General Perturbations of Sparse Signals in Compressed Sensing

Matthew A. Herman and Thomas Strohmer

Department of Mathematics, University of California, Davis, CA 95616-8633, USA.
{mattyh,strohmer}@math.ucdavis.edu

Abstract:

We analyze the Basis Pursuit recovery of signals when observing sparse data with general perturbations. Previous studies have only considered partially perturbed observations $\boldsymbol{Ax}+\boldsymbol{e}$. Here, $\boldsymbol{x}$ is a K-sparse signal which we wish to recover, $\boldsymbol{A}$ is a measurement matrix with more columns than rows, and $\boldsymbol{e}$ is simple *additive* noise. Our model also incorporates perturbations $\boldsymbol{E}$ (which result in *multiplicative* noise) to the matrix $\boldsymbol{A}$ in the form of $(\boldsymbol{A}+\boldsymbol{E})\boldsymbol{x}+\boldsymbol{e}$. This completely perturbed framework extends the previous work of Candès, Romberg and Tao on stable signal recovery from incomplete and inaccurate measurements. Our results show that, under suitable conditions, the stability of the recovered signal is limited by the noise level in the observation. Moreover, this accuracy is within a constant multiple of the best-case reconstruction using the technique of least squares.

1. Introduction

Employing the techniques of compressed sensing (CS) to recover signals with a sparse representation has enjoyed a great deal of attention over the last 5–10 years. The initial studies considered an ideal unperturbed scenario:

$$\boldsymbol{b} \;=\; \boldsymbol{Ax}. \tag{1}$$

Here $\boldsymbol{b} \in \mathbb{C}^m$ is the observation vector, $\boldsymbol{A} \in \mathbb{C}^{m\times n}$ ($m \leq n$) is a full-rank measurement matrix or system model, and $\boldsymbol{x} \in \mathbb{C}^n$ is the signal of interest which has a K-sparse representation (i.e., it has no more than K nonzero coefficients) under some fixed basis. More recently researchers have included an *additive* noise term $\boldsymbol{e}$ into the received signal [1, 2, 4, 8], creating a *partially perturbed model*:

$$\hat{\boldsymbol{b}} \;=\; \boldsymbol{Ax}+\boldsymbol{e} \tag{2}$$

This type of noise generally models simple, uncorrelated errors in the data or at the receiver/sensor.

As far as we can tell, practically no research has been done yet on perturbations $\boldsymbol{E}$ to the matrix $\boldsymbol{A}$. Our ***completely perturbed model*** extends (2) by incorporating a perturbed sensing matrix in the form of

$$\hat{\boldsymbol{A}} = \boldsymbol{A}+\boldsymbol{E}.$$

It is important to consider this kind of noise since it can account for precision errors when applications call for physically implementing the matrix $\boldsymbol{A}$ in a sensor. When $\boldsymbol{A}$ represents a system model, such as in the context of radar [7] or telecommunications, then $\boldsymbol{E}$ can absorb errors in assumptions made about the transmission channel, as well as quantization errors arising from the discretization of analog signals. In general, these perturbations can be characterized as *multiplicative noise*, and are more difficult to analyze than simple additive noise since they are correlated with the signal of interest. To see this, simply substitute $\boldsymbol{A} = \hat{\boldsymbol{A}} - \boldsymbol{E}$ in (2); there will be an extra noise term $\boldsymbol{Ex}$. (Note that it makes no difference whether we account for the perturbation $\boldsymbol{E}$ on the "encoding side" (2), or on the "decoding side" (7). The model used here was chosen so as to agree with the conventions of classical perturbation theory which we use in Section 4.)

