# General Perturbations in Compressed Sensing

Matthew A. Herman, Thomas Strohmer

# General Perturbations in Compressed Sensing

Matthew A. Herman and Thomas Strohmer
Department of Mathematics, University of California, Davis, CA 95616–8633, USA
e-mail: {mattyh,strohmer}@math.ucdavis.edu

***Abstract*—We analyze the Basis Pursuit recovery of signals when observing $K$-sparse data with general perturbations (i.e., *additive*, as well as *multiplicative* noise). This completely perturbed model 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.**

## I. 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:

$$\boldsymbol{b} \;=\; \boldsymbol{A}\boldsymbol{x}. \tag{1}$$

Here $\boldsymbol{b} \in \mathbb{C}^m$ is the observation vector, $\boldsymbol{A} \in \mathbb{C}^{m\times n}$ $(m \le n)$ 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]–[4], creating a *partially perturbed model*:

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

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

As far as we can tell, no research has been done yet on perturbations $\boldsymbol{E}$ to the matrix $\boldsymbol{A}$.[1] 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 [6] 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);[2] there will be an extra noise term $\boldsymbol{E}\boldsymbol{x}$).

### A. Assumptions and Notation

Without loss of generality, assume the original data $\boldsymbol{x}$ to be a $K$-sparse vector for some fixed $K$. We hope to extend these results in the future to include compressible signals (i.e., those whose ordered coefficients decay according to a power law). Denote $\sigma_{\max}^{(K)}(\boldsymbol{Y})$, $\|\boldsymbol{Y}\|_2^{(K)}$, and $\mathrm{rank}^{(K)}(\boldsymbol{Y})$ 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

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

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

## II. CS $\ell_1$ Perturbation Analysis

### A. 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 [7]:

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

for some $\varepsilon' \ge 0$.[3]

The *restricted isometry property* (RIP) [8] for any matrix $\boldsymbol{A} \in \mathbb{C}^{m\times n}$ defines, for each integer $K = 1, \ldots, n$, the *restricted isometry constant* (RIC) $\delta_K$, which is the smallest nonnegative number such that

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

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 Thm. 1.2 of [1]) that the solution to (4) obeys

$$\|\boldsymbol{z}^\star - \boldsymbol{x}\|_2 \;\le\; C_{\text{BP}}\,\varepsilon' \tag{6}$$

for some constant $C_{\text{BP}}$.

[1] A related problem is considered in [5] for greedy algorithms rather than $\ell_1$-minimization, and in a multichannel rather than a single channel setting; it mentions using different matrices on the encoding and decoding sides, but its analysis is not from an error or perturbation point of view.

[2] 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 IV.

[3] Throughout this paper *absolute* errors are denoted with a prime. In contrast, *relative* perturbations, such as in (3), are not primed.

## B. Incorporating nontrivial perturbation $\boldsymbol{E}$

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

$$\boldsymbol{z}^{\star} = \underset{\hat{\boldsymbol{z}}}{\operatorname{argmin}} \|\hat{\boldsymbol{z}}\|_1 \quad \text{s.t.} \quad \|\hat{\boldsymbol{A}}\hat{\boldsymbol{z}} - \hat{\boldsymbol{b}}\|_2 \leq \varepsilon'_{\boldsymbol{A},K,\boldsymbol{b}} \tag{7}$$

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

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

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

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

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

*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 the RIC for matrix $\boldsymbol{A}$ satisfies*

$$\delta_{2K} < \frac{\sqrt{2}}{\Big(1+\varepsilon_{\boldsymbol{A}}^{(2K)}\Big)^2} - 1.$$

*Set*

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

*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*

$$\|\boldsymbol{z}^{\star} - \boldsymbol{x}\|_2 \leq C_{BP}\,\varepsilon'_{\boldsymbol{A},K,\boldsymbol{b}}, \tag{11}$$

*where*

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

*Remark 2.* Theorem 2 generalizes of Candès' results in [1] for $K$-sparse $\boldsymbol{x}$. Indeed, if matrix $\boldsymbol{A}$ is unperturbed, then $\boldsymbol{E} = \mathbf{0}$ 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 (18)) collapses to $\|\boldsymbol{e}\|_2 \leq \varepsilon'_{\boldsymbol{b}} := \varepsilon_{\boldsymbol{b}}\|\boldsymbol{b}\|_2$; both of these are identical to Candès' assumptions in (6). Finally, the constant $C_{\text{BP}}$ 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 $k \leq 2K$*

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

*and therefore*

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

We apply this fact in the least squares analysis of Section IV.

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 $C_{\text{BP}} = 9.057$. Note from (12), we see that if there had been no perturbation, then $C_{\text{BP}} = 5.530$.

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 $C_{\text{BP}} = 9.643$. Here, if $\boldsymbol{A}$ was unperturbed, then we would have had $C_{\text{BP}} = 8.473$.

