Performance Estimation of Compressed Sensing with Oracle Information

Ruiyang Song, Laming Chen, and Yuantao Gu*

January 15, 2015

Abstract

This article discusses the performance of the oracle receiver in terms of the normalized mean square error in the sparse signal recovery process. The measurement is conducted in completely perturbed scenarios where the system is disturbed simultaneously by a perturbation matrix exerted on the sensing matrix, a noise vector on the result of measurement and the input noise added directly on the signal. In a bid to achieve concrete results, the entries of the sensing matrix are specified as Bernoulli random variables. The article introduces and proves the lower and upper bounds of the mean square error of the reconstructed signal. Those theoretical bounds hold in high probability for high dimensional signals. Numerical results verified the conclusions, illustrating both the lower and upper bounds as close and meaningful estimations of the mean square error in the Bernoulli case. The result is then compared with previous works in literature on Gaussian sensing matrices. An estimation of the average recovery error is derived as a generalization of the conclusion in [13] for the Gaussian case.

Keywords: Sparse Reconstruction, Compressed Sensing, Oracle Receiver, ± 1 Bernoulli random matrices

1 Introduction

The recovery of a signal $\mathbf{x} \in \mathbb{R}^n$ from an underdetermined linear system

$$\mathbf{y} = \mathbf{A}\mathbf{x}$$

has been a fundamental problem drawing the attention of numerous scholars in the realm of signal processing.([15]) Though generally infeasible, compressed sensing ([5], [10], and [17]) has been introduced to recover sparse or compressible vectors, indicating that far fewer number of measurements $k \ll n$ should be sufficient to recover the signal with high accuracy.

*The authors are with the Department of Electronic Engineering, Tsinghua University, Beijing 100084, CHINA. The corresponding author of this paper is Yuantao Gu (gyt@tsinghua.edu.cn).

Unfortunately, measurements will inevitably suffer from the perturbation of various noises which diminish the accuracy and reliability of the reconstructed signal $\hat{\mathbf{x}}$. The noisy model $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}$ has been studied from several perspectives. Candes et al. [7] introduced a convex programming problem and proved that the solution to the optimization problem $\hat{\mathbf{x}}$ is a good recovery since $\|\mathbf{x} - \hat{\mathbf{x}}\|$ is upper bounded by a constant C_S times the upper bound of the ℓ_2 norm of $\mathbf{n}$. The constant, in particular, depends on δ_{4S}, where the denotation δ_S stands for the *S-restricted isometry constant*, where S is the sparsity of signal. ([9]). This setup is regarded as a coding process in [9] and a linear programming method is introduced to decode the signal from the noisy outcome. A Dantzig selector on the noise model has been introduced ([8]) and analyzed ([3]). This setup is also discussed in [4] where the author derived the Cramer Rao bound (CRB).

Popular models of the sensing matrix include Gaussian and Bernoulli ensembles. The Restricted Isometry Propoerties (RIP) of Gaussian and Bernoulli random matrices are studied in [21]. The authors of [6] analyzed the Gaussian case in depth, illustrating the probability of exact reconstruction and generalized the result to sub-Gaussian case. [25] also considered the noisy setup and particularly mentioned the Gaussian and Bernoulli ensembles. Concrete discussions of the Bernoulli case can also be found in [27].

Wishart distribution has been a widely applied model in the field of statistical signal processing. Properties of Wishart random matrices such as its inverse and pseudo-inverse have been well established in literature. ([19], [22]) Basing on the mathematical results, Coluccia et al. ([13]) introduced and analyzed the oracle receiver, a type of ideally optimal receiver possessing the exact information of the support set Ω. By applying theorems and conclusions concerning Wishart distribution, the author illustrated the performance estimation also using the $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}$ model, and the entries of $\mathbf{A}$ were assigned to be i.i.d. random varaibles drawn from a zero-mean Gaussian distribution.

More general noise models have also been considered. The input noise which is directly added to the signal before measurement has been discussed in [1] and [14]. The measurement in this case can be therefore described as $\mathbf{y} = \mathbf{A}(\mathbf{x} + \mathbf{e}) + \mathbf{n}$, which involves the "noise folding" effect. Another interesting model introduces the perturbation of the sensing matrix that can be elaborated as $\mathbf{y} = (\mathbf{A} + \mathbf{E})\mathbf{x} + \mathbf{n}$, where $\mathbf{E}$ is a random matrix added to the measurement matrix $\mathbf{A}$ and $\mathbf{n}$ is still the additive noise.([16]) This matrix perturbation model is also discussed in [23] where the sensing matrix is considered in general and two lower bounds of the mse are discussed, including the constrained Cramer-Rao bound (CCRB) and the HammersleyChapman-Robbins bound (HCRB). In [11], the authors analyzed the recovery performance of greedy pursuits with replacement. During the process of the recovery algorithm, the update of the support of the signal is renewed after each step of iteration.

In this article, a more general noise model with the juxtaposition of additive noise $\mathbf{n}$, matrix perturbation $\mathbf{E}$, and input noise $\mathbf{e}$ will be discussed. With the system set as

$$\mathbf{y} = (\mathbf{A} + \mathbf{E})(\mathbf{x} + \mathbf{e}) + \mathbf{n}.$$

The performance of this particular receiver can be estimated by the normalized mean square error (mse) of the reconstructed signal

$$\text{mse}_0 = \frac{\mathbb{E}[\|\mathbf{x} - \hat{\mathbf{x}}\|_{\mathbf{2}}^{\mathbf{2}}]}{\mathbb{E}[\|\mathbf{x}\|_2^2]}.$$

The analysis of the relationship between the average recovery error and the noises from three different sources will be more involved comparing to previous models.

Random matrices are of great interest in the fields of signal processing as well as numerical mathematics. Properties of typical models such as the Gaussian ensemble have been extensively studied. ([20]) Another important ensemble is the ± 1 random matrix. The singularity probability of a random ± 1 matrix is discussed in depth in [24] and [18]. In [26], a theorem presents a lower bound of the least singular value and an upper bound of the largest singular value of a tall matrix with independent sub-Gaussian isotropic columns.

This paper develops an approach basing on the applications of theories in random matrices, to achieve more concrete and explicit results with constants and parameters easier to comprehend, estimate, and handle. Another primary difference between this article and prior works is the model of sensing matrix. The article primarily dedicates in studying the performance of Bernoulli random matrices comprising of i.i.d. entries. This paper also presents a generalized form of the result on Gaussian sensing matrix in [13].

We apply the following notations in this article. Denote the support set of vector $\mathbf{v}$ as $\text{supp}(\mathbf{v})$.

For any matrix $\mathbf{B}$ suppose S is a finite subset of positive integers. Denote $\mathbf{B}_S$ the submatrix of $\mathbf{B}$ generated by picking the columns indexed in S. Similarly, for any vector $\mathbf{v}$, denote $\mathbf{v}_S$ the vector derived by taking the entries of $\mathbf{v}$ indexed in S. For simplicity, denote $\mathbf{B}_{\{j\}} \equiv \mathbf{B}_j$, $\mathbf{v}_{\{j\}} \equiv \mathbf{v}_\mathbf{j}$. Denote $s_{max}(\mathbf{B})$ and $s_{min}(\mathbf{B})$ the largest and smallest singular values of $\mathbf{B}$ (or simply s_{max} and s_{min} if the matrix is indicated in contexts).

For a square matrix $\mathbf{D}$, denote $\lambda(\mathbf{D})$ the set of eigenvalues of $\mathbf{D}$. The following notations are also introduced in order to simplify equations. For a square matrix $\mathbf{D} = \{\mathbf{D}_{ij}\}_{i,j=1}^n \in \mathbb{R}^{n\times n}$ which is indicated in contexts, let

$$\mu_1 = \mathbf{tr}(\mathbf{D}),$$
$$\mu_2 = \sum_{i,j=1}^{n} \mathbf{D}_{ij}^2 = \|\mathbf{D}\|_F^2.$$

For $x, y \in \mathbb{R}^n$, denote $\langle x, y\rangle = x^T y$ the inner product of the two vectors. The notation $\|\cdot\|_2^2$ represents the ℓ_2 norm and $\|\cdot\|_F$ the Frobenius norm.

