On the Achievability of Cramér-Rao Bound In Noisy Compressed Sensing

Rad Niazadeh, *Member, IEEE* Massoud Babaie-Zadeh, *Senior Member, IEEE,* and Christian Jutten, *Fellow, IEEE*

Abstract—Recently, it has been proved in [1] that in noisy compressed sensing, a joint typical estimator can asymptotically achieve the Cramér-Rao lower bound of the problem. To prove this result, [1] used a lemma, which is provided in [2], that comprises the main building block of the proof. This lemma is based on the assumption of Gaussianity of the measurement matrix and its randomness in the domain of noise. In this correspondence, we generalize the results obtained in [1] by dropping the Gaussianity assumption on the measurement matrix. In fact, by considering the measurement matrix as a deterministic matrix in our analysis, we find a theorem similar to the main theorem of [1] for a family of randomly generated (but deterministic in the noise domain) measurement matrices that satisfy a generalized condition known as The Concentration of Measures Inequality. By this, we finally show that under our generalized assumptions, the Cramér-Rao bound of the estimation is achievable by using the typical estimator introduced in [1].

Index Terms—Compressed Sensing, Joint Typicality, Typical Estimation, Chernof Bound

I. INTRODUCTION

Compressed Sensing (CS) which is also known as Compressive Sampling [3]–[5] is a well known method for taking linear measurements from a sparse vector. Compressed sensing proposes that one can recover a sparse signal from a few number of measurements, and so it can override the usual sampling method based on Nyquist criteria [3]. In this correspondence, we revisit the problem of signal recovery in noisy compressed sensing, in which the above mentioned measurements are blended with noise. Indeed, suppose that noisy measurements of the sparse signal are taken by a random measurement matrix in the following form:

$$\mathbf{y} = \mathbf{A}\mathbf{s} + \mathbf{n} , \tag{1}$$

in which, $\mathbf{s}$ is the original $M \times 1$ sparse signal, $\mathbf{y}$ is the $N \times 1$ vector of measurements, $\mathbf{n} \sim N(0, \sigma_n^2 I_{N \times N})$ is an $N \times 1$ Gaussian noise vector and $\mathbf{A} = [\mathbf{a}_1 \quad \mathbf{a}_2 \ldots \mathbf{a}_M]$ is an $N \times M$ measurement matrix whose elements are usually generated at random. More precisely, these elements are independent and identically distributed random variables drawn from some specific distributions (such as Gaussian, Bernoulli, etc.), so that the overall measurement matrix will be appropriate in the framework of recovery in compressive sampling [3], [4], [6], [7]. Suppose that $\mathbf{s}$ is sparse, i.e. $\|\mathbf{s}\|_0 = K \ll M$ where $\|.\|_0$ denotes the l_0-norm, i.e. the number of non-zero components of $\mathbf{s}$. Moreover, define $\tau \triangleq \mathrm{supp}(\mathbf{s})$ as a subset

R. Niazadeh is with the Electrical and Computer Engineering Department, Cornell University, Ithaca, NY 14850 (email: rn274@cornell.edu).

M. Babaie-Zadeh is with the Electrical Engineering Department, Sharif University of Technology, Tehran 14588-89694, Iran (e-mail: mbzadeh@yahoo.com).

C. Jutten is with the GIPSA-Lab, Department of Images and Signals, Institut National Polytechnique de Grenoble (INPG), 38031 Grenoble Cedex, France (e-mail: Christian.Jutten@inpg.fr).

of $\{1, 2 \ldots M\}$ that contains the indices of non-zero elements of $\mathbf{s}$, i.e. $\tau = \{i \in \{1, 2, \ldots M\} : s_i \neq 0\}$ in which s_i stands for the ith element of $\mathbf{s}$. For this model, one can also define the size parameters [1] as in [1]:

$$\alpha \triangleq \frac{K}{N} \quad \beta \triangleq \frac{M}{K} . \tag{2}$$

The main problem of compressive sampling is to estimate the unknown sparse signal from its noisy measurements which are taken as in (1). Many efforts have been done to find a practical recovery method and some acceptable solutions have been proposed in the literature whose computational cost are tolerable, such as the algorithms that are proposed in [9]–[16]. On the other hand, there is another related problem which is indeed the framework of our correspondence. In this problem, we are searching for the existence of an *efficient* estimator, an estimator that can achieve the Cramér-Rao lower bound [17] for the Mean Square Error (MSE) of the estimation. It is important to note that in this problem, computational complexity of the proposed estimator has no importance (or much less importance when comparing to practical methods), while the achievablity of Cramér-Rao bound and the *existence* of such an estimator is in the point of interest.

For our problem, two different Cramér-Rao lower bounds for MSE have been studied in [1], [18] depending on the amount of knowledge of the estimators about the sparsity structure of the original vector. The first bound, which is known as CRB-S [18], is the Cramér-Rao lower bound of a Genie Aided Estimation (GAE) problem in which the estimators know the location of the non-zero taps i.e. τ, as if a Genie has aided them with the location of the taps [18], [19]. This bound can be described in closed form as [18]:

$$\mathrm{CRB\text{-}S} = \sigma_n^2 \mathrm{Trace}\{(\mathbf{A}_\tau^T \mathbf{A}_\tau)^{-1}\} , \tag{3}$$

in which $\mathbf{A}_\tau$ is a sub-matrix of $\mathbf{A}$ that contains the columns corresponding to the indices in τ. Among all of the estimators that know the location of the taps, it can be shown that (as we will also show later in this correspondence) the efficient estimator will be the Structural Least Square Estimator (SLSE) which finds the solution of the following problem [18]:

$$\hat{\mathbf{s}}_\tau = \underset{\mathbf{s}_\tau}{\mathrm{argmin}} \|\mathbf{y} - \mathbf{A}_\tau \mathbf{s}_\tau\|_2^2 , \tag{4}$$

in which $\mathbf{s}_\tau$ is the $K \times 1$ vector of non-zero taps. The second bound, which is known as CRB-US [1], is the Cramér-Rao lower bound for the estimation problem in which the estimators have only prior knowledge about the cardinality of τ i.e. K, which indicates the degree of sparsity. It is obvious that the Cramér-Rao bound for this kind of estimation is not less than that of GAE, i.e.:

$$\mathrm{CRB\text{-}US} \geq \mathrm{CRB\text{-}S} . \tag{5}$$

---

[1]In the context of compressive sampling, the linear system in (1) is underdetermined, i.e. $M > N$. However, this assumption is not required in any of our presented analyses. Hence, our provided lemmas and theorems in this correspondence could be applied to the case of overdetermined noisy sparse recovery, which appears in many applications in communication theory, for example in sparse channel estimation [8].

Furthermore, in a recent work by Ben-Haim et al. [20] an expression for CRB-US has been stated. In fact, they have shown that the behaviour of the CRB differs depending on whether or not the unknown sparse vector has maximal support (i.e., $\|\mathbf{s}\|_0 = K$ or $\|\mathbf{s}\|_0 < K$). More accurately, they have shown that if the measurement matrix satisfies the uniqueness theorem provided by Donoho et al. [3] and Candés et al. [4], and if we consider the case of maximal support, i.e. when $\|\mathbf{s}\|_0 = K$ which is indeed our case in this correspondence, and if we consider the case of finite size sparse recovery, i.e. when M, N, and K are fixed and limited, the Cramér-Rao bound equals to that of GAE (when the sparsity pattern is known by the estimator), i.e. CRB-US equals CRB-S. However, according to our best knowledge, no evidence of exact achievability of CRB-US by the means of any practical estimator or non-practical estimators has been presented in the literature for the case of fixed and limited M, N, and K. So, if someone proposes an estimator that can achieve CRB-S instead of CRB-US while it has only prior knowledge about the sparsity degree, then it will be proven that CRB-S and CRB-US are equal to each other (as stated in [20]) and both of them are achievable by this proposed estimator.

Many efforts have been done to design an estimator with just the knowledge about the cardinality of τ that can achieve MSE as close as possible to the GAE Cramér-Rao lower bound (CRB-S). Candés et al. [19] and Haupt et al. [21] proposed estimators that can achieve CRB-S up to a factor of $\log M$ which is far from CRB-S. Interestingly, recent works done by Babadi et al. [1] and Akçakaya et al. [2] have shown that by using an impractical estimator known as "Typical Estimator", under certain constraints on $\mathbf{s}$ and $\mathbf{A}$, one can asymptotically achieve the Cramér-Rao bound of the GAE problem, i.e. CRB-S, without a priori knowing τ. By asymptotic, we mean where N, M and K tend to infinity while the size parameters in (2) remain constant. In other words, since the proposed typical estimator asymptotically achieves CRB-S, one can conclude that **a)** CRB-S and CRB-US are asymptotically equal, and **b)** this Cramér-Rao bound is achievable (note that in general, the Cramér-Rao bound of an estimation problem is not achievable, i.e. it is not generally a tight bound for MSE).