These numerical examples show how the stability constant $C_{\text{BP}}$ 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.

# III. Proofs

## A. Proof of Theorem 1

Temporarily define $l_K$ and $u_K$ as the *smallest nonnegative numbers* such that

$$(1-l_K)\|\boldsymbol{x}\|_2^2 \leq \|\hat{\boldsymbol{A}}\boldsymbol{x}\|_2^2 \leq (1+u_K)\|\boldsymbol{x}\|_2^2 \tag{14}$$

holds for any $K$-sparse vector $\boldsymbol{x}$. From the triangle inequality, (5) and (3) we have

$$\begin{aligned}
\|\hat{\boldsymbol{A}}\boldsymbol{x}\|_2^2 &\leq \big(\|\boldsymbol{A}\boldsymbol{x}\|_2 + \|\boldsymbol{E}\boldsymbol{x}\|_2\big)^2 && (15)\\
&\leq \Big(\sqrt{1+\delta_K} + \|\boldsymbol{E}\|_2^{(K)}\Big)^2\|\boldsymbol{x}\|_2^2 && (16)\\
&\leq (1+\delta_K)\Big(1+\varepsilon_{\boldsymbol{A}}^{(K)}\Big)^2\|\boldsymbol{x}\|_2^2. && (17)
\end{aligned}$$

In comparing the RHS of (14) and (17), it must be that

$$(1+u_K) \leq (1+\delta_K)\Big(1+\varepsilon_{\boldsymbol{A}}^{(K)}\Big)^2$$

as demanded by the definition of the $u_K$. Moreover, this inequality is sharp for the following reasons:

- Equality occurs in (15) whenever $\boldsymbol{E}$ is a multiple of $\boldsymbol{A}$.[4]
- Equality occurs in (16) 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 (17) 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 $\frac{\|\boldsymbol{E}\|_2^{(K)}}{\|\boldsymbol{A}\|_2^{(K)}} = \beta =: \varepsilon_{\boldsymbol{A}}^{(K)}$).

Since the triangle inequality constitutes a *least-upper bound*, and since we attain this bound, then

$$u_K := (1+\delta_K)\left(1+\varepsilon_{\boldsymbol{A}}^{(K)}\right)^2 - 1$$

satisfies the definition of $u_K$. Now the LHS of (14) is obtained in much the same way using the "reverse" triangle inequality with the same arguments. Thus

$$l_K := 1 - (1-\delta_K)\left(1-\varepsilon_{\boldsymbol{A}}^{(K)}\right)^2.$$

Since $(1-u_K) \leq (1-l_K)$ and $(1+l_K) \leq (1+u_K)$, we choose

$$\hat{\delta}_K := u_K$$

as the smallest nonnegative constant which makes (9) true. ■

### B. 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*

$$\|\boldsymbol{E}\boldsymbol{x}\|_2 + \|\boldsymbol{e}\|_2 \leq \varepsilon'_{\boldsymbol{A},K,\boldsymbol{b}} \tag{18}$$

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

*Proof:* From (1), (5) and (3) we have

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

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

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

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

[4] See the appendix for more discussion on the different forms of perturbation $\boldsymbol{E}$ which we are likely to encounter.

### C. Proof of Theorem 2

*Step 1.* We duplicate the techniques used in Candès' proof of Thm. 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

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

which follows from the triangle inequality, and the BP constraint in (7) since $\boldsymbol{x}$ is a feasible solution (i.e., it satisfies Lemma 2).

Since the other steps in the proof are essentially the same, we end up with constants $\hat{\alpha}$ and $\hat{\rho}$ in Candès' (14) (instead of $\alpha$ and $\rho$) where

$$\hat{\alpha} := \frac{2\sqrt{1+\hat{\delta}_{2K}}}{1-\hat{\delta}_{2K}}, \qquad \hat{\rho} := \frac{\sqrt{2}\,\hat{\delta}_{2K}}{1-\hat{\delta}_{2K}}. \tag{19}$$

The main difference at this point is that we assume that we are solely dealing with $K$-sparse signals $\boldsymbol{x}$. Therefore, the terms in Theorem 1.2 and its proof which involve $\|\boldsymbol{x} - \boldsymbol{x}_K\|_1$ and $e_0$ are identically zero. The final line of the proof concludes that

$$\|\boldsymbol{h}\|_2 \leq \frac{2\hat{\alpha}\varepsilon'_{\boldsymbol{A},K,\boldsymbol{b}}}{1-\hat{\rho}}. \tag{20}$$

The denominator demands that we impose the condition that $0 < 1 - \hat{\rho}$, or equivalently

$$\hat{\delta}_{2K} < \sqrt{2} - 1. \tag{21}$$

The constant $C_{\mathrm{BP}}$ is obtained by substituting $\hat{\delta}_{2K}$ from (8), and $\hat{\alpha}$ and $\hat{\rho}$ from (19) into (20).