2 Problem Modeling

The sparse signal to be reconstructed from measurement $\mathbf{x}$ can be decomposed as

$$\mathbf{x} = \mathbf{\Psi}\boldsymbol{\theta},\ \mathbf{x} \in \mathbb{R}^{\mathbf{N}}$$

and the input noise

$$\mathbf{e} = \mathbf{\Psi}\boldsymbol{\varepsilon}.$$

$\mathbf{\Psi} \in \mathbb{R}^{N\times N}$ represents an orthonormal matrix. $\boldsymbol{\theta}$ is a sparse vector:

$$\mathrm{supp}(\boldsymbol{\theta}) = \Omega,\ |\Omega| = K \ll N.$$

We model the input noise $\mathbf{e} \sim \mathcal{N}(0, \sigma_e^2\mathbf{I}_N)$. The general noise model can be formulated as

$$\mathbf{y} = (\mathbf{A} + \mathbf{E})(\mathbf{x} + \mathbf{e}) + \mathbf{n}.$$

$\mathbf{E} \in \mathbb{R}^{M\times N}$ is a perturbation matrix with elements which are i.i.d. Gaussian distributed random variables with zero-mean, and variances σ_E^2. $\mathbf{A} \in \mathbb{R}^{M\times N}(M < N)$ is the sensing matrix. $\mathbf{n}$ denotes the measurement noise. In this article, two models of the sensing matrix are considered. We analyze the case when $\mathbf{A}$ comprises of i.i.d. Bernoulli random variables with sample space $S = \{-\sigma_A, \sigma_A\}$, and probability parameter $p = 1/2$. The elements of measurement noise $\mathbf{n}$ are i.i.d. Gaussian distributed random variables with zero mean, variance σ_n^2. The nonzero components of $\boldsymbol{\theta}$ are modeled as centered random variables and the covariance matrix of $\boldsymbol{\theta}$ is $\mathbf{\Sigma}_{\boldsymbol{\theta}}$.

Consider the oracle estimator with exact knowledge of the support set Ω. The estimator can therefore reconstruct $\hat{\boldsymbol{\theta}}$ as

$$\hat{\boldsymbol{\theta}}_{\mathbf{\Omega}} = \mathbf{U}_{\mathbf{\Omega}}^{\dagger}\mathbf{y},$$
$$\hat{\boldsymbol{\theta}}_{\mathbf{\Omega}^c} = \mathbf{0}.$$

and therefore recover the signal $\mathbf{x}$. Given the effect of noise, the result of the recovery $\hat{\mathbf{x}}$ is inaccurate. The aim is to estimate the normalized average reconstruction error, i.e. mean square error

$$\mathrm{mse}_0 = \mathbb{E}(\|\mathbf{x} - \hat{\mathbf{x}}\|_2^2).$$

As the oracle receiver possesses the knowledge of the support set, only $\mathbf{e}_{\mathbf{\Omega}} \in \mathbb{R}^K$ will influence the signal recovery. Thus, for the oracle receiver, the input noise can be reduced to $\tilde{\mathbf{e}} = \mathbf{\Psi}\tilde{\boldsymbol{\varepsilon}}$, where

$$\tilde{\boldsymbol{\varepsilon}}_{\mathbf{\Omega}} = \boldsymbol{\varepsilon}_{\mathbf{\Omega}}$$
$$\tilde{\boldsymbol{\varepsilon}}_{\mathbf{\Omega}^c} = \mathbf{0}$$

We will simply represent $\boldsymbol{\varepsilon} = \tilde{\boldsymbol{\varepsilon}}$ in the following text for simplicity of notations. Also, we assume $\mathbf{\Psi} = \mathbf{I}$ for the Bernoulli case due to the same consideration.

3 Main Result

A good way of interpreting the variance σ_A^2 is to regard it as an indicator of the measurement power. It is important to observe that the noise model concerned can be reformulated as $\mathbf{y} = \mathbf{Ax} + \mathbf{z_e}$, where $\mathbf{z_e} = \mathbf{z} + \mathbf{Ae}$ and $\mathbf{z} = \mathbf{Ex} + \mathbf{Ee} + \mathbf{n}$. Note that we can reduce the influence of $\mathbf{z}$ if we are willing to increase the power of measurement, i.e. the norm of $\mathbf{A}$. However, the noise folding effect resulting from the input noise $\mathbf{e}$ can not be eliminated by increasing the power.

Denote

$$\mathbf{p_x} = \mathbb{E}(\mathbf{x^T x}),$$

which is the average intensity of signal. And denote

$$\mathbf{p_e} = \mathbb{E}(\tilde{\mathbf{e}}^{\mathbf{T}}\tilde{\mathbf{e}}) = K\sigma_e^2$$

the average power of input noise in the same bandwidth of the signal. Define the input SNR (ISNR) as in [14],

$$\text{ISNR} = \frac{\mathbf{p_x}}{\mathbf{p_e}}.$$

Also, we introduce MSNR which is the SNR concerning the measurement noise $\mathbf{n}$:

$$\text{MSNR} = \frac{M\sigma_A^2\mathbf{p_x}}{\text{tr}(\Sigma_n)}.$$

Besides, we define the average relative power of the measurement to the perturbation as

$$\eta = \frac{\sigma_A^2}{\sigma_E^2}.$$

In the following of this article, we apply several more notations. For a positive definite matrix $\mathbf{C}$, define

$$\mathbf{J_C}(\xi) = \begin{pmatrix} \|\mathbf{C}\|_\mathbf{F}^\mathbf{2} & \text{tr}(\mathbf{C}) \\ \xi^2 & \xi \end{pmatrix}.$$

and

$$\mathscr{U}_\alpha(\mathbf{C}) = \begin{pmatrix} \text{tr}(\mathbf{C}) & K \end{pmatrix} \mathbf{J}_\mathbf{C}^{-\mathbf{1}}(\alpha) \begin{pmatrix} K \\ 1 \end{pmatrix},$$

Also, denote

$$\mathcal{P}_\mathcal{S} = \frac{[1 + o(1)]K(K-1)}{2^M},$$

and

$$\mathcal{P}_\mathcal{I} = 1 - \mathcal{P}_\mathcal{S}.$$

We propose the following result including probabilistic lower and upper bounds of the normalized mean square error mse_0.

Proposition 1. *(Lower and Upper Bounds of MSE) When the entries of the sensing matrix* $\mathbf{A} \in \mathbb{R}^{M\times N}$ *are* $\pm\sigma_A = \pm\kappa_A/\sqrt{M}$ *centered Bernoulli random variables,* $\mathbf{E} \in \mathbb{R}^{M\times N}$ *is comprised of i.i.d. Gaussian random variables that follow* $\mathcal{N}(\mathbf{0}, \sigma_E^2)$*, the measurement noise* $\mathbf{n}$ *and the input noise* $\mathbf{e}$ *are white noises,* $n \sim \mathcal{N}(0, \sigma_n^2\mathbf{I})$*,* $\mathbf{e} \sim \mathcal{N}(0, \sigma_e^2\mathbf{I})$*. For some* $t > 0$ *and constants* $C, c > 0$ *that depend only on* σ_A*, let*

$$\theta_0 = (\frac{\sqrt{M} - C\sqrt{K} - t}{\sqrt{M}})^2$$

and

$$\mathscr{F}(M, K, \theta_0) = \gamma\sigma_{\mathbf{A}}^{\mathbf{2}}\mathscr{U}_{\alpha_{\mathbf{0}}}.$$

Suppose the signal $\mathbf{x}$ *(*$\|x\|_0 = K$*) is sparse and the support set* Ω *is provided by an oracle, then the columns of* $\mathbf{A}_{\mathbf{\Omega}}$ *are independent and the normalized mean square error of the recovery is lower bounded by*

$$\mathrm{mse}_0 \geq \frac{1}{\gamma\eta} + \frac{1}{\gamma}\frac{1}{\mathrm{MSNR}} + (1 + \frac{1}{\gamma\eta})\frac{1}{\mathrm{ISNR}}$$

with overwhelmingly high probability

$$\mathcal{P}_{\mathcal{L}} = 1 - \mathcal{P}_{\mathcal{S}}$$

and lowered bounded by

$$\mathrm{mse}_0 \leq \frac{\mathscr{F}}{\gamma\eta} + (1 + \frac{\mathscr{F}}{\gamma\eta})\frac{1}{\mathrm{ISNR}} + \frac{\mathscr{F}}{\gamma}\frac{1}{\mathrm{MSNR}}$$

with high probability

$$\mathcal{P}_{\mathcal{U}} = 1 - \frac{1}{\gamma^2} - 2\exp(-ct^2) - \mathcal{P}_{\mathcal{S}}.$$

Proof. The proof of Proposition 1 is postponed to Appendix 9.7. □

Remark 1. *Proposition 1 presents a lower bound of the normalized average recovery error of the optimal estimator that holds with overwhelmingly high probability, which is in turn a lower bound of the normalized MSE of any recovery strategy.*

Remark 2. *This proposition demonstrates a probablistic upper bound of the MSE in the Bernoulli case. By taking* $\gamma^2 = M^2/K^2$ *and parameter* t *sufficiently large, we can derive an upper bound that holds with high probability*

With Proposition 1, lower and upper bounds that hold with high probability are introduced for Bernoulli sensing matrices. By applying some applicable conditions, the expression of the lower and upper bounds will be more concise. Recall $M = \gamma K$, where $\gamma^2 \gg 1, K \gg 1$.