The typical estimation in [1], [2] is based on checking the *Joint Typicality* of the noisy observations vector with all possible choices of τ, and then decoding the one which is jointly typical with the observed $\mathbf{y}$. Definition of joint typicality is introduced in [1], [2] and we will review it later in this correspondence.[2] After detecting the support of $\mathbf{s}$, typical estimator estimates the unknown vector $\mathbf{s}$ by using a structural least square estimation method, i.e. it finds the solution of (4). In [1], the proof of the achievability of the Cramér-Rao bound by using the typical estimator is based on a lemma (Lemma 3.3 of [2]), which bounds the probability of two error events in the mentioned estimation process. The first of these probabilities is the probability of the event that the support of $\mathbf{s}$ is not jointly typical with $\mathbf{y}$ which we denote[3] by $(\tau \nsim \mathbf{y})$ and the second one is the probability of the event that a subset $J \subset \{1, 2, \ldots M\} \neq \tau$ with cardinality K is jointly typical with $\mathbf{y}$ which we denote by $(J \sim \mathbf{y})$. Using this lemma, [1] shows that if the average power of $\mathbf{s}$ is limited and if

$$\alpha < \frac{1}{9 + 4\log(\beta - 1)} \ , \tag{6}$$

then the joint typical estimator achieves the Cramér-Rao bound as

$N \to \infty$.

It is important to mention that the proof of the above mentioned statement in [1] depends on the assumption that the elements of the measurement matrix are drawn randomly from a Gaussian distribution, in addition to the assumption that this matrix is stochastic in the noise domain. By this, we mean that this assumption will impose the consideration of the elements of measurement matrix as random variables in our analysis, just like the elements of noise vector. On the contrary, these assumptions are unnecessary in the ordinary framework of compressed sensing, while we are looking to find a stable recovery method. In fact, it is common to use *non-stochastic* but *randomly generated* measurement matrices in this context, while assuming that the noise vector is stochastic (because the estimator knows the exact measurement matrix, but it is not aware of the noise vector). Additionally, among all randomly generated matrices, appropriate measurement matrices are those that satisfy a constraint called *The Concentration of Measures Inequality*[4], i.e. the following condition [24]:

$$\mathbb{P}\{| \ \|\mathbf{A}\mathbf{x}\|^2 - \|\mathbf{x}\|^2 | \geq \epsilon\|\mathbf{x}\|^2|\} \leq 2e^{-Nc_0(\epsilon)}, \ \epsilon \in (0,1) \qquad , \tag{7}$$

where the probability is taken over random space for $N \times M$ random-generated matrix $\mathbf{A}$, $\epsilon \in (0,1)$ is arbitrary, $c_0(\epsilon)$ is a constant depending only on ϵ and such that for all $\epsilon \in (0,1)$, $c_0(\epsilon) > 0$ and $\mathbf{x}$ is an arbitrary fixed vector in $\mathbb{R}^M$. Because of this mentioned difference in the assumptions made in [1] and ordinary assumptions made in the framework of compressed sensing, one may wonder that the results obtained in [1] may be also valid in the case of a larger family of measurement matrices than just the Gaussian matrices. Indeed, we will introduce a family of random-generated matrices which satisfies a modified version of concentration of measures inequality, i.e. the following condition:

$$\mathbb{P}\{| \ \|\mathbf{A}\mathbf{x}\|^2 - N\|\mathbf{x}\|^2 | \geq \epsilon N\|\mathbf{x}\|^2|\} \leq 2e^{-Nc_0(\epsilon)}, \ \epsilon \in (0,1) \qquad , \tag{8}$$

in which all the variables are the same as those in (7). Perhaps, the most prominent example of matrices that satisfy (8), are those with elements drawn independently and identically distributed according to $N(0,1)$ [24]; but, there is no force on having Gaussian entries in the measurement matrix. More precisely, one can also use matrices whose entries are independent realizations of ± 1 Bernoulli random variables

$$A_{i,j} = \begin{cases} +1 & \text{with probability } 1/2, \\ -1 & \text{with probability } 1/2. \end{cases} \tag{9}$$

or related distributions such as

$$A_{i,j} = \begin{cases} +\sqrt{3} & \text{with probability } \frac{1}{6}, \\ 0 & \text{with probability } \frac{2}{3}, \\ -\sqrt{3} & \text{with probability } \frac{1}{6}. \end{cases} \tag{10}$$

and yet these matrices satisfy (8). In addition to example random matrices described in (9) and (10), there are many other examples of random matrices that satisfy the condition in (8) and have an important role in statistical signal processing, communications[5], and in particular compressive sampling. In fact, there is a well known class of linear projections, mostly known as *data-base friendly* random projections [25], that satisfies the condition in (8), and at the same time can exploit the full allotment of dimensionality of

---

[2]It is worth noting that the concepts of typicality and typical estimation have been first introduced in the literature of Shannon's work on information theory [22], [23]. With some changes, this concept is adapted to the field of compressive sampling in [1], [2].

[3]We will use the notations "$\sim$" and "$\nsim$" for indicating a jointly typical or a non jointly typical pair in the rest of this correspondence.

[4]This condition is a preliminary condition for Restricted Isometry Property (RIP) which is a well-known sufficient condition in the area of compressed sensing for robust and stable recovery of the original sparse vector via l_1-minimization [6], [9].

[5]Many applications of using such non-Gaussian random projection, such as sparse channel estimation [8], have been reported in the literature.

a high-dimensional point set. Random i.i.d Gaussian matrices and those in (9) and (10) are considered as examples within this class. Hence, as satisfying (8) is a general property of commonly used random projection in signal processing and compressive sampling, it may be interesting to generalize the results obtained in [1] for this class of matrices. Then, one can conclude that the Cramér-Rao bound of the estimation is also asymptotically achievable by using the typical estimator introduced in [1] and [2], while we use non-Gaussian matrices that satisfy the condition depicted in (8), which is a common and general condition for measurement matrices in compressed sensing according to the literature.

In this correspondence, according to the above discussion, we investigate the results obtained in [1], and then we generalize the conditions for the problem of asymptotic achievability of Cramér-Rao bound in noisy compressed sensing. More accurately, by using an alternative approach to this problem comparing to the one used in [1] and [2], i.e. by assuming that the measurement matrix, $\mathbf{A}$, is not stochastic in the noise domain, we will find a lemma similar to Lemma 3.3 in [2] and prove it using a different method compared to the original one (by using Chernof tail bounds for probability [26]). Since Lemma 3.3 of [2] has been used as the main building block to obtain the results of [1], one wonders if those results (achievablity of CRB-S and asymptotic equivalence of CRB-S and CRB-US) may be incorrect under our new assumption ($\mathbf{A}$ is just generated at random, but it is deterministic when compared to noise) and hence if they should be revised. In this purpose, we first re-state our proved lemma in the case of randomly generated (but deterministic in noise domain) measurement matrices that satisfy (8). Subsequently, we see that the final obtained form have very minor differences from Lemma 3.3 in [2], while it is valid under the assumption that $\mathbf{A}$ is a deterministic randomly generated matrix. Finally, we re-study the results of [1] and see that although the main lemma used in [1] has been changed in our analysis, fortunately, all of the results in [1] remain valid. In other words, in noisy compressed sensing and under our modified version of concentration of measures inequality condition, the Cramér-Rao bound is asymptomatically achievable by using a typical estimator described in [1], and the constraint in (6) will also be valid without any changes.

This correspondence paper is organized as follows. In the next section, we will first review the definition of joint typicality and the typical estimator introduced in [1]. Moreover, the main theorem of [1] and the Lemma 3.3 in [2] will be re-studied. Indeed, we provide a new form of the mentioned lemma under our new assumptions, in which the measurement matrix is considered as a randomly generated matrix that satisfies (8), although is deterministic in the noise domain. In Section III, the Cramér-Rao lower bound on MSE for the compressed sensing problem in a noisy setting will be discussed and we will show that the results obtained in [1] remain valid under our generalized assumptions. So the Cramér-Rao bound of the GAE problem and that of the problem in which estimators have only prior knowledge about the degree of sparsity are asymptotically equal if the measurement matrix satisfies (8), although it may not be Gaussian or random in the noise domain. In all of the above discussions, we will use the model described in (1) and we will assume that the matrix $\mathbf{A}$ is randomly generated, but since it is *known* to the estimator, it should be treated as a deterministic matrix.