*Step 2.* We still need to show that the hypothesis of Theorem 2 implies (21). This is easily verified by substituting the assumption of $\delta_{2K} < \sqrt{2}\left(1+\varepsilon_{\boldsymbol{A}}^{(2K)}\right)^{-2} - 1$ into (8) (with $K \to 2K$) and the proof is complete. ■

### D. 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. ■

## IV. Classical $\ell_2$ Perturbation Analysis

Let the subset $T \subseteq \{1,\dots,n\}$ 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 $K \le m$ 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 $\boldsymbol{z}^{\#}$. Using Golub and Van Loan's model (Thm 5.3.1, [9]) as a guide, assume

$$\max\left\{\frac{\|\boldsymbol{E}\|_2^{(K)}}{\|\boldsymbol{A}\|_2^{(K)}}, \frac{\|\boldsymbol{e}\|_2}{\|\boldsymbol{b}\|_2}\right\} \;<\; \frac{\sigma_{\min}^{(K)}(\boldsymbol{A})}{\sigma_{\max}^{(K)}(\boldsymbol{A})}. \tag{22}$$

Notice that the reciprocal of the RHS of (22) is represented in Theorem 2 and Lemma 2 as the constant

$$c \;=\; \frac{\sqrt{1+\delta_K}}{\sqrt{1-\delta_K}} \;=\; \frac{\sigma_{\max}^{(K)}(\boldsymbol{A})}{\sigma_{\min}^{(K)}(\boldsymbol{A})}.$$

This ratio can be thought of as the worst condition number over all $K$-column submatrices of $\boldsymbol{A}$. Actually, for very small $\delta_K$ we have $c \approx 1$, which implies that every $K$-column submatrix forms an approximately orthonormal set.

*Remark 3.* Assumption (22) is fairly easy to satisfy. In fact, the hypothesis of Theorem 2 immediately implies that $\|\boldsymbol{E}\|_2^{(K)}/\|\boldsymbol{A}\|_2^{(K)} < \sqrt{1-\delta_K}/\sqrt{1+\delta_K}$ for all $\varepsilon_{\boldsymbol{A}}^{(K)} \in [0,1]$ (simply set $k = K$ in (13) of Lemma 1). Further, the reasonable condition of $\varepsilon_{\boldsymbol{b}} \le \left(\sqrt{2}\left(1+\varepsilon_{\boldsymbol{A}}^{(2K)}\right)^2 - 1\right)^{1/2}$ is sufficient to ensure $\varepsilon_{\boldsymbol{b}} < \sqrt{1-\delta_{2K}}/\sqrt{1+\delta_{2K}}$ so that assumption (22) holds. Note that this assumption has no bearing on CS recovery, nor is it a constraint due to BP. It is simply made to enable an analysis of the least squares solution which we use as a best-case comparison below.

Following the steps in [9] with the appropriate modifications for our situation we obtain

$$\|\boldsymbol{z}^{\#} - \boldsymbol{x}\|_2 \;\le\; \|\boldsymbol{A}_T^{\dagger}\|_2\left(\frac{\|\boldsymbol{E}\boldsymbol{x}\|_2}{\|\boldsymbol{A}\boldsymbol{x}\|_2} + \frac{\|\boldsymbol{e}\|_2}{\|\boldsymbol{b}\|_2}\right)\|\boldsymbol{b}\|_2,$$

where $\boldsymbol{A}_T^{\dagger} = (\boldsymbol{A}_T^*\boldsymbol{A}_T)^{-1}\boldsymbol{A}_T^*$ is the left inverse of $\boldsymbol{A}_T$. Its spectral norm is

$$\|\boldsymbol{A}_T^{\dagger}\|_2 \;=\; 1/\sigma_K(\boldsymbol{A}_T),$$

where $\sigma_K(\boldsymbol{A}_T)$ is the smallest singular value of $\boldsymbol{A}_T$. Also, by definition of $\sigma_{\min}^{(K)}(\boldsymbol{A})$ it must be for any $T$-restriction of $\boldsymbol{A}$ with $|T| = K$ that

$$\sqrt{1-\delta_K} \;\le\; \sigma_K(\boldsymbol{A}_T),$$

which implies

$$\|\boldsymbol{A}_T^{\dagger}\|_2 \;\le\; 1\big/\sqrt{1-\delta_K}.$$

Finally, applying the same steps as in the proof of Lemma 2, we have a bound on the perturbation error of the least squares solution in terms of the same $\varepsilon'_{\boldsymbol{A},K,\boldsymbol{b}}$ in (10):

$$\|\boldsymbol{z}^{\#} - \boldsymbol{x}\|_2 \;\le\; C_{\text{LS}}\,\varepsilon'_{\boldsymbol{A},K,\boldsymbol{b}}, \tag{23}$$

with $C_{\text{LS}} := 1/\sqrt{1-\delta_K}$.