1.1 Assumptions and Notation

Without loss of generality, assume the original data $\boldsymbol{x}$ to be a K-sparse vector for some fixed K. Denote $\sigma_{\max}^{(K)}(\boldsymbol{Y})$, $\|\boldsymbol{Y}\|_2^{(K)}$, and $\text{rank}^{(K)}(\boldsymbol{Y})$ respectively as the maximum singular value, spectral norm, and rank over all K-column submatrices of a matrix $\boldsymbol{Y}$. Similarly, $\sigma_{\min}^{(K)}(\boldsymbol{Y})$ is the minimum singular value over all K-column submatrices of $\boldsymbol{Y}$. Let the perturbations in (2) be relatively bounded by

$$\frac{\|\boldsymbol{E}\|_2^{(K)}}{\|\boldsymbol{A}\|_2^{(K)}} \leq \varepsilon_{\boldsymbol{A}}^{(K)}, \quad \frac{\|\boldsymbol{e}\|_2}{\|\boldsymbol{b}\|_2} \leq \varepsilon_{\boldsymbol{b}} \tag{3}$$

with $\|\boldsymbol{A}\|_2^{(K)}, \|\boldsymbol{b}\|_2 \neq 0$. In the real world we are only interested in the case where both $\varepsilon_{\boldsymbol{A}}^{(K)}, \varepsilon_{\boldsymbol{b}} < 1$.

2. CS ℓ_1 Perturbation Analysis

2.1 Previous Work

In the *partially perturbed scenario* (i.e., $\boldsymbol{E} = \boldsymbol{0}$ in (2)) we are concerned with solving the *Basis Pursuit* (BP) problem [3]:

$$\boldsymbol{z}^\star = \underset{\hat{\boldsymbol{z}}}{\text{argmin}}\, \|\hat{\boldsymbol{z}}\|_1 \quad \text{s.t.} \quad \|\boldsymbol{A}\hat{\boldsymbol{z}} - \hat{\boldsymbol{b}}\|_2 \leq \varepsilon' \tag{4}$$

for some $\varepsilon' \geq 0$.

The *restricted isometry property* (RIP) [2] for any matrix $\boldsymbol{A} \in \mathbb{C}^{m\times n}$ defines, for each integer $K = 1, 2, \ldots$, the *restricted isometry constant* (RIC) δ_K, which is the smallest nonnegative number such that

$$(1-\delta_K)\|\boldsymbol{x}\|_2^2 \;\le\; \|\boldsymbol{A}\boldsymbol{x}\|_2^2 \;\le\; (1+\delta_K)\|\boldsymbol{x}\|_2^2 \qquad (5)$$

holds for any K-sparse vector $\boldsymbol{x}$. In the context of the RIC, we observe that $\|\boldsymbol{A}\|_2^{(K)} = \sigma_{\max}^{(K)}(\boldsymbol{A}) = \sqrt{1+\delta_K}$, and $\sigma_{\min}^{(K)}(\boldsymbol{A}) = \sqrt{1-\delta_K}$.

Assuming K-sparse $\boldsymbol{x}$, $\delta_{2K} < \sqrt{2}-1$ and $\|\boldsymbol{e}\|_2 \le \varepsilon'$, Candès has shown in Theorem 1.2 of [1] that the solution to (4) obeys

$$\|\boldsymbol{z}^\star - \boldsymbol{x}\|_2 \;\le\; C_{\mathrm{BP}}\,\varepsilon' \qquad (6)$$

for some constant C_{BP}.

2.2 Incorporating nontrivial perturbation $\boldsymbol{E}$

Now assume the ***completely perturbed*** situation with $\boldsymbol{E}, \boldsymbol{e} \neq \boldsymbol{0}$ in (2). In this case the BP problem of (4) can be generalized to include a different decoding matrix $\hat{\boldsymbol{A}}$:

$$\boldsymbol{z}^\star = \operatorname*{argmin}_{\hat{\boldsymbol{z}}} \|\hat{\boldsymbol{z}}\|_1 \quad \text{s.t.} \quad \|\hat{\boldsymbol{A}}\hat{\boldsymbol{z}} - \hat{\boldsymbol{b}}\|_2 \;\le\; \varepsilon'_{\boldsymbol{A},K,\boldsymbol{b}} \qquad (7)$$

for some $\varepsilon'_{\boldsymbol{A},K,\boldsymbol{b}} \ge 0$. The following two theorems summarize our results.