Furthermore, if $\gamma \gg 1$ holds, $\theta_0 \approx 1 - \frac{2C}{\sqrt{\gamma}}$, we have the following corollary:

Corollary 1. *When $\gamma \gg 1$, let $1/\hat{\gamma} = 1/\gamma(1+1/\gamma)$. The results in Proposition 1 can be approximated as*

$$\frac{1}{\gamma\eta} + \frac{1}{\gamma}\frac{1}{\mathrm{MSNR}} + \frac{1+\gamma\eta}{\gamma\eta}\frac{1}{\mathrm{ISNR}} \leq \mathrm{mse}_0 \leq \frac{1}{\hat{\gamma}\eta} + \frac{1}{\hat{\gamma}}\frac{1}{\mathrm{MSNR}} + \frac{1+\hat{\gamma}\eta}{\hat{\gamma}\eta}\frac{1}{\mathrm{ISNR}}.$$

in high probability.

Remark 3. *When the matrix perturbation $E = 0$, we have $1/\eta = 0$. Hence, the inequalities can be simplified as*

$$\frac{1}{\gamma}\frac{1}{\mathrm{MSNR}} + \frac{1}{\mathrm{ISNR}} \leq \mathrm{mse}_0 \leq \frac{1}{\hat{\gamma}}\frac{1}{\mathrm{MSNR}} + \frac{1}{\mathrm{ISNR}}.$$

Furthermore, if $\mathbf{E} = \mathbf{0}$ and $\mathbf{e} = \mathbf{0}$, we have $\mathrm{ISNR} = \infty$. Hence,

$$\frac{1}{\gamma}\frac{1}{\mathrm{MSNR}} \leq \mathrm{mse}_0 \leq \frac{1}{\hat{\gamma}}\frac{1}{\mathrm{MSNR}},$$

where the lower and upper bounds of the normalized mean square error is proportional to $1/\mathrm{MSNR}$. Similarly, if $\mathbf{E} = \mathbf{0}$ and $\mathbf{n} = \mathbf{0}$, we have $\mathrm{MSNR} = \infty$. Thus,

$$\mathrm{mse}_0 = \frac{1}{\mathrm{ISNR}}.$$

I.e., when the measurement noise $\mathbf{n} = \mathbf{0}$, the normalized average recovery error is exactly $1/\mathrm{ISNR}$.

Remark 4. *When the measurement noise and input noise are zero, i.e. $\mathbf{n} = \mathbf{0}$, $\mathbf{e} = \mathbf{0}$, we have $\mathrm{ISNR} = \infty$ and $\mathrm{MSNR} = \infty$. In this case,*

$$\frac{1}{\gamma\eta} \leq \mathrm{mes}_0 \leq \frac{1}{\hat{\gamma}\eta}.$$

Hence, the lower and upper bounds of mse_0 are proportional to $1/\eta = \sigma_E^2/\sigma_A^2$.

4 Background

4.1 Column independece

The full-rankness of the sensing matrix is beneficial for the reconstruction of signals. The singularity probability of Bernoulli random matrices has been discussed in [18] and [24]. Fortunately, it is almost sure that a tall Bernoulli matrix is full-rank. Indeed, it is proved by Kahn et al. that the probability of the columns of a tall Bernoulli matrix being linear-dependent is almost zero.

Lemma 1. *([18]) The probability that the columns of A_Ω are linear dependent*

$$\mathcal{P}_{\mathcal{S}} = \frac{[1+o(1)]K(K-1)}{2^M}$$

Given the condition that $N > M > K$, the columns of A_Ω are almost surely linear independent. We will continue to apply the notation $\mathcal{P}_{\mathcal{S}}$ in the following text.

4.2 Trace of the inverse of a positive definite matrix

. Estimating the trace of a particular matrix is meaningful for the analysis of recovered signal. The following lemma presented by [2] demonstrated the lower and upper bounds of the inverse of a positive definite matrix.

Lemma 2. *([2]) Let* $\mathbf{A}$ *be an n-by-n symmetric positive definite matrix,* $\mu_1 = \mathbf{tr}(\mathbf{A})$*,* $\mu_2 = \|\mathbf{A}\|_F^2 = \sum_{i,j=1}^{n} \mathbf{A_{ij}^2}$ *, and the set of eigenvalues* $\lambda(\mathbf{A}) \subset [\alpha, \beta]$ *with* $\alpha > 0$*, then*

$$\begin{pmatrix} \mu_1 & n \end{pmatrix} \mathbf{J_A^{-1}}(\beta) \begin{pmatrix} n \\ 1 \end{pmatrix} \leq \mathbf{tr}(\mathbf{A^{-1}}) \leq \begin{pmatrix} \mu_1 & n \end{pmatrix} \mathbf{J_A^{-1}}(\alpha) \begin{pmatrix} n \\ 1 \end{pmatrix} \text{ where}$$

$$\mathbf{J_A}(\xi) = \begin{pmatrix} \|\mathbf{A}\|_\mathbf{F}^\mathbf{2} & \mathrm{tr}(\mathbf{A}) \\ \xi^2 & \xi \end{pmatrix}$$

4.2.1 The Smallest Singular Value of a Rectangular Bernoulli Random Matrix

Lemma 2 illustrated the lower and upper bounds of $\mathbf{tr}(\mathbf{A^{-1}})$, which are directly related to the lower and upper bounds of $\mathbf{A}$'s eigenvalues. This article will particularly apply the second inequility, i.e. the upper bound part, in the analysis to follow. Given that this article focuses on tall Bernoulli sensing matrices, it is necessary to estimate the least singular value of this class of random matrices. In the work of Vershynin[26], a probabilistic result is shown concerning the lower bound of least singular value related to the size of matrix, constants determined solely by the matrix, as well as probability parameter t.

Before introducing the lemma, it is necessary to specify the definition of an *isotropic random vector*.

Definition 1. *([26]) A random vector* $\mathbf{X} \in \mathbb{R}^\mathbf{n}$ *is called isotropic if*

$$\mathbb{E}(\langle \mathbf{X}, \mathbf{x} \rangle^2) = \|\mathbf{x}\|_\mathbf{2}^\mathbf{2} \text{ for all } \mathbf{x} \in \mathbb{R}^\mathbf{n}$$

Obviously, a ± 1 Bernoulli random vector is isotropic, a simple result also shown in the same article of Vershynin's.

The following lemma presents a probabilistic result on tall matrices with independent sub-Gaussian isotropic columns.

Lemma 3. *([26]) Let* $\mathbf{A}$ *be an N-by-n matrix (n≤N) whose columns are independent sub-Gaussian isotropic random vectors in* $\mathbb{R}^N$ *with* $\|\mathbf{A_j}\|_2 = \sqrt{N}$*. Then for every* $t \geq 0$*, the inequality* $\sqrt{N} - C\sqrt{n} - t \leq s_{min} \leq s_{max} \leq \sqrt{N} + C\sqrt{n} + t$ *with probability at least* $1 - 2exp(-ct^2)$*, where* $C = C_K^{'}, c = c_K^{'} > 0$ *depend only on the sub-Gaussian norm* $\max_j \|\mathbf{A_j}\|_\mathbf{2}$ *of the columns.*

Interestingly, for ± 1 Bernoulli matrices, $\|\mathbf{A_j}\|_2 = \sqrt{M}$. Therefore, $C,\ c$ are constants depending solely on $\sqrt{M}$.

The following lemma from literature explicitly indicates the expression of the expectation of the pseudo-inverse of a Wishart matrix $(\mathbf{U_\Omega U_\Omega})^\dagger$

Lemma 4. *([13]) When the sensing matrix is Gaussian, and $M > K + 3$,*

$$\mathbb{E}[(\mathbf{U_\Omega U_\Omega})^\dagger] = \frac{K}{M(M-K-1)\sigma_A^2}\mathbf{I}$$

5 Theoretical Analysis

The measurement contaminated simultaneously by the perturbation matrix $\mathbf{E}$, the additional noise vector $\mathbf{n}$ and the input noise $\mathbf{e}$ can be specified as

$$\mathbf{y} = \mathbf{A\Psi}\boldsymbol{\theta} + (\mathbf{E\Psi(\boldsymbol{\theta}+\boldsymbol{\varepsilon})} + \mathbf{n}) + \mathbf{A\Psi}\boldsymbol{\varepsilon} = \mathbf{U}\boldsymbol{\theta} + \mathbf{z_e},$$

where $\mathbf{U} = \mathbf{A\Psi}$, $\mathbf{z_e} = \mathbf{z} + \mathbf{A\Psi}\boldsymbol{\varepsilon}$, $\mathbf{z} = \mathbf{E\Psi}(\boldsymbol{\theta}+\boldsymbol{\varepsilon}) + \mathbf{n}$.

According to the analysis in [13],

$$\text{mse} = \mathbb{E}(\|\mathbf{x} - \hat{\mathbf{x}}\|_2^2) = \mathbb{E}(\|\boldsymbol{\theta}_\Omega - \hat{\boldsymbol{\theta}}_\Omega\|_2^2).$$

Since $\mathbf{A}$, $\boldsymbol{\theta}$ and $\mathbf{n}$ are independent, $\mathbf{U_\Omega}$ and $\mathbf{z}$ are independent. When the columns of $\mathbf{U_\Omega}$ are independent, and

$$\boldsymbol{\theta}_\Omega - \hat{\boldsymbol{\theta}}_\Omega = \mathbf{U}_{\boldsymbol{\Omega}}^\dagger\mathbf{z}.$$

A concrete expression of $\mathbb{E}(\mathbf{z^T z})$ is presented in the following proposition.