### A. Comparison of LS with BP

Now, we can compare the accuracy of the least squares solution in (23) 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 $C_{\text{LS}}$ with $C_{\text{BP}}$ 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 $\boldsymbol{e}$.

## V. 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. 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 $\varepsilon_{\boldsymbol{A}}^{(K)}$. 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.

## APPENDIX

### DIFFERENT CASES OF PERTURBATION $\boldsymbol{E}$

There are essentially two classes of perturbations $\boldsymbol{E}$ which we care most about: *random* and *structured*. The nature of these perturbation matrices will have a significant effect on the value of $\|\boldsymbol{E}\|_2^{(K)}$, which is used in determining $\varepsilon_{\boldsymbol{A}}^{(K)}$ in (3). In fact, explicit knowledge of $\boldsymbol{E}$ can significantly improve the worst-case assumptions presented throughout this paper. However, if there is no extra knowledge on the nature of $\boldsymbol{E}$, then we can only rely on the "worst case" deviation $\|\boldsymbol{E}\|_2^{(K)} \leq \|\boldsymbol{E}\|_2$.

### A. Random Perturbations

Random matrices, such as Gaussian, Bernoulli, and partial Fourier matrices, are often amenable to analysis with the RIP (see, e.g., [10], [11], [12]). For instance, suppose that $\boldsymbol{E}$ is simply a scaled version of a random matrix $\boldsymbol{R}$ so that $\boldsymbol{E} = \beta\boldsymbol{R}$ with $0 < \beta \ll 1$. Denote $\delta_K^{\boldsymbol{R}}$ as the RIC associated with the matrix $\boldsymbol{R}$. Then for all $K$-sparse $\boldsymbol{x}$ the RIP for matrix $\boldsymbol{E}$ obeys

$$\beta^2(1-\delta_K^{\boldsymbol{R}})\|\boldsymbol{x}\|_2^2 \leq \|\boldsymbol{E}\boldsymbol{x}\|_2^2 \leq \beta^2(1+\delta_K^{\boldsymbol{R}})\|\boldsymbol{x}\|_2^2,$$

which immediately gives us

$$\|\boldsymbol{E}\|_2^{(K)} = \beta\sqrt{1+\delta_K^{\boldsymbol{R}}},$$

and thus

$$\frac{\|\boldsymbol{E}\|_2^{(K)}}{\|\boldsymbol{A}\|_2^{(K)}} = \beta\frac{\sqrt{1+\delta_K^{\boldsymbol{R}}}}{\sqrt{1+\delta_K}} =: \varepsilon_{\boldsymbol{A}}^{(K)}.$$

### B. Structured Perturbations

When encountering a structured perturbation (e.g., a Toeplitz matrix, or a banded matrix), its nature can be exploited to find a bound $\|\boldsymbol{E}\|_2^{(K)} \leq C$. For example, suppose that $\boldsymbol{E}$ is a partial circulant matrix obtained by selecting $m$ rows uniformly at random from an $n \times n$ circulant matrix. An error in the modeling of a communication channel could be represented by such a partial circulant matrix.

A complete circulant matrix has the property that each row is simply a right-shifted version of the row above it. Therefore, knowledge of any row gives information about the entries of all of the rows. This is also true for a partial circulant matrix. Thus, with this information we may be able to find a reasonable upper bound on $\|\boldsymbol{E}\|_2^{(K)}$.

## 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*, vol. I, no. 346, pp. 589–592, 2008.

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

[3] D. L. Donoho, M. Elad, and V. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 6–18, Jan. 2006.

[4] J. A. Tropp, "Just relax: Convex programming methods for identifying sparse signals in noise," *IEEE Trans. Inf. Theory*, vol. 51, no. 3, pp. 1030–1051, Mar. 2006.

[5] 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," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5-6, pp. 1069–5869, Dec. 2008.

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

[7] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal Sci. Comput.*, vol. 20, no. 1, pp. 33–61, 1999.

[8] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[9] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. Baltimore: Johns Hopkins University Press, 1996.

[10] E. J. Candès and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies," *IEEE Trans. Inf. Theory*, vol. 52, pp. 5406–5425, 2006.

[11] S. Mendelson, A. Pajor, and N. Tomczak-Jaegermann, "Uniform uncertainty principle for Bernoulli and subgaussian ensembles," 2009, to appear, *Constr. Approx.*

[12] M. Rudelson and R. Vershynin, "On sparse reconstruction from Fourier and Gaussian measurements," *Communications on Pure and Applied Mathematics*, vol. 61, pp. 1025–1045, 2008.