Theorem 1 (RIP for $\hat{\boldsymbol{A}}$)**.** *For any $K = 1, 2, \ldots,$ assume and fix the RIC δ_K associated with $\boldsymbol{A}$, and the relative perturbation $\varepsilon_{\boldsymbol{A}}^{(K)}$ associated with $\boldsymbol{E}$ in (3). Then the RIC*

$$\hat{\delta}_K \;:=\; \big(1+\delta_K\big)\Big(1+\varepsilon_{\boldsymbol{A}}^{(K)}\Big)^2 - 1 \qquad (8)$$

for matrix $\hat{\boldsymbol{A}} = \boldsymbol{A} + \boldsymbol{E}$ is the smallest nonnegative constant such that

$$(1-\hat{\delta}_K)\|\boldsymbol{x}\|_2^2 \;\le\; \|\hat{\boldsymbol{A}}\boldsymbol{x}\|_2^2 \;\le\; (1+\hat{\delta}_K)\|\boldsymbol{x}\|_2^2 \qquad (9)$$

holds for any K-sparse vector $\boldsymbol{x}$.

Remark 1. The flavor of the RIP is defined with respect to the square of the operator norm. That is, $(1-\delta_K)$ and $(1+\delta_K)$ are measures of the ***square*** of minimum and maximum singular values of $\boldsymbol{A}$, and similarly for $\hat{\boldsymbol{A}}$. In keeping with the convention of classical perturbation theory however, we defined $\varepsilon_{\boldsymbol{A}}^{(K)}$ in (3) just in terms of the operator norm (not its square). Therefore, the quadratic dependence of $\hat{\delta}_K$ on $\varepsilon_{\boldsymbol{A}}^{(K)}$ in (8) makes sense. Moreover, in discussing the spectrum of $\hat{\boldsymbol{A}}$, we see that it is really a *linear function* of $\varepsilon_{\boldsymbol{A}}^{(K)}$.

Theorem 2 (Completely perturbed observation)**.** *Fix the relative perturbations $\varepsilon_{\boldsymbol{A}}^{(K)}$, $\varepsilon_{\boldsymbol{A}}^{(2K)}$ and $\varepsilon_{\boldsymbol{b}}$ in (3). Assume that the RIC for matrix $\boldsymbol{A}$ satisfies $\delta_{2K} < \sqrt{2}\big(1+\varepsilon_{\boldsymbol{A}}^{(2K)}\big)^{-2} - 1$. Set*

$$\varepsilon'_{\boldsymbol{A},K,\boldsymbol{b}} := \Big(c\,\varepsilon_{\boldsymbol{A}}^{(K)} + \varepsilon_{\boldsymbol{b}}\Big)\,\|\boldsymbol{b}\|_2, \qquad (10)$$

where $c = \frac{\sqrt{1+\delta_K}}{\sqrt{1-\delta_K}}$. If $\boldsymbol{x}$ is K-sparse, then the solution to the BP problem (7) obeys

$$\|\boldsymbol{z}^\star - \boldsymbol{x}\|_2 \;\le\; C_{BP}\,\varepsilon'_{\boldsymbol{A},K,\boldsymbol{b}}, \qquad (11)$$

where

$$C_{BP} := \frac{4\sqrt{1+\delta_{2K}}\,\Big(1+\varepsilon_{\boldsymbol{A}}^{(2K)}\Big)}{1-(\sqrt{2}+1)\left(\big(1+\delta_{2K}\big)\Big(1+\varepsilon_{\boldsymbol{A}}^{(2K)}\Big)^2 - 1\right)}. \qquad (12)$$