Proposition 2. *When the columns of $\mathbf{U_\Omega}$ are independent,*

$$\begin{aligned}\text{mse} &= \mathbb{E}\|\hat{\mathbf{x}} - \mathbf{x}\|_\mathbf{2}^\mathbf{2} \\ &= \mathbb{E}[\mathbf{z^T}(\mathbf{U_\Omega U_\Omega^T})^\dagger\mathbf{z}] + K\sigma_e^2.\end{aligned}$$

Proof. The proof of Proposition 2 is postponed to Appendix 9.1. □

Proposition 3.

$$\mathbb{E}(\mathbf{z^T z}) = \text{tr}(\boldsymbol{\Sigma}_\mathbf{n}) + KM\sigma_E^2\sigma_e^2 + M\sigma_E^2\mathbf{p_x}.$$

When $\mathbf{n}$ is white noise, $\text{tr}(\boldsymbol{\Sigma}_\mathbf{n}) = M\sigma_n^2$, we have

$$\mathbb{E}(\mathbf{z^T z}) = M(\sigma_n^2 + K\sigma_E^2\sigma_e^2 + \sigma_E^2\mathbf{p_x}).$$

Proof. The proof of Proposition 3 is postponed to Appendix 9.2. □

5.0.2 Lower Bound

In order to avoid complicated calculation, assume $\mathbf{\Psi} = \mathbf{I_N}$ in the following analysis. Thus, $\mathbf{U} = \mathbf{A\Psi} = \mathbf{A}$. Hence,

$$\mathbf{y} = \mathbf{A}\boldsymbol{\theta} + (\mathbf{E}(\theta + \boldsymbol{\varepsilon}) + \mathbf{n}) + \mathbf{A}\boldsymbol{\varepsilon},$$

and

$$\text{mse} = \mathbb{E}(\|\mathbf{x} - \hat{\mathbf{x}}\|_2^2) = \mathbb{E}(\|\boldsymbol{\theta}_\Omega - \hat{\boldsymbol{\theta}}_\Omega\|_2^2),$$

$$\boldsymbol{\theta}_\Omega - \hat{\boldsymbol{\theta}}_\Omega = \mathbf{A}_{\mathbf{\Omega}}^{\dagger}\mathbf{z}.$$

Similarly, the mean square error of the recovered signal is

$$\text{mse} = \mathbb{E}(\|\mathbf{A}_{\mathbf{\Omega}}^{\dagger}\mathbf{z}\|_2^2) = \mathbb{E}[\mathbf{z}^{\mathbf{T}}\mathbb{E}[(\mathbf{A_\Omega}\mathbf{A}_{\mathbf{\Omega}}^{\mathbf{T}})^{\dagger}]\mathbf{z}] + K\sigma_e^2.$$

Denote $\mathbf{B} = \mathbf{A_\Omega}\mathbf{A}_{\mathbf{\Omega}}^{\mathbf{T}}$. Assume the nonzero eigenvalues of $\mathbf{B}$ are $\{\lambda_1, \cdots, \lambda_K\}$, $\mathbf{\Lambda} = \text{diag}\{\lambda_1, \cdots, \lambda_K\}$. Notice that $\mathbf{B}$ is positive semi-definite. Thus, $\lambda_i > 0$, $i = 1, \cdots, K$. Therefore, the set of nonzero eigenvalues of $\mathbf{B}^{\dagger}$ is $\mathbf{\Lambda}^{\dagger} = \{1/\lambda_1, \cdots, 1/\lambda_K\}$.

Denote $\mathbf{C} = \mathbf{A}_{\mathbf{\Omega}}^{\mathbf{T}}\mathbf{A_\Omega}$. Notice that $\mathbf{C}$ is a K-by-K symmetric matrix. With overwhelming probability ($\mathcal{P}_{\mathcal{I}} = 1 - \mathcal{P}_{\mathcal{S}}$), the columns of $\mathbf{A_\Omega}$ are independent. Thus, with probability

$$\mathcal{P}_{\mathcal{I}} = 1 - \frac{[1 + o(1)]K(K-1)}{2^M} \approx 1,$$

$\mathbf{C}$ is positive definite matrix and that tr($\mathbf{B}^{\dagger}$)=tr($\mathbf{C}^{-1}$).

The following proposition illustrates the trace of $\mathbf{B}^{\dagger}$ is lower-bounded by $\kappa_A^{-2}K$.

Proposition 4. *The trace of* $\mathbf{B}^{\dagger}$ *is lower-bounded by* $\kappa_A^{-2}K$*,*

$$\text{tr}(\mathbf{B}^{\dagger}) \geq \kappa_A^{-2}K$$

Proof. The proof of Proposition 4 is presented in Appendix 9.3. □

5.0.3 Upper Bound

This section concentrates on discussing the upper bound of the mean square error of the reconstructed signal. In the following discussions, the relationship between M and K is further specified. Assume $M = \gamma K$, where $\gamma^2 \gg 1$ (the simulation results presented later show that $\gamma >$ 5 is sufficient for $\gamma^2 \gg 1$). Thus,

$$N > M = \gamma K > K \gg 1,$$

$$\frac{K^2}{M^2} \ll 1.$$

In the following analysis, denote

$$\mu_1 = \text{tr}(\mathbf{C}),$$

$$\mu_2 = \sum_{i,j=1}^{K} \mathbf{C}_{ij}^2.$$

When $\mathbf{C}$ is positive definite, Lemma 2 therefore leads to the following inequality

$$\mathrm{tr}(B^{\dagger}) = \mathrm{tr}(\mathbf{C^{-1}}) \leq \mathscr{U}_{\alpha}(\mathbf{C}).$$

$\mathscr{U}_{\alpha}$ as defined above is the upper bound of $\mathbf{tr}(\mathbf{B}^{\dagger})$, α is the smallest eigenvalue of $\mathbf{C}$. Hence, the square of the least singular value of $\mathbf{A_{\Omega}}$. The following lemma provides a concentration analysis of μ_2, examining how reliable it is to estimate μ_2 with its expectation $\mathbb{E}(\mu_2)$.

Lemma 5. *(Concentration analysis) If $\gamma^2 = M^2/K^2 \gg 1$, then with high probability, μ_2 is close to its average $\mathbb{E}(\mu_2)$. More rigorously,*

$$P(|\mu_2 - \mathbb{E}(\mu_2)| \geq \kappa_A^4) \ll 1.$$

Proof. The proof of Lemma 5 is shown in Appendix 9.4. □

Note that the upper bound $\mathscr{U}_{\alpha}$ is a funciton of α. Unfortunately, α is a random variable related to the specific matrix $\mathbf{A}$ and the only information available is its lower bound. Therefore, it is significant to find the probability that $\mathscr{U}_{\alpha}$ is monotonically decreasing, i.e. $P(\partial \mathscr{U}_{\alpha}/\partial \alpha < 0)$ in order to derive an upper bound effective for all $\alpha > \alpha_0$, where α_0 is the probabilistic lower bound of α.

Lemma 6. *When $\mathbf{C}$ is positive definite, if $K \gg 1$, the upper bound of $\mathbf{tr}(\mathbf{C^{-1}})$, i.e. $\mathscr{U}_{\alpha}$ decreases as α increases with high probability:*

$$P(\frac{\partial \mathscr{U}_{\alpha}}{\partial \alpha} < 0) \approx 1$$

Proof. The proof of Lemma 6 is available in Appendix 9.5. □

In order to apply the method utilized in the proof of Proposition 3, we need to find out the expectation of $\mathrm{tr}(B^{\dagger})$.

Proposition 5. *If the columns of $\mathbf{A_{\Omega}}$ are independent, the following inequality holds*

$$\mathbb{E}(\mathbf{tr}(\mathbf{B}^{\dagger})) = \mathbb{E}(\mathbf{tr}(\mathbf{C^{-1}})) \leq \frac{K}{M\sigma_A^2}\mathscr{F}(M, K, \theta_0),$$

where

$$\theta_0 = (\frac{\sqrt{M} - C\sqrt{K} - t}{\sqrt{M}})^2$$

with probability at least $p_0 + p_1 - 1$, where

$$p_1 \geq 1 - \frac{1}{\gamma^2},$$
$$p_0 = 1 - 2exp(-ct^2).$$

$C = C_K^{'}, c = c_K^{'} > 0$ *are constants depending merely on* $\sqrt{M}$.

Proof. The proof of Proposition 5 is shown in Appendix 9.6. □

6 Related work

The most closely related work to our result is [13]. The author demonstrated the performance analysis of Gaussian sensing matrix when the information of the support set is available at the receiver.