II. STATEMENT AND PROOF OF THE MAIN THEOREM

First, consider the noisy compressed sensing model in (1). As in [1], we use the following definition for joint typicality:

Definition 2.1 (Joint Typicality): Suppose that $\xi \subset \{1, 2, \ldots M\}$ and $|\xi| = K$, in which $|\cdot|$ denotes the cardinality of a set. Let $\mathbf{A}_\xi$ denote the $N \times K$ sub-matrix of $\mathbf{A}$ including those columns of $\mathbf{A}$ that

correspond to the indices in ξ. Let also $\Pi_{\mathbf{A}_\xi} \triangleq \mathbf{A}_\xi(\mathbf{A}_\xi^T \mathbf{A}_\xi)^{-1}\mathbf{A}_\xi^T$ and $\Pi_{\mathbf{A}_\xi}^\perp \triangleq \mathbf{I} - \Pi_{\mathbf{A}_\xi}$. ξ and $\mathbf{y}$ are said to be jointly typical with order ϵ, denoted by $(y \sim \xi)_\epsilon$, if and only if:

$$|\frac{1}{N}\|\Pi_{\mathbf{A}_\xi}^\perp \mathbf{y}\|^2 - \frac{N-K}{N}\sigma_n^2| < \epsilon . \tag{11}$$

In order to generalize the results in [1], we neglect the assumption that $\mathbf{A}$ is a Gaussian random matrix in the noise domain. Indeed, we assume that $\mathbf{A}$ is a randomly generated matrix, but is known to the estimator, and hence should be considered as a deterministic matrix. Accordingly, we first introduce the following theorem, which is similar to Lemma 3.3 in [2] and only depends on our new assumption on measurement matrix:

Theorem 2.1 (Bounds on the Probabilities of Typicality): Assume that in (1), $\tau = \text{supp}(\mathbf{s})$. Additionally, assume that $\xi \subset \{1, 2, \ldots M\}$ and $|\xi| = K$. Considering an arbitrary small enough $\epsilon > 0$, the following expressions hold as $N \to \infty$:

$$\mathbb{P}\{|\frac{1}{N}\|\Pi_{\mathbf{A}_\tau}^\perp \mathbf{y}\|^2 - \frac{N-K}{N}\sigma_n^2| > \epsilon\} \overset{\text{exponentially}}{\longrightarrow} 0 \ , \tag{12}$$

$$\mathbb{P}\{|\frac{1}{N}\|\Pi_{\mathbf{A}_\xi}^\perp \mathbf{y}\|^2 - \frac{N-K}{N}\sigma_n^2| < \epsilon\} \leq$$
$$\exp\left\{ -\frac{N-K}{4}\left[\frac{\frac{1}{N}\sum_{i\in\tau\backslash\xi}\sum_{j\in\tau\backslash\xi}s_i s_j \mathbf{a'}_i^T \mathbf{a'}_j - \epsilon}{\frac{2}{N}\sum_{i\in\tau\backslash\xi}\sum_{j\in\tau\backslash\xi}s_i s_j \mathbf{a'}_i^T \mathbf{a'}_j + \sigma_n'^2}\right]^2\right\} . \tag{13}$$

in which $\sigma_n'^2 = (1-\alpha)\sigma_n^2$ and $\mathbf{a'}_i = (\mathbf{U}_\xi \mathbf{a}_i)_{N-K}$, where $\mathbf{U}_\xi$ is a unitary matrix extracted from the eigenvalue decomposition of $\Pi_{\mathbf{A}_\xi}^\perp$, i.e. $\Pi_{\mathbf{A}_\xi}^\perp = \mathbf{U}_\xi \mathbf{D}\mathbf{U}_\xi^T$ and $\mathbf{D}$ is a diagonal matrix. The $(.)_m$ operator denotes a vector comprising of the first m elements of the operand.

Proof of (12) : The proof of this part is the same as the proof of the first part of Lemma 3.3 in [2] with some minor modifications. For the sake of readability, we will go through the steps of this proof. In these steps, we will try to find the PDF (Probability Density Function) of $\|\Pi_{\mathbf{A}_\tau}^\perp \mathbf{y}\|^2$ assuming that $\mathbf{A}$ is *known* and deterministic, while the noise vector is random.

Due to the fact that $\Pi_{\mathbf{A}_\tau}$ is the projector transform onto $\mathcal{S} = \text{span}\{\text{columns of } \mathbf{A}_\tau\}$ and since $\text{supp}(\mathbf{s}) = \tau$, we have:

$$\Pi_{\mathbf{A}_\tau}^\perp \mathbf{y} = \Pi_{\mathbf{A}_\tau}^\perp(\mathbf{A}_\tau \mathbf{s}_\tau + \mathbf{n}) = \mathbf{0} + \Pi_{\mathbf{A}_\tau}^\perp \mathbf{n} .$$

$\Pi_{\mathbf{A}_\tau}^\perp$ is a symmetric matrix, therefore we can decompose it as $\mathbf{U}_\tau \mathbf{D}\mathbf{U}_\tau^T$, in which $\mathbf{D}$ is a diagonal matrix and $\mathbf{U}_\tau$ is a unitary matrix ($\mathbf{U}_\tau \mathbf{U}_\tau^T = \mathbf{I}$). $\Pi_{\mathbf{A}_\tau}^\perp$ is an $N \times N$ matrix (which obviously has N eigenvalues). In addition to that, [1] shows that $\mathbf{A}_\tau$ is full-rank with probability 1. This means that $\mathcal{S} = \text{span}\{\text{columns of } \mathbf{A}_\tau\}$ is a K dimensional subspace of $\mathbb{R}^N$ as $N \to \infty$. Moreover, for every $\mathbf{y} \in \mathcal{S}$, we have $\Pi_{\mathbf{A}_\tau}^\perp \mathbf{y} = 0$ and so, the K basis vectors of $\mathcal{S}$ are K linearly independent eigenvectors of $\Pi_{\mathbf{A}_\tau}^\perp$ corresponding to the eigenvalue 0. Additionally, for every $\mathbf{y} \in \mathcal{S}' = \{\text{orthogonal compliment of } \mathcal{S} \text{ in } \mathbb{R}^N\}$, we have $\Pi_{\mathbf{A}_\tau}^\perp \mathbf{y} = \mathbf{y}$. In a similar way, we can show that $\mathcal{S}'$ is an $N-K$ dimensional subspace of $\mathbb{R}^N$ as $N \to \infty$ and so, the $N-K$ basis vectors of $\mathcal{S}'$ are the $N-K$ linearly independent eigenvectors of $\Pi_{\mathbf{A}_\tau}^\perp$ corresponding to the eigenvalue 1. Consequently, the main diagonal of $\mathbf{D}$ consists of $N-K$ 1's and K 0's. Moreover we have:

$$\|\Pi_{\mathbf{A}_\tau}^\perp \mathbf{y}\|^2 = \|\Pi_{\mathbf{A}_\tau}^\perp \mathbf{n}\|^2 = \|\mathbf{U}_\tau \mathbf{D}\mathbf{U}_\tau^T \mathbf{n}\|^2 =$$
$$\mathbf{n}^T \mathbf{U}_\tau \mathbf{D}\mathbf{U}_\tau^T \mathbf{U}_\tau \mathbf{D}\mathbf{U}_\tau^T \mathbf{n} =$$
$$(\mathbf{D}\mathbf{n'})^T \mathbf{D}\mathbf{n'} = \|\mathbf{D}\mathbf{n'}\|^2 , \tag{14}$$

in which $\mathbf{n'} = \mathbf{U}_\tau^T \mathbf{n}$ is a white Gaussian random vector (according to the fact that $\mathbf{U}_\tau^T$ is just a deterministic rotation transform), i.e.,

$\mathbf{n}' \sim N(0, \sigma_n^2 \mathbf{I})$. Without loss of generality, we can assume that the $N-K$ first elements of $\mathbf{D}$ are 1. So we can say that:

$$\varphi_1 = \|\mathbf{D}\mathbf{n}'\|^2 = |n'_1|^2 + |n'_2|^2 \cdots + |n'_{N-K}|^2 . \tag{15}$$

Since φ_1 is the sum of squares of $N-K$ independent Gaussian random variables with mean 0 and variance σ_n^2, it is a χ^2 random variable of order $N-K$ with parameter σ_n^2. i.e.,

$$\mathbb{E}\{\varphi_1\} = (N-K)\sigma_n^2 \quad \mathrm{var}\{\varphi_1\} = 2(N-K)\sigma_n^4 .$$