Remark 2. Theorem 2 generalizes of Candès' results in [1] for K-sparse $\boldsymbol{x}$. Indeed, if matrix $\boldsymbol{A}$ is unperturbed, then $\boldsymbol{E} = \boldsymbol{0}$ and $\varepsilon_{\boldsymbol{A}}^{(K)} = 0$. It follows that $\hat{\delta}_K = \delta_K$ in (8), and the RIPs for $\boldsymbol{A}$ and $\hat{\boldsymbol{A}}$ coincide. Moreover, the condition in Theorem 2 reduces to $\delta_K < \sqrt{2}-1$, and the total perturbation (see (17)) collapses to $\|\boldsymbol{e}\|_2 \le \varepsilon'_{\boldsymbol{b}} := \varepsilon_{\boldsymbol{b}}\|\boldsymbol{b}\|_2$; both of these are identical to Candès' assumptions in (6). Finally, the constant C_{BP} in (12) reduces to the same as outlined in the proof of [1].

It is also interesting to examine the spectral effects due to the assumptions of Theorem 2. Namely, we want to be assured that the rank of submatrices of $\boldsymbol{A}$ are unaltered by the perturbation $\boldsymbol{E}$.

Lemma 1. *If the hypothesis of Theorem 2 is satisfied, then for any $k \le 2K$*

$$\sigma_{\max}^{(k)}(\boldsymbol{E}) \;<\; \sigma_{\min}^{(k)}(\boldsymbol{A}), \qquad (13)$$

and therefore

$$\mathit{rank}^{(k)}(\hat{\boldsymbol{A}}) \;=\; \mathit{rank}^{(k)}(\boldsymbol{A}).$$

This fact is necessary (although, not explicitly stated) in the least squares analysis Section 4.

The utility of Theorems 1 and 2 can be understood with two simple numerical examples. Suppose that measurement matrix $\boldsymbol{A}$ in (2) is designed to have an RIC of $\delta_{2K} = 0.100$. Assume, however, that its physical implementation will experience a worst-case relative error of $\varepsilon_{\boldsymbol{A}}^{(2K)} = 5\%$. Then from (8) we can design a matrix $\hat{\boldsymbol{A}}$ with RIC $\hat{\delta}_{2K} = 0.213$ to be used in (7) which will yield a solution whose accuracy is guaranteed by (11) with $C_{\mathrm{BP}} = 9.057$. Note from (12), we see that if there had been no perturbation, then $C_{\mathrm{BP}} = 5.530$.

Consider now a different example. Suppose instead that $\delta_{2K} = 0.200$ and $\varepsilon_{\boldsymbol{A}}^{(2K)} = 1\%$. Then $\hat{\delta}_{2K} = 0.224$ and $C_{\mathrm{BP}} = 9.643$. Here, if $\boldsymbol{A}$ was unperturbed, then we would have had $C_{\mathrm{BP}} = 8.473$.

These numerical examples show how the stability constant C_{BP} of the BP solution gets worse with perturbations to $\boldsymbol{A}$. It must be stressed however, that they represent worst-case instances. It is well-known in the CS community that better performance is normally achieved in practice.

2.3 Numerical Simulations

Numerical simulations were conducted as follows. Gaussian matrices of size 128×512 were randomly generated in MATLAB. The entries of matrix $\boldsymbol{A}$ were normally distributed $\mathcal{N}(0, \sigma_A^2)$ where $\sigma_A^2 = 1/128$, while those of matrix $\boldsymbol{E}$ were $\mathcal{N}(0, \sigma_E^2)$ with $\sigma_E^2 = \varepsilon_{\boldsymbol{A}}^2/128$. The parameter $\varepsilon_{\boldsymbol{A}}$ is a measure of the relative perturbation of matrix $\boldsymbol{A}$ and took on values $\{0, 0.01, 0.05, 0.10\}$. Next, a random