Lemma 7. *([13]) For the simple noise model* $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}$ *where the entries of* $\mathbf{A}$ *are i.i.d. Gaussian random variables with zero mean and variance* σ_A^2*. If* $M > K + 3$*, the average reconstruction error by the oracle estimator, when the support* Ω *of* $\mathbf{x}$ *is available at the receiver is*

$$\mathbb{E}[\|\hat{\mathbf{x}} - \mathbf{x}\|_2^2] = \frac{K}{M(M-K-1)} \frac{\mathbf{tr}(\Sigma_n)}{\sigma_A^2}$$

when $\mathbf{n}$ *is white Gaussian noise, the expression can be simplified as*

$$\mathbb{E}[\|\hat{\mathbf{x}} - \mathbf{x}\|_2^2] = \frac{K}{M-K-1} \frac{\sigma_n^2}{\sigma_A^2}$$

The average recovery error of any reconstruction algorithm is lower bounded by that of the oracle estimator.

We introduce the following proposition as a generalization of Lemma 7 considering $\mathbf{n}$, $\mathbf{e}$, and $\mathbf{E}$ simultaneously.

Proposition 6. *When the entries of the sensing matrix* $\mathbf{A} \in \mathbb{R}^{M\times N}$*, and the perturbation matrix* $\mathbf{E} \in \mathbb{R}^{M\times N}$*,* $M < N$*, are i.i.d. Gaussian random variables that follow* $\mathcal{N}(0, \sigma_A^2)$ *and* $\mathcal{N}(0, \sigma_E^2)$*, respectively.* $\mathbf{x} = \mathbf{\Psi}\boldsymbol{\theta}$*, where* $\mathrm{supp}(\boldsymbol{\theta}) = \Omega$ *and* $K = |\Omega| < M - 3$*. When the columns of* $\mathbf{U}_{\mathbf{\Omega}}$ *are linearly independent, the average reconstruction error*

$$\begin{aligned}
\mathrm{mse}_0 &= \frac{K}{M-K-1}\frac{1}{\mathrm{MSNR}} + (\frac{K}{M-K-1}\frac{1}{\eta} + 1)\frac{1}{\mathrm{ISNR}} \\
&\quad + \frac{1}{\eta}\frac{K}{M-K-1} \\
&\approx \frac{1}{\gamma-1}\frac{1}{\mathrm{MSNR}} + (\frac{1}{\gamma-1}\frac{1}{\eta} + 1)\frac{1}{\mathrm{ISNR}} \\
&\quad + \frac{1}{\eta}\frac{1}{\gamma-1}.
\end{aligned} \tag{1}$$

Proof. The proof of Proposition 6 is postponed to Appendix 9.8. □

Remark 5. *Notice that in Proposition 6, when the perturbation matrix* $\mathbf{E} = \mathbf{0}$*,* $\sigma_E = 0$*, and the input noise* $\sigma_e = 0$*, the expression becomes the expression in Lemma 1.*

7 Numerical Results

In the simulation of the measurement system with Bernoulli sensing matrix, the nonzero part of the signal $\boldsymbol{\theta}_\Omega$ is assumed to be a non-correlated normal distributed vector with zero mean and $\mathbf{\Sigma}_\theta = \sigma_\theta^2 I$. All the results are averaged in 1000 trials.

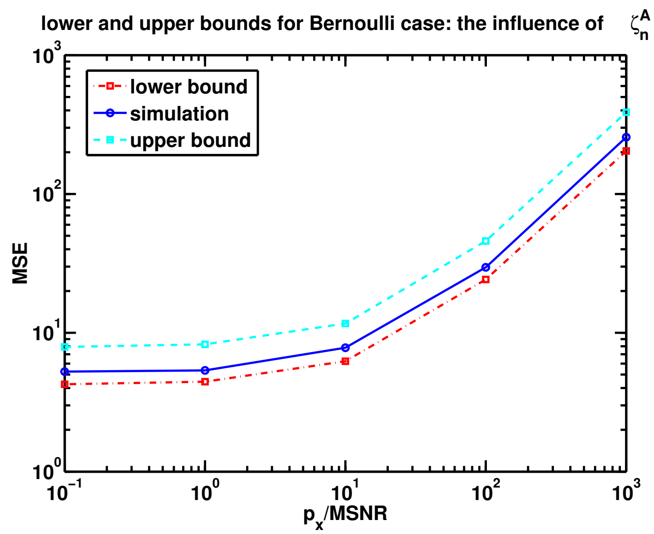

Figure 1: Oracle reconstruction error in Bernoulli case. Simulation vs. theoretical lower and upper bounds. $N = 625, K = 20, M = 100$.

During the simulations, assume $C_K^, = 1, t = 1$. Let $M = 100, N = 625, K = 20$.

The first simulation (Fig. 1) in this section concentrates on exploring the influence of MSNR. $\mathbf{p_x}$MSNR ranges from 10^{-1} to 10^3.

The second simulation focuses on the impact of η. In the experiments, $1/\eta$ is taken from 10^{-1} to 10^3. (See Fig. 2)

The third experiment (Fig. 3) concentrates on the influence of ISNR. Particularly, 1/ISNR is taken from 10^{-3} to 10^1.

According to the simulation results, the theoretical lower and upper bounds are proved correct and reliable. Moreover, since those bounds are actually very close to the simulation result in practice, they can serve as good estimations of the mean square error.

8 Conclusion

Proposition 1 leads to the main conclusion in the Bernoulli case. If the dimension of the sensing matrix satisfy the following inequalities:

$$1 \ll K < \gamma K = M < N,$$
$$\gamma^2 \gg 1,$$

then with high probability, $\text{rank}(\mathbf{A}_{\mathbf{\Omega}}) = K = |\Omega|$, and the mean square error of the recovered signal is lower and upper bounded by

$$\text{mse}_0 \geq \frac{1}{\gamma\eta} + \frac{1}{\gamma}\frac{1}{\text{MSNR}} + \frac{1+\gamma\eta}{\gamma\eta}\frac{1}{\text{ISNR}}$$

and

$$\text{mse}_0 \leq \frac{1}{\hat{\gamma}\eta} + \frac{1}{\hat{\gamma}}\frac{1}{\text{MSNR}} + \frac{1+\hat{\gamma}\eta}{\hat{\gamma}\eta}\frac{1}{\text{ISNR}}.$$

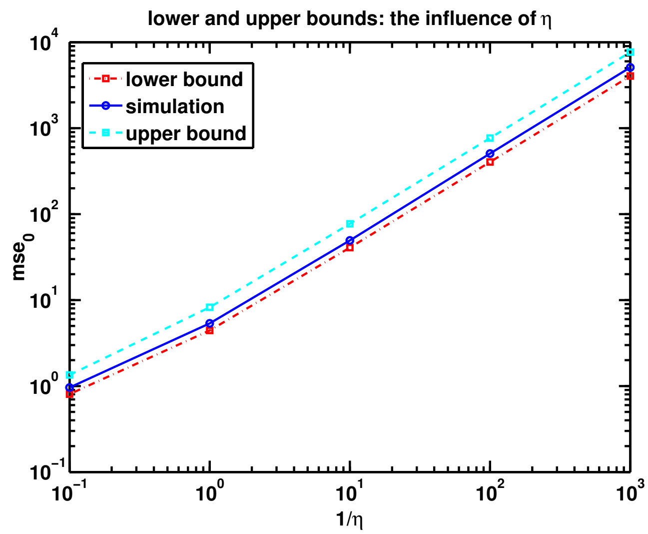

Figure 2: Oracle reconstruction error in Bernoulli case. Simulation vs. theoretical lower and upper bounds. $N = 625, K = 20, M = 100$.

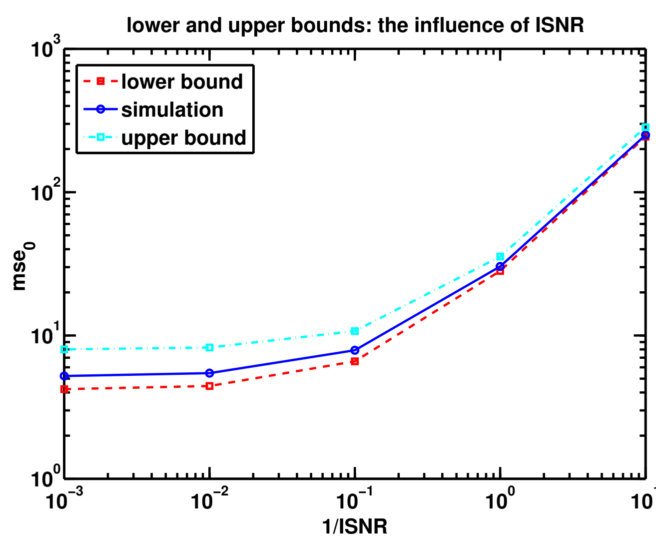

Figure 3: Oracle reconstruction error in Bernoulli case. Simulation vs. theoretical lower and upper bounds. $N = 625, K = 20, M = 100$.

where

$$\theta_0 = (\frac{\sqrt{M} - C\sqrt{K} - t}{\sqrt{M}})^2$$
$$\mathscr{F}(M, K, \theta_0) = \gamma \sigma_{\mathbf{A}}^{\mathbf{2}} \mathscr{U}_{\alpha_{\mathbf{0}}}$$

and

$$\frac{1}{\hat{\gamma}} = \frac{1}{\gamma}(1 + \frac{1}{\gamma}).$$

The first inequality holds with probability $\mathcal{P}_{\mathcal{L}} = 1 - \mathcal{P}_{\mathcal{S}}$ while the second inequality is held with probability at least $\mathcal{P}_{\mathcal{U}} = 1 - 1/\gamma^2 - 2\exp(-ct^2) - \mathcal{P}_{\mathcal{S}}$. The constants $C = C_K^{'}, c = c_K^{'} > 0$ are constants depending only on $\sqrt{M}$. Future research may focus on deriving an accurate equation of the average reconstruction error in the Bernoulli case. It is also meritorious to explore a generalized expression for the sub-Gaussian case without unduly sacrificing concreteness of results.