This χ^2 random variable has a moment generating function $\Phi_{\varphi_1}(s)$, which is defined by $\Phi_{\varphi_1}(s) \triangleq \mathbb{E}\{e^{\varphi_1 s}\}$, and for every s satisfying the condition $1 - 2s\sigma_n^2 > 0$ can be expressed as [27]:

$$\Phi_{\varphi_1}(s) = \frac{1}{(1 - 2s\sigma_n^2)^{\frac{N-K}{2}}} . \tag{16}$$

We can rewrite the probability in (12) as follows:

$$\mathbb{P}\{|\frac{1}{N}\|\Pi_{\mathbf{A}_\tau}^\perp \mathbf{y}\|^2 - \frac{N-K}{N}\sigma_n^2| > \epsilon\} =$$
$$\mathbb{P}\{|\varphi_1 - (N-K)\sigma_n^2| > N\epsilon\} \leq$$
$$\mathbb{P}\{\varphi_1 > N\epsilon + (N-K)\sigma_n^2\} + \mathbb{P}\{\varphi_1 < -N\epsilon + (N-K)\sigma_n^2\} . \tag{17}$$

So, by using Chernof bounds on the tail probability [26], i.e.:

$$\forall \nu > 0 : \mathbb{P}\{\varphi_1 > \delta\} \leq e^{-\nu\delta}\Phi_{\varphi_1}(\nu) \tag{18}$$
$$\forall \nu < 0 : \mathbb{P}\{\varphi_1 < \delta\} \leq e^{-\nu\delta}\Phi_{\varphi_1}(\nu) , \tag{19}$$

we can bound the probabilities in (48). By applying (18) and (16), and also considering the constraints needed for these equations, we have:

$$\forall\ 0 < \nu < \frac{1}{2\sigma_n^2} : \mathbb{P}\{\varphi_1 > N\epsilon + (N-K)\sigma_n^2\} \leq$$
$$\frac{1}{(1 - 2\nu\sigma_n^2)^{\frac{N-K}{2}}} \exp\{-\nu(N\epsilon + (N-K)\sigma_n^2)\} \triangleq f(\nu) . \tag{20}$$

By taking the derivative of $f(\nu)$ and finding its minimum in order to obtain the tightest bound, we find that this minimum occurs at $\nu^* = \frac{1}{2\sigma_n^2}\frac{N\epsilon}{N\epsilon + (N-K)\sigma_n^2}$. Moreover, it is easy to check that ν^* satisfies the constraints imposed by (18) and (16), i.e. $\nu^* > 0$ and $1 - 2\sigma_n^2\nu^* = \frac{(N-K)\sigma_n^2}{N\epsilon + (N-K)\sigma_n^2} > 0$. Hence we have:

$$\mathbb{P}\{\varphi_1 > N\epsilon + (N-K)\sigma_n^2\} \leq\ f(\nu^*) =$$
$$\left\{\frac{N\epsilon + (N-K)\sigma_n^2}{(N-K)\sigma_n^2}\right\}^{\frac{N-K}{2}} \exp\{-\frac{N\epsilon}{2\sigma_n^2}\} =$$
$$\exp\left\{\frac{(N-K)}{2}\ln(1 + \frac{N\epsilon}{(N-K)\sigma_n^2}) - \frac{N\epsilon}{2\sigma_n^2}\right\} =$$
$$\exp\left\{-\frac{(N-K)}{2}\left(-\ln(1 + \frac{\epsilon}{\sigma_n'^2}) + \frac{\epsilon}{\sigma_n'^2}\right)\right\} , \tag{21}$$

in which $\sigma_n'^2 = \frac{(N-K)}{N}\sigma_n^2$. Using the inequality $\ln(1 + \frac{\epsilon}{\sigma_n'^2}) \leq \frac{\epsilon}{\sigma_n'^2}$, we can say that the bound in (21) decreases exponentially to 0 as $N \to \infty$. Similarly, using (19) and following the same approach as in the proof of (21), we can bound $\mathbb{P}\{\varphi_1 < -N\epsilon + (N-K)\sigma_n^2\}$ and so we will have:

$$\mathbb{P}\{\varphi_1 < -N\epsilon + (N-K)\sigma_n^2\} \leq$$
$$\exp\left\{-\frac{(N-K)}{2}\left(-\ln(1 - \frac{\epsilon}{\sigma_n'^2}) - \frac{\epsilon}{\sigma_n'^2}\right)\right\} . \tag{22}$$

Using the inequality $\ln(1 - \frac{\epsilon}{\sigma_n'^2}) \leq -\frac{\epsilon}{\sigma_n'^2}$, it is seen that the bound in (22) approaches 0 exponentially as $N \to \infty$. Consequently the probability in (12) will tend at least exponentially to 0, and so the proof is complete. $\square$

Proof of (13): Similar to the previous part we have:

$$\Pi_{\mathbf{A}_\xi}^\perp \mathbf{y} = \Pi_{\mathbf{A}_\xi}^\perp (\mathbf{A}_\tau \mathbf{s}_\tau + \mathbf{n}) =$$
$$\Pi_{\mathbf{A}_\xi}^\perp (\sum_{i \in \tau \cap \xi} s_i \mathbf{a}_i + \sum_{i \in \tau \setminus \xi} s_i \mathbf{a}_i + \mathbf{n}) = \Pi_{\mathbf{A}_\xi}^\perp (\sum_{i \in \tau \setminus \xi} s_i \mathbf{a}_i + \mathbf{n}) .$$

In the same way, we can decompose $\Pi_{\mathbf{A}_\xi}^\perp = \mathbf{U}_\xi \mathbf{D} \mathbf{U}_\xi^T$, in which $\mathbf{D}$ is similar to the one in the previous part and $\mathbf{U}_\xi$ is a unitary matrix. Then we have:

$$\|\Pi_{\mathbf{A}_\xi}^\perp \mathbf{y}\|^2 = \|\mathbf{U}_\xi \mathbf{D} \mathbf{U}_\xi^T (\sum_{i \in \tau \setminus \xi} s_i \mathbf{a}_i + \mathbf{n})\|^2 = \|\mathbf{D}\mathbf{n}''\|^2 ,$$

in which, $\mathbf{n}'' = \mathbf{U}_\xi^T \mathbf{n} + \sum_{i \in \tau \setminus \xi} s_i \mathbf{U}_\xi^T \mathbf{a}_i$ is a Gaussian random vector with mean $\bar{\mathbf{n}}'' = \mathbb{E}\{\mathbf{n}''\} = \sum_{i \in \tau \setminus \xi} s_i \mathbf{U}_\xi^T \mathbf{a}_i$ and auto-covariance matrix $\mathbb{E}\{(\mathbf{n}'' - \bar{\mathbf{n}}'')(\mathbf{n}'' - \bar{\mathbf{n}}'')^T\} = C_{\mathbf{n}''} = \sigma_n^2 \mathbf{I}$, which are results of the fact that $\mathbf{A}$ is deterministic. It is important to note that the remaining proof of this part of Lemma 3.3 in [2] (which is so similar to our proposed lemma) is based on the Gaussian assumption on $\mathbf{A}$, in addition to the assumption that this matrix is random in the domain of noise[6]; nevertheless, our proof is free from such assumptions while we assume that the measurement matrix is deterministic. As a result, this assumption will help us to generalize our results for other types of randomly generated measurement matrices that are common in the compressed sensing area, as will be shown later in this correspondence.