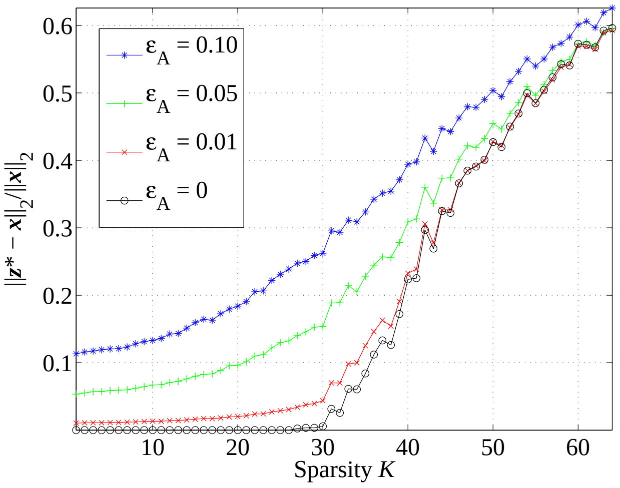

Figure 1: Average (100 trials) relative error of BP solution $\boldsymbol{z}^\star$ with respect to K-sparse $\boldsymbol{x}$ vs. Sparsity K for different relative perturbations $\varepsilon_{\boldsymbol{A}}$ of $\boldsymbol{A} \in \mathbb{C}^{128\times 512}$ (and $\varepsilon_{\boldsymbol{b}} = 0$) .

vector $\boldsymbol{x}$ of sparsity $K = 1, \ldots, 64$ was randomly generated (nonzero entries uniformly distributed with $\mathcal{N}(0,1)$) and $\hat{\boldsymbol{b}} = \boldsymbol{A}\boldsymbol{x}$ in (2) was created (note, we set $\boldsymbol{e} = \boldsymbol{0}$ so as to focus on the effect of perturbation $\boldsymbol{E}$). Given $\hat{\boldsymbol{b}}$ and $\hat{\boldsymbol{A}} = \boldsymbol{A} + \boldsymbol{E}$, the BP program (7) was implemented with cvx software [5]. For each value of $\varepsilon_{\boldsymbol{A}}$ and K, 100 trials were performed.

Fig. 1 shows the average relative error $\|\boldsymbol{z}^\star - \boldsymbol{x}\|_2/\|\boldsymbol{x}\|_2$ as a function of K for each $\varepsilon_{\boldsymbol{A}}$. As a reference, the ideal, noise-free case can be seen for $\varepsilon_{\boldsymbol{A}} = 0$. It is interesting to notice that all perturbations, including $\varepsilon_{\boldsymbol{A}} = 0$, experience significant jumps simultaneously at several places, such as $K = 31, 42, 43, 44$, etc. Now fix a particular value of $K \le 30$ and compare the relative error for the three nonzero values of $\varepsilon_{\boldsymbol{A}}$. It is clear that the error scales roughly linearly with $\varepsilon_{\boldsymbol{A}}$. This empirical study essentially confirms the conclusion of Theorem 2, that the stability of the BP solution scales linearly with $\varepsilon_{\boldsymbol{A}}^{(K)}$ (i.e., the singular values of $\boldsymbol{E}$).

Note that better performance in theory and in simulation can be achieved if BP is used solely to determine the support of the solution. Then we can use least squares to find a better result. This is similar to the the best-case, oracle least squares solution discussed in Section 4.

3. Proofs

3.1 Proof Sketch of Theorem 1

From the triangle inequality, (5) and (3) we have

$$\begin{aligned} \|\hat{\boldsymbol{A}}\boldsymbol{x}\|_2^2 &\le \left(\|\boldsymbol{A}\boldsymbol{x}\|_2 + \|\boldsymbol{E}\boldsymbol{x}\|_2\right)^2 && (14)\\ &\le \left(\sqrt{1+\delta_K} + \|\boldsymbol{E}\|_2^{(K)}\right)^2 \|\boldsymbol{x}\|_2^2 && (15)\\ &\le (1+\delta_K)\left(1+\varepsilon_{\boldsymbol{A}}^{(K)}\right)^2 \|\boldsymbol{x}\|_2^2. && (16) \end{aligned}$$