9 Appendix

9.1 Proof of Proposition 2

Proof. The preassumption that the columns of $\mathbf{U}_{\mathbf{\Omega}}$ are linearly independent indicates that $\mathbf{U}_{\mathbf{\Omega}}^{\dagger} = (\mathbf{U}_{\mathbf{\Omega}}^{\mathbf{T}} \mathbf{U}_{\mathbf{\Omega}})^{-\mathbf{1}} \mathbf{U}_{\mathbf{\Omega}}^{\mathbf{T}}$. Recall that

$$\mathbf{x} = \mathbf{\Psi} \boldsymbol{\theta}$$
$$\tilde{\mathbf{e}} = \mathbf{\Psi} \boldsymbol{\varepsilon}$$
$$\mathbf{F} = \mathbf{E} \mathbf{\Psi}$$

Thus,

$$\begin{aligned} \mathbf{y} &= (\mathbf{A} + \mathbf{E})(\mathbf{x} + \mathbf{e}) + \mathbf{n} \\ &= \mathbf{U}_{\mathbf{\Omega}} \theta + (\mathbf{F} \boldsymbol{\theta} + \mathbf{F} \varepsilon + \mathbf{n}) + \mathbf{A} \mathbf{e} \end{aligned}$$

$$\begin{aligned} \mathbb{E}[\|\hat{\mathbf{x}} - \mathbf{x}\|_{\mathbf{2}}^{\mathbf{2}}] &= \mathbb{E}[\|\mathbf{U}_{\mathbf{\Omega}}^{\dagger} \mathbf{z}_{\mathbf{e}}\|_{\mathbf{2}}^{\mathbf{2}}] \\ &= \mathbb{E}[\mathbf{z}^{\mathbf{T}} (\mathbf{U}_{\mathbf{\Omega}} \mathbf{U}_{\mathbf{\Omega}}^{\mathbf{T}})^{\dagger} \mathbf{z}] \\ &+ \mathbb{E}[\boldsymbol{\varepsilon}^{\mathbf{T}} \mathbf{U}_{\mathbf{\Omega}}^{\dagger} \mathbf{z} + \mathbf{z}^{\mathbf{T}} (\mathbf{U}_{\mathbf{\Omega}}^{\dagger})^{\mathbf{T}} \boldsymbol{\varepsilon}] + \mathbb{E}[\boldsymbol{\varepsilon}^{\mathbf{T}} \boldsymbol{\varepsilon}] \end{aligned}$$

where $\mathbf{z}_{\mathbf{e}} = \mathbf{z} + \mathbf{A}\mathbf{e} = \mathbf{z} + \mathbf{U}_{\mathbf{\Omega}} \varepsilon$. Notice that $\mathbf{z} = \mathbf{F}\boldsymbol{\theta} + \mathbf{F}\varepsilon + \mathbf{n}$, is independent from $\mathbf{U}_{\mathbf{\Omega}}$.

$$\begin{aligned} \text{mse} &= \mathbb{E}[\mathbf{z}^{\mathbf{T}} \mathbb{E}[(\mathbf{U}_{\mathbf{\Omega}} \mathbf{U}_{\mathbf{\Omega}}^{\mathbf{T}})^{\dagger}] \mathbf{z}] \\ &+ 2\mathbb{E}[\boldsymbol{\varepsilon}^{\mathbf{T}} \mathbf{U}_{\mathbf{\Omega}}^{\dagger} \mathbf{z}] + K\sigma_e^2 \\ &= \mathbb{E}[\mathbf{z}^{\mathbf{T}} \mathbb{E}[(\mathbf{U}_{\mathbf{\Omega}} \mathbf{U}_{\mathbf{\Omega}}^{\mathbf{T}})^{\dagger}] \mathbf{z}] + K\sigma_e^2 \end{aligned}$$

□

9.2 Proof of Proposition 3

Proof.

$$\mathbb{E}(\mathbf{z^T z}) = \mathbb{E}[(\mathbf{Ex} + \mathbf{E\tilde{e}} + \mathbf{n})^\mathbf{T}(\mathbf{Ex} + \mathbf{E\tilde{e}} + \mathbf{n})]$$

Notice that $\mathbf{E}, \mathbf{x}, \mathbf{\tilde{e}}, \textbf{and } \mathbf{n}$ are independent

$$\mathbb{E}(\mathbf{z^T z}) = \mathbb{E}[\mathbf{n^T n}] + \mathbb{E}[\mathbf{\tilde{e}^T}(\mathbf{E^T E})\mathbf{\tilde{e}}] + \mathbb{E}[\mathbf{x^T}(\mathbf{E^T E})\mathbf{x}]$$

Apparently,

$$\mathbb{E}(\mathbf{E^T E}) = M\sigma_E^2 \mathbf{I_N}$$

Therefore,

$$\mathbb{E}(\mathbf{z^T z}) = \mathrm{tr}(\mathbf{\Sigma_n}) + KM\sigma_E^2\sigma_e^2 + M\sigma_E^2\mathbf{p_x}$$

When $\mathbf{n}$ is white noise, $\mathrm{tr}(\mathbf{\Sigma_n}) = M\sigma_n^2$

$$\mathbb{E}(\mathbf{z^T z}) = M(\sigma_n^2 + K\sigma_E^2\sigma_e^2 + \sigma_E^2\mathbf{p_x})$$

□

9.3 Proof of Proposition 4

Proof.

$$\mathbf{B} = \mathbf{V\Sigma V^T}$$

$$\mathbf{B}^\dagger = \mathbf{V\Sigma}^\dagger\mathbf{V^T}$$

$$\lambda_1 + \cdots + \lambda_K = \mathrm{tr}(\mathbf{B}) = \kappa_A^2 K$$

The Cauchian inequality gurantees that

$$(\lambda_1 + \cdots + \lambda_K)(\frac{1}{\lambda_1} + \cdots + \frac{1}{\lambda_K}) \geq K^2$$

Therefore,

$$\frac{1}{\lambda_1} + \cdots + \frac{1}{\lambda_K} \geq \kappa_A^{-2}K$$

□

9.4 Proof of Lemma 5

Proof.

$$
\begin{aligned}
c_{ij} &= \sum_{k=1}^{M} a_{ki} a_{kj} \\
c_{ii} &= \kappa_A^2 \\
c_{ij}^2 &= \frac{\kappa_A^4}{M} + \sum_{s,t=1,s\neq t}^{M} a_{si} a_{sj} a_{ti} a_{tj} \\
\mathbb{E}(c_{ij}^2) &= \frac{\kappa_A^4}{M}, \text{if } i \neq j \\
\mu_2 &= \sum_{i,j=1}^{K} c_{ij}^2 \\
&= \kappa_A^4 [\frac{K(K-1)}{M} + K] \\
&\quad + \sum_{i,j=1,i\neq j}^{K} \sum_{s,t=1,s\neq t}^{M} a_{si} a_{sj} a_{ti} a_{tj} \\
\mathbb{E}(\mu_2) &= \kappa_A^4 [\frac{K(K-1)}{M} + K] \\
\mathbb{E}(\mu_2^2) &= \mathbb{E}(\mu_2)^2 + \frac{\kappa_A^8 K(K-1)(M-1)}{M^3} \\
\mathrm{Var}(\mu_2) &= \frac{\kappa_A^8 K(K-1)(M-1)}{M^3} \leq \kappa_A^8 \frac{K^2}{M^2} \ll 1
\end{aligned}
$$

Therefore,

$$\mathbb{E}(\mu_2) = \kappa_A^4 [\frac{K(K-1)}{M} + K]$$

is a good estimation of μ_2. More specifically, Chebyshev's inequality leads to

$$P(|\mu_2 - \mathbb{E}(\mu_2)| \geq \kappa_A^4) \leq \mathrm{Var}(\mu_2)/\kappa_A^8 \ll 1$$

□

9.5 Proof of Lemma 6

Proof.

$$\frac{\partial \mathscr{U}_\alpha}{\partial \alpha} = \frac{(\kappa_A^4 K - \mu_2)(K\alpha^2 - 2\kappa_A^2 K\alpha + \mu_2)}{(\mu_2 \alpha - \kappa_A^2 K \alpha^2)^2}$$

