \approx 0.09$. Here, $\nu \approx 0.3$. In face recognition, the sublinear growth of $\|x_0\|_0$ comes from the fact that the observation should ideally be a linear combination of only images of the same subject. Various estimates of the required number of images, $k_1$, appear in the literature, ranging from 5 to 9. Here, we fix $k_1 = 7$, and generate the (clean) test image synthetically as a linear combination of $k_1$ training images from a single subject. For each resolution considered, and for each error fraction, we generate 75 trials. Figure 6 plots the fraction of successes as a function of the fraction of corruption. As predicted by Theorem 1, the fraction of errors that can be corrected again approaches 1 as the data size increases.

4. DISCUSSIONS AND FUTURE WORK

This work analyzes one scenario, motivated by a practical imaging application, in which the performance of $\ell^1$-minimization greatly exceeds what might be expected based on existing theory. We believe that similar analysis of other practical applications may likewise reveal phenomena of broad practical and theoretical interest. Even for this simple model, there is still much to be done. In particular, while the $\ell^1$-minimizer is known to be stable under noise, it would be interesting to provide an explicit stability bound, as a function of $\nu$. We would also like to investigate the relevance of this result to compressive image acquisition problems, by analyzing how much error tolerance remains after randomly projecting $y$ onto a low dimensional subspace.

5. REFERENCES