Moreover, this inequality is sharp for the following reasons:

- Equality occurs in (14) if $\boldsymbol{E}$ is a multiple of $\boldsymbol{A}$.
- Equality occurs in (15) whenever $\boldsymbol{x}$ is in the direction of the vector associated with the value $(1 + \delta_K)$ in the RIP for $\boldsymbol{A}$.
- Equality occurs in (16) since, in this hypothetical case, we assume that $\boldsymbol{E} = \beta\boldsymbol{A}$ for some $0 < \beta < 1$. Therefore, the relative perturbation $\varepsilon_{\boldsymbol{A}}^{(K)}$ in (3) no longer represents a worst-case deviation (i.e., the ratio $\frac{\|\boldsymbol{E}\|_2^{(K)}}{\|\boldsymbol{A}\|_2^{(K)}} = \beta =: \varepsilon_{\boldsymbol{A}}^{(K)}$).

The full details of this proof can be found in [6] □

3.2 Bounding the perturbed observation

Before proceeding, we need some sense of the size of the total perturbation incurred by $\boldsymbol{E}$ and $\boldsymbol{e}$. We don't know *a priori* the exact values of $\boldsymbol{E}$, $\boldsymbol{x}$, or $\boldsymbol{e}$. But we can find an upper bound in terms of the relative perturbations in (3). The main goal in the following lemma is to remove the total perturbation's dependence on the input $\boldsymbol{x}$.

Lemma 2 (Total perturbation bound)**.** *Set* $\varepsilon'_{\boldsymbol{A},K,\boldsymbol{b}} := \left(c\varepsilon_{\boldsymbol{A}}^{(K)} + \varepsilon_{\boldsymbol{b}}\right)\|\boldsymbol{b}\|_2$*, where* $c = \frac{\sqrt{1+\delta_K}}{\sqrt{1-\delta_K}}$*, and* $\varepsilon_{\boldsymbol{A}}^{(K)}$ *and* $\varepsilon_{\boldsymbol{b}}$ *are defined in (3). Then the total perturbation obeys*

$$\|\boldsymbol{E}\boldsymbol{x}\|_2 + \|\boldsymbol{e}\|_2 \le \varepsilon'_{\boldsymbol{A},K,\boldsymbol{b}} \qquad (17)$$

for all K*-sparse* $\boldsymbol{x}$*.*

Proof. From (1), (5) and (3) we have

$$\begin{aligned} \|\boldsymbol{E}\boldsymbol{x}\|_2 + \|\boldsymbol{e}\|_2 &= \left(\frac{\|\boldsymbol{E}\boldsymbol{x}\|_2}{\|\boldsymbol{A}\boldsymbol{x}\|_2} + \frac{\|\boldsymbol{e}\|_2}{\|\boldsymbol{b}\|_2}\right)\|\boldsymbol{b}\|_2 \\ &\le \left(\frac{\|\boldsymbol{E}\|_2^{(K)}\|\boldsymbol{x}\|_2}{\sqrt{1-\delta_K}\,\|\boldsymbol{x}\|_2} + \frac{\|\boldsymbol{e}\|_2}{\|\boldsymbol{b}\|_2}\right)\|\boldsymbol{b}\|_2 \\ &\le \left(c\varepsilon_{\boldsymbol{A}}^{(K)} + \varepsilon_{\boldsymbol{b}}\right)\|\boldsymbol{b}\|_2 \end{aligned}$$

for all $\boldsymbol{x}$ which are K-sparse. □

Note that the results in this paper can easily be expressed in terms of the perturbed observation by replacing

$$\|\boldsymbol{b}\|_2 \le \frac{\|\hat{\boldsymbol{b}}\|_2}{1-\varepsilon_{\boldsymbol{b}}}.$$

This can be useful in practice since one normally only has access to $\hat{\boldsymbol{b}}$.