To continue our proof, without loss of generality we can assume that the first $N-K$ elements of the main diagonal of $\mathbf{D}$ are 1 and so:

$$\varphi_2 = \|\mathbf{D}\mathbf{n}''\|^2 = |n_1''|^2 + |n_2''|^2 \cdots + |n_{N-K}''|^2 , \tag{23}$$

in which, $n_i'' \sim N(m_i, \sigma_n^2)$ for every $1 \leq i \leq N-K$. In addition to these, we have:

$$\sum_{i=1}^{N-K} m_i^2 = \|\sum_{i \in \tau \setminus \xi} (\mathbf{U}_\xi^T \mathbf{a}_i)_{N-K} s_i\|^2 = \sum_{i \in \tau \setminus \xi} \sum_{j \in \tau \setminus \xi} s_i s_j \mathbf{a'}_i^T \mathbf{a'}_j , \tag{24}$$

in which, the $(\cdot)_{N-K}$ operator denotes a sub-vector of the first $N-K$ elements and $\mathbf{a'}_i = (\mathbf{U}_\xi^T \mathbf{a}_i)_{N-K}$. For the sake of simplicity of notations, we define $\gamma^2 = \frac{1}{N}\sum_{i \in \tau \setminus \xi} \sum_{j \in \tau \setminus \xi} s_i s_j \mathbf{a'}_i^T \mathbf{a'}_j$. Now, the sum of the squares of $N-K$ independent Gaussian random variables n_i'', each having mean m_i, is a *non-central* χ^2 random variable of order $N-K$ with parameters σ_n^2 and $\sum_{i=1}^{N-K} m_i^2$. So we have:

$$\mathbb{E}\{\varphi_2\} = (N-K)\sigma_n^2 + N\gamma^2 , \quad \mathrm{var}\{\varphi_2\} = 2(N-K)\sigma_n^4 + 4\sigma_n^2 N^2 \gamma^4 .$$

Additionally, this χ^2 random variable has a moment generating function $\Phi_{\varphi_2}(s)$, which is defined by $\Phi_{\varphi_2}(s) \triangleq \mathbb{E}\{e^{\varphi_2 s}\}$, and for every s satisfying $1 - 2s\sigma_n^2 > 0$ can be expressed as [27]:

$$\Phi_{\varphi_2}(s) = \frac{1}{(1 - 2s\sigma_n^2)^{\frac{N-K}{2}}} \exp\left(\frac{s\sum_{i=1}^{N-K} m_i^2}{1 - 2s\sigma_n^2}\right) . \tag{25}$$

By *centralizing* the probability in (13) with respect to the mean of φ_2, we can rewrite the probability in (13) as:

$$\mathbb{P}\{|\frac{1}{N}\|\Pi_{\mathbf{A}_\xi}^\perp \mathbf{y}\|^2 - \frac{N-K}{N}\sigma_n^2| < \epsilon\} =$$
$$\mathbb{P}\{|\varphi_2 - (N-K)\sigma_n^2| < N\epsilon\} \leq \mathbb{P}\{\varphi_2 - (N-K)\sigma_n^2 < N\epsilon\} =$$
$$\mathbb{P}\{\varphi_2 - (N-K)\sigma_n^2 < N(\gamma^2 - \bar{\epsilon})\} =$$
$$\mathbb{P}\{\varphi_2 - (N-K)\sigma_n^2 - N\gamma^2 < -N\bar{\epsilon}\}, \tag{26}$$

in which $\bar{\epsilon} = \gamma^2 - \epsilon > 0$ (we assume that ϵ is small enough so that $\epsilon < \gamma^2$). Similar to the proof of (12), we will use Chernof

---

[6] This approach is very common in the framework of information theory, when one tries to show the achievability of a rate in a channel [22].

bounds stated in (18) and (19) to bound the probability in (26). More accurately, by the use of (19) we get:

$$\forall \nu < 0 : \ \mathbb{P}\{\varphi_2 < (N-K)\sigma_n^2 + N\gamma^2 - N\bar{\epsilon}\} \leq$$
$$\exp\left(-\nu[(N-K)\sigma_n^2 + N\gamma^2 - N\bar{\epsilon}]\right)\Phi_{\varphi_2}(\nu) \triangleq g(\nu) \ . \quad (27)$$

By plugging in the value of $\Phi_{\varphi_2}(\nu)$ from (25) for every ν satisfying $1 - 2\nu\sigma_n^2 > 0$, $g(\nu)$ is equal to:

$$g(\nu) = \frac{1}{(1 - 2\nu\sigma_n^2)^{\frac{N-K}{2}}} \times$$
$$\exp\left(\frac{\nu\sum_{i=1}^{N-K} m_i^2}{1 - 2\nu\sigma_n^2} - \nu[(N-K)\sigma_n^2 + N\gamma^2 - N\bar{\epsilon}]\right) \ . \quad (28)$$

As shown in the Appendix (Lemma A.1), by taking the derivative of $g(\nu)$ with respect to ν, one can see that this function will reach its minimum value at ν^*, calculated as following:

$$\nu^* = \frac{2\gamma^2 - 2\bar{\epsilon} + \sigma_n'^2 - \sqrt{(\sigma_n'^2 + 2\gamma^2)^2 - 4\gamma^2\bar{\epsilon}}}{4\sigma_n^2(\gamma^2 - \bar{\epsilon} + \sigma_n'^2)} \ , \quad (29)$$

in which $\sigma_n'^2 = \frac{N-K}{N}\sigma_n^2$. Moreover, this ν^* is negative (and hence satisfies the constraint $1 - 2\nu^*\sigma_n^2 > 0$), as stated by Lemma A.1. By plugging (29) and the expressions obtained in the Appendix for $1 - 2\nu^*\sigma_n^2$, $\frac{1}{1-2\nu^*\sigma_n^2}$ and $\frac{\gamma^2\nu^*}{1-2\nu^*\sigma_n^2}$ in (27), we will have the tightest Chernof bound for the probability in (27) as:

$$g(\nu^*) = \left[\frac{\sqrt{(\sigma_n'^2 + 2\gamma^2)^2 - 4\gamma^2\bar{\epsilon}} - \sigma_n'^2}{2\gamma^2}\right]^{\frac{N-K}{2}} \times$$
$$\exp\left\{-N\left[\frac{\sigma_n'^2 + 2\gamma^2 - \bar{\epsilon} - \sqrt{(\sigma_n'^2 + 2\gamma^2)^2 - 4\gamma^2\bar{\epsilon}}}{2\sigma_n^2}\right]\right\} \ . \quad (30)$$

After some manipulations, this bound can be re-written as:

$$\exp\left\{-\frac{N}{2}\left[\frac{\sigma_n'^2 + 2\gamma^2 - \bar{\epsilon} - (\sigma_n'^2 + 2\gamma^2)\sqrt{1 - \frac{4\gamma^2\bar{\epsilon}}{(\sigma_n'^2+2\gamma^2)^2}}}{\sigma_n^2}\right.\right.$$
$$\left.\left. - \frac{N-K}{N}\ln\left(\frac{(\sigma_n'^2 + 2\gamma^2)\sqrt{1 - \frac{4\gamma^2\bar{\epsilon}}{(\sigma_n'^2+2\gamma^2)^2}} - \sigma_n'^2}{2\gamma^2}\right)\right]\right\} \ . \quad (31)$$

Before proceeding any further, we will introduce the following lemma:

Lemma 2.2: for any $x \in \mathbb{R}$ and $0 \leq x \leq 1$ we have:
- $\sqrt{1-x} \ \leq 1 - \frac{1}{2}x$
- $\ln(1-x) \leq -x - \frac{1}{2}x^2$

the proof is elementary and is left to the reader.

It is important to note that $\frac{4\gamma^2\bar{\epsilon}}{(\sigma_n'^2+2\gamma^2)^2} = \frac{4\gamma^4 - 4\gamma^2\epsilon}{4\gamma^4 + \sigma_n'^4 + 2\gamma^2\sigma_n'^2}$, and so we have that $0 \leq \frac{4\gamma^2\bar{\epsilon}}{(\sigma_n'^2+2\gamma^2)^2} \leq 1$. Hence, by using the first part of Lemma (2.2) and some further manipulations, we can bound (30) as:

$$g(\nu^*) \leq \exp\left\{-\frac{N}{2}\left[\frac{\sigma_n'^2 + 2\gamma^2 - \bar{\epsilon} - (\sigma_n'^2 + 2\gamma^2) + \frac{2\gamma^2\bar{\epsilon}}{(\sigma_n'^2+2\gamma^2)}}{\sigma_n^2}\right.\right.$$
$$\left.\left. - \frac{N-K}{N}\ln\left(\frac{(\sigma_n'^2 + 2\gamma^2) - \frac{2\gamma^2\bar{\epsilon}}{(\sigma_n'^2+2\gamma^2)} - \sigma_n'^2}{2\gamma^2}\right)\right]\right\} \ . \quad (32)$$

After simplifying (32) we have:

$$g(\nu^*) \leq \exp\left\{-\frac{N}{2}\left[-\frac{\sigma_n'^2\bar{\epsilon}}{\sigma_n^2(\sigma_n'^2 + 2\gamma^2)}\right.\right.$$
$$\left.\left. - \frac{N-K}{N}\ln\left(1 - \frac{\bar{\epsilon}}{(\sigma_n'^2 + 2\gamma^2)}\right)\right]\right\} =$$
$$\exp\left\{-\frac{N-K}{2}\left[-\frac{\bar{\epsilon}}{(\sigma_n'^2 + 2\gamma^2)} - \ln\left(1 - \frac{\bar{\epsilon}}{(\sigma_n'^2 + 2\gamma^2)}\right)\right]\right\} \ . \quad (33)$$