With *Lemma* 4 and *Chebyshev's Inequality*,

$$P(|\mu_2 - \mathbb{E}(\mu_2)| \geq \epsilon) \leq \frac{\mathrm{Var}(\mu_2)}{\epsilon^2}$$

Let $\epsilon = \frac{\kappa_A^4 K(K-1)}{M}$

$$
\begin{aligned}
P(\mu_2 \le \kappa_A^4 K) &\le P(|\mu_2 - \kappa_A^4(\frac{K(K-1)}{M} - K)| \ge \frac{\kappa_A^4 K(K-1)}{M}) \\
&\le \frac{\mathrm{Var}(\mu_2)}{\epsilon^2} \\
&= \frac{M-1}{MK(K-1)} \\
&\approx \frac{1}{K^2}
\end{aligned}
$$

therefore,

$$P(\frac{\partial \mathscr{U}_\alpha}{\partial \alpha} < 0) \ge P(\mu_2 > \kappa_A^4 K) \ge 1 - \frac{M-1}{MK(K-1)} \approx 1$$

□

9.6 Proof of Proposition 5

Proof. Consider the following three events:

$$
\begin{aligned}
X_1 :\ & \{\alpha \ge \alpha_0 = \sigma_A^2(\sqrt{M} - C\sqrt{K} - t)^2\} \\
X_2 :\ & \{|\mu_2 - \mathbb{E}(\mu_2)| \le \kappa_A^4\} \\
X_3 :\ & \{\frac{\partial \mathscr{U}_\alpha}{\partial \alpha} < 0\}
\end{aligned}
$$

Consider events X_2 and X_3 first. *Lemma* 4 indicates that

$$P(X_2) \ge p_1$$

Lemma 5 leads to

$$P(X_3) \ge p_2 = 1 - \frac{M-1}{MK(K-1)}$$

It is necessary to investigate the relationship between X_2 and X_3

$$P(X_2) = P(K - 1 + \frac{K(K-1)}{M} < \frac{\mu_2}{\kappa_A^4} < K + 1 + \frac{K(K-1)}{M})$$

Notice that

$$\{\kappa_A^4 K < \mu_2\} \subset X_3$$

Since $M = \gamma K$, $K >> 1$,

$$(K-1) + \frac{K(K-1)}{M} > K$$

Thus,

$$X_2 \subset X_3$$

The probability that X_2 and X_3 coincide is

$$P(X_2, X_3) = P(X_2) \geq p_1$$

Then consider the event X_1. An estimation of the least singular value s_{min} can be readily derived by applying *Lemma* 3:

$$s_{min} \geq s_0 = \kappa_A \frac{\sqrt{M} - C\sqrt{K} - t}{\sqrt{M}}$$

with probability at least

$$p_0 = 1 - 2exp(-ct^2)$$

Therefore,

$$\alpha \geq \alpha_0 = \sigma_A^2(\sqrt{M} - C\sqrt{K} - t)^2$$

with probability at least p_0. We have

$$P(X_1) \geq p_0$$

Now we consider the probability that X_1, X_2, and X_3 coincide, which is equivalent to X_1 and X_2 coincide

$$P(X_1, X_2, X_3) = P(X_1, X_2)$$

Therefore,

$$P(X_1, X_2, X_3) = P(X_1, X_2)$$

$$\geq P(X_1) + P(X_2) - 1 \geq p_0 + p_1 - 1$$

When X_1, X_2, and X_3 coincide,

$X_1 = 1$ gurantees $\alpha \geq \alpha_0$, $X_3 = 1$ ensures that $U(\alpha)$ is *monotonically decreasing*. Thus,

$$\begin{aligned} \mathrm{tr}(\mathbf{B}^\dagger) &= \mathrm{tr}(\mathbf{C^{-1}}) \\ &\leq \mathscr{U}_\alpha \\ &\leq \mathscr{U}_{\alpha_0} \\ &= \frac{K}{\alpha_0}(1 - K\frac{(\alpha_0 - \kappa_A^2)^2}{\mu_2 - \kappa_A^2 K\alpha_0}) \end{aligned}$$

$X_2 = 1$ ensures that $\mu_2 \leq \mathbb{E}(\mu_2) + \kappa_A^4$. So we have

$$\begin{aligned} \mathscr{U}_{\alpha_0} &\leq \frac{K}{M\sigma_A^2\theta_0}(1 - K\frac{(\theta_0 - 1)^2}{K + 1 + \frac{K(K-1)}{M} - K\theta_0}) \\ &:= \frac{K}{M\sigma_A^2}\mathscr{F}(M, K, \theta_0) \end{aligned}$$

Therefore,

$$\mathbb{E}(\mathbf{tr}(\mathbf{B}^\dagger)) = \mathbb{E}(\mathbf{tr}(\mathbf{C^{-1}})) \leq \frac{K}{M\sigma_A^2}\mathscr{F}(M, K, \theta_0)$$

,
where

$$\theta_0 = (\frac{\sqrt{M} - C\sqrt{K} - t}{\sqrt{M}})^2$$

with probability at least

$$P = p_0 + p_1 - 1 = 1 - \frac{K(K-1)(M-1)}{M^3} - 2exp(-ct^2)$$

. Since p_0 and p_1 are very close to 1, this probability P also tends to 1.

Finally, consider the constants $C = C_K^{'}$ and $c = c_K^{'} > 0$. As per *Lemma* 3, $C = C_K^{'} = \frac{C(\sqrt{M})}{\sqrt{M}}$ and $c = c_K^{'} = \frac{c(\sqrt{M})}{\sqrt{M}} > 0$ are constants depending only on $\sqrt{M}$. □

9.7 Proof of Proposition 1

Proof. 1) Proof of Lower Bound: Denote $\mathbf{B} = \mathbf{A_\Omega A_\Omega^T}$

$$\mathbb{E}(\mathbf{B}^\dagger) = \mathbb{E}((\mathbf{U_\Omega U_\Omega^T})^\dagger)$$

Given the symmetry of B, apply the decomposition

$$\mathbb{E}(\mathbf{B}^\dagger) = \mathbf{Q\Lambda_0^\dagger Q^T}$$

where $\mathbf{Q}$ is an orthonormal matrix. Proposition 2 and Lemma 2 illustrates that with probability

$$\mathcal{P_I} = 1 - \mathcal{P_S}$$

the mse can be elaborated as

$$\text{mse} = \mathbb{E}[\mathbf{z^T}\mathbb{E}[(\mathbf{U_\Omega U_\Omega^T})^\dagger]\mathbf{z}] + \mathbf{p_e} = \mathbb{E}[\mathbf{z^T Q^T \Lambda_0^\dagger Q z}] + \mathbf{p_e}$$

where $\mathbf{\Lambda}_0^\dagger = \text{diag}\{1/\lambda_{1,0}, \cdots, 1/\lambda_{K,0}\}$
Notice that $\mathbf{Q}$ and $\mathbf{\Lambda}_0^\dagger$ are deterministic matrices.
Denote $\mathbf{w} = \mathbf{Qz}$. Therefore,

$$\begin{aligned}
\text{mse} &= \mathbb{E}[\mathbf{w^T}\mathbf{\Lambda}_0^\dagger\mathbf{w}] + \mathbf{p_e} \\
&= \mathbb{E}(\frac{\mathbf{w_1^2}}{\lambda_{1,0}} + \cdots + \frac{\mathbf{w_K^2}}{\lambda_{K,0}}) + \mathbf{p_e} \\
&= \frac{\mathbb{E}(\mathbf{w_1^2})}{\lambda_{1,0}} + \cdots + \frac{\mathbb{E}(\mathbf{w_K^2})}{\lambda_{K,0}} + \mathbf{p_e}
\end{aligned}$$

Since $\mathbf{w^T w} = \mathbf{z^T z}$, Proposition 3 indicates that

$$\begin{aligned}
\mathbb{E}(\mathbf{w^T w}) &= \mathbb{E}(\mathbf{w_1^2}) + \cdots + \mathbb{E}(\mathbf{w_M^2}) \\
&= \mathbb{E}(\mathbf{z^T z}) \\
&= M(\sigma_n^2 + K\sigma_E^2\sigma_e^2 + \sigma_E^2\mathbf{p_x})
\end{aligned}$$

Given

$$\mathbf{w}_i = \mathbf{Q}_{i1}\mathbf{z}_1 + \mathbf{Q}_{i2}\mathbf{z}_2 + \cdots + \mathbf{Q}_{iM}\mathbf{z}_M$$

and

$$\mathbf{z}_j = \mathbf{E}_{j1}(\mathbf{x}_1 + \tilde{\mathbf{e}}_1) + \mathbf{E}_{j2}(\mathbf{x}_2 + \tilde{\mathbf{e}}_2) + \cdots + \mathbf{E}_{jN}(\mathbf{x}_N + \tilde{\mathbf{e}}_N) + \mathbf{n}_j,$$

we have

$$\begin{aligned}
\mathbb{E}(\mathbf{z}_j^2) &= \sigma_E^2(\mathbf{p_x} + \mathbf{p_e}) + \sigma_n^2 \\
\mathbb{E}(\mathbf{w}_i^2) &= \mathbf{Q}_{i1}^2\mathbb{E}(\mathbf{z}_1^2) + \mathbf{Q}_{i2}^2\mathbb{E}(\mathbf{z}_2^2) + \cdots + \mathbf{Q}_{iM}^2\mathbb{E}(\mathbf{z}_M^2) \\
&= \sigma_E^2(\mathbf{p_x} + \mathbf{p_e}) + \sigma_n^2
\end{aligned}$$