3.3 Proof Sketch of Theorem 2

We duplicate the techniques used in Candès' proof of Theorem 1.2 in [1], but with decoding matrix $\boldsymbol{A}$ replaced by $\hat{\boldsymbol{A}}$. Set the BP minimizer in (7) as $\boldsymbol{z}^\star = \boldsymbol{x}+\boldsymbol{h}$. Here, $\boldsymbol{h}$ is the perturbation from the true solution $\boldsymbol{x}$ induced by $\boldsymbol{E}$ and $\boldsymbol{e}$. Instead of Candès' (9), we determine that the image of $\boldsymbol{h}$ under $\hat{\boldsymbol{A}}$ is bounded by

$$\begin{aligned} \|\hat{\boldsymbol{A}}\boldsymbol{h}\|_2 &\le \|\hat{\boldsymbol{A}}\boldsymbol{z}^\star - \hat{\boldsymbol{b}}\|_2 + \|\hat{\boldsymbol{A}}\boldsymbol{x} - \hat{\boldsymbol{b}}\|_2 \\ &\le 2\,\varepsilon'_{\boldsymbol{A},K,\boldsymbol{b}} \end{aligned}$$

which follows from the BP constraint in (7) as well as $\boldsymbol{x}$ being a feasible solution (i.e., it satisfies Lemma 2). The rest of this proof can be found in [6] □

3.4 Proof of Lemma 1

Assume the hypothesis of Theorem 2. It is easy to show that this implies

$$\|\boldsymbol{E}\|_2^{(2K)} \; < \; \sqrt[4]{2} \; - \; \sqrt{1 + \delta_{2K}}.$$

Simple algebraic manipulation then confirms that

$$\sqrt[4]{2} \; - \; \sqrt{1 + \delta_{2K}} \; < \; \sqrt{1 - \delta_{2K}} \; = \; \sigma_{\min}^{(2K)}(\boldsymbol{A}).$$

Therefore, (13) holds with $k = 2K$. Further, for any $k \leq 2K$ we have $\sigma_{\max}^{(k)}(\boldsymbol{E}) \leq \sigma_{\max}^{(2K)}(\boldsymbol{E})$ and $\sigma_{\min}^{(2K)}(\boldsymbol{A}) \leq \sigma_{\min}^{(k)}(\boldsymbol{A})$, which proves the lemma. □

4. Classical ℓ_2 Perturbation Analysis

Let the subset $T \subseteq \{1, \ldots, n\}$ have cardinality $|T| = K$, and note the following T-*restrictions*: $\boldsymbol{A}_T \in \mathbb{C}^{m \times K}$ denotes the submatrix consisting of the columns of $\boldsymbol{A}$ indexed by the elements of T, and similarly for $\boldsymbol{x}_T \in \mathbb{C}^K$.

Suppose the "oracle" case where we already know the support T of K-sparse $\boldsymbol{x}$. By assumption, we are only interested in the case where $K \leq m$ in which $\boldsymbol{A}_T$ has full rank. Given the completely perturbed observation of (2), the least squares problem consists of solving:

$$\boldsymbol{z}_T^{\#} \; = \; \operatorname*{argmin}_{\hat{\boldsymbol{z}}_T} \|\hat{\boldsymbol{A}}_T \hat{\boldsymbol{z}}_T - \hat{\boldsymbol{b}}\|_2.$$

Since we know the support T, it is trivial to extend $\boldsymbol{z}_T^{\#}$ to $\boldsymbol{z}^{\#} \in \mathbb{C}^n$ by zero-padding on the complement of T. Our goal is to see how the perturbations $\boldsymbol{E}$ and $\boldsymbol{e}$ affect $\boldsymbol{z}^{\#}$.