It is also important to note that as $\frac{\bar{\epsilon}}{(\sigma_n'^2+2\gamma^2)} = \frac{\gamma^2-\epsilon}{(\sigma_n'^2+2\gamma^2)}$, we have $0 \leq \frac{\bar{\epsilon}}{(\sigma_n'^2+2\gamma^2)} \leq 1$. Now, by applying the second part of Lemma 2.2 we can make an upper bound for (33) as the following:

$$g(\nu^*) \leq \exp\left\{-\frac{N-K}{4}\left[\frac{\bar{\epsilon}}{(\sigma_n'^2 + 2\gamma^2)}\right]^2\right\} \ . \quad (34)$$

Therefore, according to (34), (27) and (26), we have finally come to the following result:

$$\mathbb{P}\{|\frac{1}{N}\|\Pi_{\mathbf{A}_\xi}^{\perp}\mathbf{y}\|^2 - \frac{N-K}{N}\sigma_n^2| < \epsilon\} \leq$$
$$\exp\left\{-\frac{N-K}{4}\left[\frac{\bar{\epsilon}}{(\sigma_n'^2 + 2\gamma^2)}\right]^2\right\} =$$
$$\exp\left\{-\frac{N-K}{4}\left[\frac{\frac{1}{N}\sum_{i\in\tau\setminus\xi}\sum_{j\in\tau\setminus\xi} s_i s_j \mathbf{a}_i'^T \mathbf{a}_j' - \epsilon}{\frac{2}{N}\sum_{i\in\tau\setminus\xi}\sum_{j\in\tau\setminus\xi} s_i s_j \mathbf{a}_i'^T \mathbf{a}_j' + \sigma_n'^2}\right]^2\right\} \ , \quad (35)$$

and this will complete the proof of (13). $\square$

It is important to note that if we see the proof of Theorem 2.1, then we will conclude that this theorem holds asymptotically in probability, i.e. if you test the validity of Theorem 2.1 for infinite numbers of randomly generated $\mathbf{A}$, then this theorem may not be valid for just finite numbers of $\mathbf{A}$. Moreover, as $N \to \infty$ the size of this finite set will tend to zero. Accordingly, one can say that as $N \to \infty$, Theorem 2.1 may not be valid for just asymptotic *zero* number of randomly generated $\mathbf{A}$, or simply it is asymptotically valid. However, as we will see later in Section III, we want to consider the achievability of Cramér-Rao bound in asymptotic case, and so this asymptotic validation should be enough.

In addition to what has been stated in Theorem 2.1, when the size of the problem tends to infinity and $\mathbf{A}$ satisfies the introduced concentration of measures inequality depicted in (8) (for instance, its elements are drawn i.i.d from $N(0,1)$ or distributions such as the ones introduced in (9) and (10)), one may find an equivalent bound using the following lemma:

Lemma 2.3: If the elements of $\mathbf{A}$ are randomly and independently generated according to a distribution that satisfies (8), then we have:

$$\sum_{i\in\tau\setminus\xi}\sum_{j\in\tau\setminus\xi} s_i s_j \mathbf{a}_i'^T \mathbf{a}_j' \to (N-K)\sum_{i\in\tau\setminus\xi}|s_i|^2 \ , \quad (36)$$

in which τ, ξ, $\mathbf{s}$ and $\mathbf{a}_i'$ are defined as in Theorem 2.1 .

Proof: Suppose that $\mathbf{x_1}$ and $\mathbf{x_2}$ are two arbitrary fixed vectors in $\mathbb{R}^M$. Then for every $\epsilon \in (0,1)$, the following inequalities hold with a probability that tends exponentially to 1 as N tends to ∞:

$$(1-\epsilon)N\|\mathbf{x}_1\|^2 \leq \|\mathbf{A}\mathbf{x}_1\|^2 \leq (1+\epsilon)N\|\mathbf{x}_1\|^2 \ , \quad (37)$$
$$(1-\epsilon)N\|\mathbf{x}_2\|^2 \leq \|\mathbf{A}\mathbf{x}_2\|^2 \leq (1+\epsilon)N\|\mathbf{x}_2\|^2 \ , \quad (38)$$
$$(1-\epsilon)N\|\mathbf{x}_1 - \mathbf{x}_2\|^2 \leq \|\mathbf{A}(\mathbf{x}_1 - \mathbf{x}_2)\|^2 \leq (1+\epsilon)N\|\mathbf{x_1} - \mathbf{x_2}\|^2 \ . \quad (39)$$

Using (37), (38) and (39), it is straightforward to show that:

$$(1+\epsilon)N\mathbf{x}_1^T\mathbf{x}_2 - \epsilon N(\|\mathbf{x}_1\|^2 + \|\mathbf{x}_2\|^2) \leq$$
$$(\mathbf{A}\mathbf{x}_1)^T(\mathbf{A}\mathbf{x}_2) \leq$$
$$(1-\epsilon)N\mathbf{x}_1^T\mathbf{x}_2 + \epsilon N(\|\mathbf{x}_1\|^2 + \|\mathbf{x}_2\|^2) \ . \quad (40)$$

By setting $\mathbf{x_1}$ and $\mathbf{x_2}$ with 1's in their i-th and j-th elements respectively and 0's in their other elements, for $i \neq j$ we will have:

$$-2\epsilon \leq \frac{1}{N}\mathbf{a}_i^T\mathbf{a}_j \leq 2\epsilon \ , \quad (41)$$

and in the case of $i = j$, we will have:

$$1 - \epsilon \leq \frac{1}{N}\mathbf{a}_i^T\mathbf{a}_i \leq 1 + \epsilon \ . \quad (42)$$

These events hold valid with a probability that tends exponentially to 1 as N tends to ∞ for a fix value of i and j. By applying the union bound on all $\binom{M}{2} = \frac{M(M-1)}{2}$ choices for i and j, if $\epsilon \to 0$ then the following equation holds for every i and j, with a probability that still tends to 1 as N increases:

$$\frac{1}{N}\mathbf{a}_i^T\mathbf{a}_j \to 0 \text{ if } i \neq j \text{ , } \frac{1}{N}\mathbf{a}_i^T\mathbf{a}_j \to 1 \text{ if } i = j \text{ .} \quad (43)$$

Now, consider the matrix $(\mathbf{U}_\xi^T\mathbf{A})_{N-K}$. We want to show this $(N-K)\times M$ matrix will also satisfy the modified version of the concentration of measures inequality. In other words, we want to show that for every $\mathbf{x} \in \mathbb{R}^M$, the following equation holds with a probability that tends exponentially to 1 as N tends ∞:

$$(1-\epsilon)(N-K)\|\mathbf{x}\|^2 \leq \|(\mathbf{U}_\xi^T\mathbf{A})_{N-K}\mathbf{x}\|^2 \leq (1+\epsilon)(N-K)\|\mathbf{x}\|^2 \text{ .} \quad (44)$$

To show this, we have:

$$\|(\mathbf{U}_\xi^T\mathbf{A})_{N-K}\mathbf{x}\|^2 = \mathbf{x}^T\mathbf{A}^T(\mathbf{U}_\xi^T)_{N-K}^T(\mathbf{U}_\xi^T)_{N-K}\mathbf{A}\mathbf{x} \text{ .} \quad (45)$$