With Proposition 2, $\mathrm{tr}(\mathbf{B}^\dagger) \geq \kappa_A^{-2}K$, then

$$\mathrm{tr}(\mathbb{E}[\mathbf{B}^\dagger]) = \frac{1}{\lambda_{1,0}} + \cdots + \frac{1}{\lambda_{K,0}} \geq \kappa_A^{-2}K$$

Therefore,

$$\begin{aligned}
\mathrm{mse} &\geq \frac{K}{M}\frac{[\sigma_{\mathbf{E}}^{\mathbf{2}}(\mathbf{p_x} + \mathbf{p_e}) + \sigma_{\mathbf{n}}^{\mathbf{2}}]}{\sigma_A^2} + \mathbf{p_e} \\
&= [\frac{1}{\gamma\eta} + \frac{1}{\gamma}\frac{1}{\mathrm{MSNR}} + (1 + \frac{1}{\gamma\eta})\frac{1}{\mathrm{ISNR}}]\mathbf{p_x}
\end{aligned}$$

with probability

$$\mathcal{P}_{\mathcal{L}} = \mathcal{P}_{\mathcal{I}}$$

9.7.1 Proof of Upper Bound

Assume that the columns of $\mathbf{A_\Omega}$ are independent, with Proposition 5,

$$\mathbb{E}(\mathbf{tr}(\mathbf{B}^\dagger)) = \mathbb{E}(\mathbf{tr}(\mathbf{C^{-1}})) \leq \frac{K}{M\sigma_A^2}\mathscr{F}(M, K, \theta_0),$$

where

$$\theta_0 = (\frac{\sqrt{M} - C\sqrt{K} - t}{\sqrt{M}})^2$$

with probability at least

$$\begin{aligned}
P &= 1 - \frac{K(K-1)(M-1)}{M^3} - 2\exp(-ct^2) \\
&\geq 1 - \frac{1}{\gamma^2} - 2\exp(-ct^2)
\end{aligned}$$

Considering the probability column independence assumption, we have the probability that this upper bound holds exceeds

$$\mathcal{P}_{\mathcal{U}} = P - \mathcal{P}_{\mathcal{S}}$$

$C = C_K^{'}, c = c_K^{'} > 0$ are constants depending only on $\sqrt{M}$ then the result can be derived using a similar method as the proof of lower bound. □

9.8 Proof of Proposition 6

Proof. Proposition 2 and Lemma 5 leads to

$$\begin{aligned}\text{mse} &= \mathbb{E}[\mathbf{z}^{\mathbf{T}}\mathbb{E}[(\mathbf{U}_{\mathbf{\Omega}}\mathbf{U}_{\mathbf{\Omega}}^{\mathbf{T}})^{\dagger}]\mathbf{z}] + K\sigma_e^2 \\ &= \frac{K}{M(M-K-1)\sigma_A^2}\mathbb{E}(\mathbf{z}^{\mathbf{T}}\mathbf{z}) + \mathbf{p_e}\end{aligned}$$

With Proposition 3,

$$\begin{aligned}\text{mse} &= \frac{K}{M(M-K-1)\sigma_A^2}[\text{tr}(\mathbf{\Sigma_n}) + KM\sigma_E^2\sigma_e^2 + M\sigma_E^2\mathbf{p_x}] + \mathbf{p_e} \\ &= [\frac{K}{M-K-1}\frac{1}{\text{MSNR}} + (\frac{K}{M-K-1}\frac{1}{\eta} + 1)\frac{1}{\text{ISNR}} \\ &\quad + \frac{1}{\eta}\frac{K}{M-K-1}]\mathbf{p_x}.\end{aligned}$$

□

References

[1] Arias-Castro, E., & Eldar, Y. C. (2011). Noise folding in compressed sensing. *Signal Processing Letters, IEEE*, 18(8), 478-481.

[2] Bai, Zhaojun, and Gene H. Golub. *Bounds for the trace of the inverse and the determinant of symmetric positive definite matrices.* Annals of Numerical Mathematics 4 (1996): 29-38.

[3] Bickel, P. J., Ritov, Y. A., & Tsybakov, A. B. (2009). Simultaneous analysis of Lasso and Dantzig selector. *The Annals of Statistics*, 1705-1732.

[4] Ben-Haim, Z., & Eldar, Y. C. (2010). The Cramer-Rao bound for estimating a sparse parameter vector. *Signal Processing, IEEE Transactions on*, 58(6), 3384-3389.

[5] Candes, E. J. (2014). Mathematics of sparsity (and few other things). *ICM 2014 Proceedings*, to appear.

[6] Candes, E., & Recht, B. (2011). Simple bounds for low-complexity model reconstruction. *Mathematical Programming Series A* 141(1–2), 577–589.

[7] Candes, E. J., Romberg, J. K., & Tao, T. (2006). Stable signal recovery from incomplete and inaccurate measurements. *Communications on pure and applied mathematics*, 59(8), 1207-1223.

[8] Candes, E., & Tao, T. (2007). The Dantzig selector: statistical estimation when p is much larger than n. The Annals of Statistics, 2313-2351.

[9] Candes, E. J., & Tao, T. (2005). Decoding by linear programming. *Information Theory, IEEE Transactions on*, 51(12), 4203-4215.

[10] Candes, E. J., Romberg, J., & Tao, T. (2006). Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *Information Theory, IEEE Transactions on*, 52(2), 489-509.

[11] Chen, L., & Gu, Y. (2013). Oracle-Order Recovery Performance of Greedy Pursuits With Replacement Against General Perturbations. *Signal Processing, IEEE Transactions on*, 61(18), 4625-4636.

[12] Cook, R. D., & Forzani, L. (2011). On the mean and variance of the generalized inverse of a singular wishart matrix. *Electronic Journal of Statistics*, 5, 146-158.

[13] Coluccia, G., Roumy, A., & Magli, E. (2014, May). Exact performance analysis of the oracle receiver for compressed sensing reconstruction. *In Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on* (pp. 1005-1009). IEEE.

[14] Davenport, M. A., Laska, J. N., treichler, J. R., & Baraniuk, R. G. (2012). The Pros and Cons of Compressive Sensing for Wideband Signal Acquisition: Noise Folding versus Dynamic Range. *IEEE transactions on signal processing*, 60(9), 4628-4642.

[15] Ding, J., Chen, L., & Gu, Y. (2013). Perturbation analysis of orthogonal matching pursuit. *Signal Processing, IEEE Transactions on*, 61(2), 398-410.

[16] Ding, J., Chen, L., & Gu, Y. (2012, March). Robustness of orthogonal matching pursuit for multiple measurement vectors in noisy scenario. *In Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on* (pp. 3813-3816). IEEE.

[17] Donoho, D. L. (2006). Compressed sensing. *Information Theory, IEEE Transactions on*, 52(4), 1289-1306.

[18] Kahn, J., Komlos, J., & Szemeredi, E. (1995). On the probability that a random ±1-matrix is singular. *Journal of the American Mathematical Society*, 8(1), 223-240.

[19] Maiwald, D., & Kraus, D. (2000). Calculation of moments of complex Wishart and complex inverse Wishart distributed matrices. *IEE Proceedings-Radar, Sonar and Navigation*, 147(4), 162-168.

[20] Mehta, M. L. (2004). *Random matrices* (Vol. 142). Academic press.

[21] Rauhut, H. (2010). Compressive sensing and structured random matrices. *Theoretical foundations and numerical methods for sparse recovery,* 9, 1-92.

[22] von Rosen, D. (1988). Moments for the inverted Wishart distribution. *Scandinavian Journal of Statistics*, 97-109.

[23] Tang, Y., Chen, L., & Gu, Y. (2013). On the performance bound of sparse estimation with sensing matrix perturbation. *Signal Processing, IEEE Transactions on*, 61(17), 4372-4386.

[24] Tao, T., & Vu, V. (2006). On random1 matrices: singularity and determinant. *Random Structures & Algorithms*, 28(1), 1-23.

[25] Tsaig, Y., & Donoho, D. L. (2006). Extensions of compressed sensing. *Signal processing*, 86(3), 549-571.

[26] Vershynin, R. (2010). Introduction to the non-asymptotic analysis of random matrices. *arXiv preprint arXiv*:1011.3027.

[27] Zhang, G., Jiao, S., Xu, X., & Wang, L. (2010, June). Compressed sensing and reconstruction with bernoulli matrices. In *Information and Automation (ICIA), 2010 IEEE International Conference on* (pp. 455-460). IEEE.