More discussion on the oracle least squares analysis can be found in [6]. In the end, we find using the same $\varepsilon'_{\boldsymbol{A},K,\boldsymbol{b}}$ in (10) that its stability is

$$\|\boldsymbol{z}^{\#} - \boldsymbol{x}\|_2 \; \leq \; C_{\mathrm{LS}} \, \varepsilon'_{\boldsymbol{A},K,\boldsymbol{b}} \qquad (18)$$

where $C_{\mathrm{LS}} := 1/\sqrt{1 - \delta_K}$.

4.1 Comparison of LS with BP

Now, we can compare the accuracy of the least squares solution in (18) with the accuracy of the BP solution found in (11). In both cases the error bound is of the form

$$C \varepsilon'_{\boldsymbol{A},K,\boldsymbol{b}}.$$

A detailed numerical comparison of C_{LS} with C_{BP} is not entirely valid, nor illuminating. This is due to the fact that we assumed the oracle setup in the the least squares analysis, which is the best that one could hope for. In this sense, the least squares solution we examined here can be considered a "best, worst-case" scenario. In contrast, the BP solution really should be thought of as a "worst, of the worst-case" scenarios.

The important thing to glean is that the accuracy of the BP solution, like the least squares solution, is on the order of the noise level $\varepsilon'_{\boldsymbol{A},K,\boldsymbol{b}}$ in the perturbed observation. This is an important finding since, in general, no other recovery algorithm can do better than the oracle least squares solution. These results are analogous to the comparison by Candès, Romberg and Tao in [2], although they only consider the case of additive noise $\boldsymbol{e}$.

5. Conclusion

We introduced a general perturbed model for CS, and found the conditions under which BP could stably recover the original data. This completely perturbed model extends previous work by including a multiplicative noise term in addition to the usual additive noise term. We only considered K-sparse signals, however these results can be extended to also include compressible signals (see [6]).

Simple numerical examples were given which demonstrated how the multiplicative noise reduced the accuracy of the recovered BP solution. In terms of the spectrum of the perturbed matrix $\hat{\boldsymbol{A}}$, we showed that the penalty on $\hat{\delta}_K$ was a graceful, linear function of the relative perturbation $\varepsilon_{\boldsymbol{A}}^{(K)}$. Numerical simulations were performed with $\varepsilon_{\boldsymbol{b}} = 0$ and appear to confirm the conclusion of Theorem 2, that the BP solution scales linearly with $\varepsilon_{\boldsymbol{A}}^{(K)}$.

We also found that the rank of $\hat{\boldsymbol{A}}$ did not exceed the rank of $\boldsymbol{A}$ under the assumed conditions. This permitted an analysis of the oracle least squares solution which showed that its accuracy, like the BP solution, was limited by the total noise in the observation.

Acknowledgment

This work was partially supported by NSF Grant No. DMS-0811169 and NSF VIGRE Grant No. DMS-0636297.

References:

[1] E. J. Candès. The restricted isometry property and its implications for compressed sensing. *Académie des Sciences*, I(346):589–592, 2008.

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

[3] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM Journal Sci. Comput.*, 20(1):33–61, 1999.

[4] D. L. Donoho, M. Elad, and V. Temlyakov. Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Trans. Inf. Theory*, 52(1):6–18, Jan. 2006.

[5] M. Grant, S. Boyd, and Y. Ye. cvx: Matlab software for disciplined convex programming. http://www.stanford.edu/~boyd/cvx/.

[6] M. A. Herman and T. Strohmer. General Deviants: An analysis of perturbations in compressed sensing. http://www.math.ucdavis.edu/~mattyh/publications.html.

[7] M. A. Herman and T. Strohmer. High-resolution radar via compressed sensing. *To appear in IEEE Trans. Signal Processing*, Jun. 2009.

[8] J. A. Tropp. Just relax: Convex programming methods for identifying sparse signals in noise. *IEEE Trans. Inf. Theory*, 51(3):1030–1051, Mar. 2006.