To simplify (45), lets see how the matrix $\mathbf{U}_\xi$ is constructed. First, choose a set of indices in $\{1, 2, \ldots, M\}$ such as $\mathscr{L} \subset \{1, 2, ...M\}$, so that $|\mathscr{L}| = N$, and also $\xi \subset \mathscr{L}$ and $\tau \subset \mathscr{L}$. Then, we choose N columns of $\mathbf{A}_{\text{norm}} = \frac{1}{\sqrt{N}}\mathbf{A}$ corresponding to the indices in $\mathscr{L}$. Following (43), we can say that the columns of $\mathbf{A}_{\text{norm}}$ corresponding to the indices in ξ are an *approximate* orthonormal basis for the span of columns of $\mathbf{A}_\xi$ with a probability that tends exponentially to 1 as $N \to \infty$, and this approximation will become more accurate as ϵ is chosen smaller. Therefore, these columns can be considered asymptotically as approximations for the orthonormal eigenvectors of the symmetric matrix $\Pi_{\mathbf{A}_\xi}$ corresponding to zero eigenvalue, and again these approximations will become more accurate as ϵ is chosen smaller. Similarly, the columns of $\mathbf{A}_{\text{norm}}$ corresponding to the indices in $\mathscr{L}\backslash\xi$ can be considered as an approximate orthonormal basis for the kernel space of $\mathbf{A}_\xi$ with a probability that tends exponentially to 1, and so they are approximations for orthonormal eigenvectors of $\Pi_{\mathbf{A}_\xi}^\perp$ corresponding to the eigenvalue 1. Consequently, by the definition of $\mathbf{U}_\xi$ (i.e. its first $N-K$ columns are orthonormal eigenvectors of $\Pi_{\mathbf{A}_\xi}^\perp$ corresponding to eigenvalue 1 and the next K column are orthonormal eigenvectors of $\Pi_{\mathbf{A}_\xi}^\perp$ corresponding to zero eigenvalue) and the approximate orthogonal property of the selected columns of $\mathbf{A}$ as $N \to \infty$ (equation (43)), and by doing some simple manipulations, we have:

$$(\mathbf{U}_\xi^T)_{N-K}^T(\mathbf{U}_\xi^T)_{N-K} \approx \frac{N-K}{N}\mathbf{I}_{M\times M} \text{ ,} \quad (46)$$

and this approximation will become more accurate as ϵ is chosen smaller. By substituting the approximation stated in (46) with corresponding term in (45) we have:

$$\|(\mathbf{U}_\xi^T\mathbf{A})_{N-K}\mathbf{x}\|^2 \approx \frac{N-K}{N}\|\mathbf{A}\mathbf{x}\|^2 \text{ ,} \quad (47)$$

and again, this approximation will be more accurate with smaller ϵ. So, following (8), one can say that for small enough ϵ the equation (44) holds with a probability that tends exponentially to 1 as N grows [7]. Now, using (44) and similar to what we have stated about the columns of $\mathbf{A}$, we can conclude that the columns of $(\mathbf{U}_\xi^T\mathbf{A})_{N-K} = [\mathbf{a}'_1, \mathbf{a}'_2, \ldots, \mathbf{a}'_M]$ satisfy the following equation as $N \to \infty$:

$$\frac{1}{(N-K)}\mathbf{a}'^T_i\mathbf{a}'_j \to 0 \text{ if } i \neq j \text{ , } \frac{1}{(N-K)}\mathbf{a}'^T_i\mathbf{a}'_j \to 1 \text{ if } i = j \text{ .} \quad (48)$$

[7]Note that for small enough ϵ we require large enough N (following what has been stated in (8)), so that the concentration of measures inequality will be satisfied with high probability.

Substituting (48) in (24) will complete the proof. □

Now, using Lemma 2.3, we can rewrite the bound in (13) after some manipulations as:

$$\mathbb{P}\{|\frac{1}{N}\|\Pi_{\mathbf{A}_\xi}^\perp\mathbf{y}\|^2 - \frac{N-K}{N}\sigma_n^2| < \epsilon\} \leq$$
$$\exp\left\{-\frac{N-K}{4}\left[\frac{\sum_{i\in\tau\backslash\xi}|s_i|^2 - \epsilon'}{2\sum_{i\in\tau\backslash\xi}|s_i|^2 + \sigma_n^2}\right]^2\right\} \text{ ,} \quad (49)$$

in which $\epsilon' = \frac{N}{N-K}\epsilon$. Interestingly, the asymptotic bound obtained in (49) is very similar to the bound obtained in Lemma 3.3 of [2]. In fact, the bound obtained in [2] is as the following:

$$\mathbb{P}\{|\frac{1}{N}\|\Pi_{\mathbf{A}_\xi}^\perp\mathbf{y}\|^2 - \frac{N-K}{N}\sigma_n^2| < \epsilon\} \leq$$
$$\exp\left\{-\frac{N-K}{4}\left[\frac{\sum_{i\in\tau\backslash\xi}|s_i|^2 - \epsilon'}{\sum_{i\in\tau\backslash\xi}|s_i|^2 + \sigma_n^2}\right]^2\right\} \text{ .} \quad (50)$$

Although these bounds are not identical, but they are very similar.

III. OVERVIEW OF CRAMÉR-RAO LOWER BOUND AND THE JOINTLY TYPICAL ESTIMATOR

In this section, we will discuss the problem of estimating $\mathbf{s}$ from noisy observations. The estimation process has two phases. In the first phase, the estimator will detect $\tau = \text{supp}(\mathbf{s}) = \{i_1, i_2, \ldots i_K\}$ which is the location of the taps. The second phase includes estimating $\mathbf{s}_\tau = [s_{i_1}, s_{i_2}, \ldots, s_{i_K}]^T$ which is the value of the taps. In our discussion, we are going to survey the Cramér-Rao lower bound of the estimation problem. By using the idea of two-phase estimation, we consider two special kinds of estimation process in this work. In the first case, the estimator has a complete prior knowledge of τ, i.e. a genie has aided us with τ. In the second case, we have no prior knowledge of τ except for its cardinality, K, which shows the level of sparsity. We will then derive that these two bounds are asymptotically equal to each other and are achievable by typical estimation, as shown in [1] although the main theorem used in [1] has been changed.

The model in (1) can be rewritten as:

$$\mathbf{y} = \mathbf{A}\mathbf{s} + \mathbf{n} = \mathbf{A}_\tau\mathbf{s}_\tau + \mathbf{n} \text{ .} \quad (51)$$

Now, if the estimator knows τ and wants to estimate $\mathbf{s}_\tau$ from $\mathbf{y}$ and τ, then the Cramér-Rao bound of the estimation can be computed using the following theorem, stated in [1], [28]:

Theorem 3.1 (Cramér-Rao bound of genie aided estimation): Considering the model depicted in (1) and estimators of the form $f(\mathbf{y}, \tau) = \hat{\mathbf{s}}_\tau$, the Fisher information matrix of the GAE, which is defined as:

$$J_{\text{GAE}} = \mathbb{E}\{\left[\frac{\partial}{\partial\mathbf{s}_\tau}\log\mathbb{P}(\mathbf{y}|\mathbf{s}_\tau)\right]\left[\frac{\partial}{\partial\mathbf{s}_\tau}\log\mathbb{P}(\mathbf{y}|\mathbf{s}_\tau)\right]^T\} \text{ ,} \quad (52)$$

is equal to:

$$J_{\text{GAE}} = \frac{1}{\sigma_n^2}\mathbf{A}_\tau^T\mathbf{A}_\tau \text{ ,} \quad (53)$$

and so we have the following Cramér-Rao bound[8] for the estimator $\hat{\mathbf{s}}_\tau = f(\mathbf{y}, \tau)$:

$$\mathbb{E}\{(\mathbf{s}_\tau - \hat{\mathbf{s}}_\tau)(\mathbf{s}_\tau - \hat{\mathbf{s}}_\tau)^T\} \geq J^{-1} = \sigma_n^2(\mathbf{A}_\tau^T\mathbf{A}_\tau)^{-1} \quad (54)$$
$$\mathbb{E}\{\|\mathbf{s}_\tau - \hat{\mathbf{s}}_\tau\|^2\} \geq \sigma_n^2\text{Trace}[(\mathbf{A}_\tau^T\mathbf{A}_\tau)^{-1}] = \text{CRB-S} \text{ .} \quad (55)$$

Proof: The proof is given in [1] and [28]. □

In a GAE, by using a simple least square estimator for the model of (51) we can achieve the Cramér-Rao bound mentioned in (55), i.e. this estimator is *efficient*. In a more mathematical way, we have the following theorem:

[8]The equation $\mathbf{A} \geq \mathbf{B}$ means that $\mathbf{A} - \mathbf{B}$ is non-negative definite.

Theorem 3.2 (Structural Least Square Estimator (SLSE)): Consider the following genie aided estimator

$$\hat{\mathbf{s}}_\tau = f(\mathbf{y}, \tau) = \text{argmin} \ \|\mathbf{y} - \mathbf{A}_\tau \mathbf{x}_\tau\| = (\mathbf{A}_\tau^T \mathbf{A}_\tau)^{-1} \mathbf{A}_\tau^T \mathbf{y} \ ,$$

then we have:

$$\mathbb{E}\{\|\mathbf{s}_\tau - \hat{\mathbf{s}}_\tau\|^2\} = \sigma_n^2 \text{Trace}[(\mathbf{A}_\tau^T \mathbf{A}_\tau)^{-1}] \tag{56}$$

Proof: The proof is similar to the proof of achievability of the Cramér-Rao bound by the least square estimator where noise is Gaussian [17] and is omitted due to the lack of space. □
When considering the asymptotic case in the estimation process, one may use the equivalent limit of the bound in (55) using the following lemma:

Lemma 3.3: If the elements of $\mathbf{A}$ are generated independently and identically distributed according to a distribution that satisfies (8), considering the model in (51), we will have:

$$\text{CRB-S} \longrightarrow \alpha\sigma_n^2 = \frac{K}{N}\sigma_n^2 \ . \tag{57}$$

Proof: The proof of this lemma is given in [1] for the special case that elements of $\mathbf{A}$ are i.i.d Gaussian random variables. Generalization of this proof for the family of distributions that satisfy (8) is elementary and is left to the reader. □
Now, we are going to investigate the relation between CRB-S and CRB-US (which is Cramér-Rao bound of the estimators with just knowledge about the cardinality of τ) under the assumption that the measurement matrix, $\mathbf{A}$, is a randomly generated but deterministic matrix that satisfies our modified concentration of measures inequality described in (8). As was mentioned before, CRB-S and CRB-US seem to be different bounds at the first glance. But interestingly, as was shown in [1], in the asymptotic case they tend to each other. The proof of this statement in [1] is based on the Lemma 3.3 in [2] which is based on the Gaussianity of the measurement matrix and its randomness in the noise domain. So, one may wonder if the results in [1] are still correct under our new generalized assumptions, which fortunately is, as we will discuss later in this section. For showing this, we investigate the method of estimation in [1] which is based on a combinatorial search for finding the support of original sparse vector. Before proceeding any further, we will state the definition of this estimator as in [1]:

Definition 3.1 (Joint Typicality Estimator): The Joint Typicality Estimator finds a set of indices, $\zeta \subset \{1, 2, \dots M\}$ with cardinality of K which is jointly typical with $\mathbf{y}$ with order of ϵ. After that, it will produce the estimate $\hat{\mathbf{s}}_\zeta$ as:

$$(\mathbf{A}_\zeta^T \mathbf{A}_\zeta)^{-1} \mathbf{A}_\zeta^T \mathbf{y} \ . \tag{58}$$

If the estimator does not find a unique solution for ζ, it will return an all-zero vector as its output.

In the main theorem of [1], it is shown that under certain constraints, the MSE of the jointly typical estimator is upper bounded by $\alpha\sigma_n^2$. But the proof of this property is strongly based on Lemma 3.3 in [2], which cannot be used under our new assumptions, as was mentioned before. Instead, we use our variant of this lemma (Theorem 2.1 and especially its asymptotic form in (49)). According to the fact that this variant and the original form in [2] are not much different from each other, we can show that the main theorem in [1] remains valid without any necessary changes. More accurately, we have the following theorem:

Theorem 3.4 (Revised version of Main Theorem in [1]): Consider the model described in (51) and suppose that $\mathbf{A}$ is a randomly generated, but a deterministic matrix in the noise domain that satisfies (8). Let $\hat{\mathbf{s}}_\zeta$ be the output of the Jointly Typical Estimator defined in Definition 3.1. Additionally, let $\mu(\mathbf{s}) \triangleq \min_{i \in \tau} |s_i|$. If

- $\frac{K\mu^4(\mathbf{s})}{\log(K)} \to \infty$ as $N \to \infty$,
- $\|\mathbf{s}\|_2^2$ grows polynomially in N ,
- $\alpha < \frac{1}{9 + 4\log(\beta - 1)}$,

then we have:

$$\mathbb{E}\{\|\mathbf{s}_\tau - \hat{\mathbf{s}}_\zeta\|^2\} \le \alpha\sigma_n^2 \ , \tag{59}$$

as $N \to \infty$ for a fixed α and β.

Proof: Our proof, is exactly the same as the proof in [1] with some minor changes. First, similar to the mentioned proof, we try to upper bound the MSE of the estimation. Indeed, by repeating the first steps described by equations (17)-(22) of [1], applying the new form of the Lemma 3.3 in [2] which contains the bounds in (12) and (13) and also by using the asymptotic form of Theorem 2.1 described in (49), we can upper bound the MSE of joint typical estimator, i.e. $\mathbb{E}\{\|\mathbf{s}_\tau - \hat{\mathbf{s}}_\zeta\|^2\}$, by:

$$\alpha\sigma_n^{'2} + (K\sigma_n^{'2} + \|\mathbf{s}\|^2) \sum_{k'=1}^{K} \binom{K}{k'} \binom{M-K}{k'} \times$$
$$\exp\left\{ -\frac{N-K}{4}\left(\frac{k'\mu^2(\mathbf{s}) - \epsilon'}{2k'\mu^2(\mathbf{s}) + \sigma_n^2}\right)^2 \right\} \ . \tag{60}$$

Similar to [1], we use the inequality

$$\binom{K}{k'} \le \exp\left(k'\log(\frac{Ke}{k'})\right) \ , \tag{61}$$

to upper bound the k'-th term in the summation of (60) by:

$$\exp\left(K\frac{k'}{K}\log\left(\frac{e}{\frac{k'}{K}}\right) + K\frac{k'}{K}\log\left(\frac{(\beta-1)e}{\frac{k'}{K}}\right) - C_0 K\left(\frac{K\frac{k'}{K}\mu^2(\mathbf{s}) - \epsilon'}{2K\frac{k'}{K}\mu^2(\mathbf{s}) + \sigma_n^2}\right)^2\right) \ , \tag{62}$$

in which $C_0 \triangleq \frac{N-K}{4K}$. Again, similar to [1] we define:

$$f(z) \triangleq Kz\log\left(\frac{e}{z}\right) + Kz\log\left(\frac{(\beta-1)e}{z}\right) - C_0 K\left(\frac{Kz\mu^2(\mathbf{s}) - \epsilon'}{2Kz\mu^2(\mathbf{s}) + \sigma_n^2}\right)^2 \ . \tag{63}$$

Now, by Lemmas 3.4, 3.5 and 3.6 of [2] we can easily conclude that $f(z)$ attains its maximum at either $z = 1$ or $z = \frac{1}{K}$ if $\frac{K\mu^4(\mathbf{s})}{\log(K)} \to \infty$ as $N \to \infty$. So, we can upper bound (60) as:

$$\alpha\sigma_n^2 + (K\sigma_n^2 + \|\mathbf{s}\|^2) \sum_{k'=1}^{K} \exp\left\{\max\left\{f(1), f(\frac{1}{K})\right\}\right\} =$$
$$\alpha\sigma_n^2 + \exp\left\{\log(K^2\sigma_n^2 + K\|s\|^2) + \max\left\{f(1), f(\frac{1}{K})\right\}\right\} \ . \tag{64}$$

Additionally, we have:

$$f(1) = K(2 + \log(\beta - 1)) - C_0 K\left(\frac{K\mu^2(\mathbf{s}) - \epsilon'}{2K\mu^2(\mathbf{s}) + \sigma_n^2}\right)^2 \ , \tag{65}$$

and

$$f(\frac{1}{K}) = 2\log(K) + 2 + \log(\beta - 1) - C_0 K\left(\frac{\mu^2(\mathbf{s}) - \epsilon'}{2\mu^2(\mathbf{s}) + \sigma_n^2}\right)^2 \ . \tag{66}$$

It is obvious that $f(\frac{1}{K})$ grows linearly to $-\infty$ as $N \to \infty$. Additionally, if $C_0 > 2 + \log(\beta - 1)$ or equivalently $\alpha < \frac{1}{9 + 4\log(\beta - 1)}$ then $f(1)$ will also grow linearly to $-\infty$ as $N \to \infty$. Hence, the exponent of the second term in (64) tends to $-\infty$ as long as $\|\mathbf{s}\|^2$ grows polynomially with respect to N. So we have the following inequality when $N \to \infty$

$$\mathbb{E}\{\|\mathbf{s}_\tau - \hat{\mathbf{s}}_\zeta\|^2\} \le \alpha\sigma_n^2 \ , \tag{67}$$

which completes the proof. □

Now, by comparing the result of Theorem 3.4 with (57) and (5), we come to the conclusion that under the assumption we made about $\mathbf{A}$ (its distribution satisfies (8)), the CRB-S and CRB-US are asymptotically equal. Additionally, they can be asymptotically achieved using the Jointly Typical Estimator.

IV. Conclusion

In this correspondence paper, we examined the problem of the achievability of the Cramér-Rao bound in noisy compressed sensing under some new assumptions on the measurement matrix. Indeed, we relax our analysis from the Gaussianity constraint on the measurement matrix and its randomness in the domain of noise. Instead, we assumed that this matrix is randomly generated according to a distribution that satisfies some sort of concentration of measures inequality (described in (8)), but is deterministic in the noise domain. Mainly, we focused on the proof of Lemma 3.3 in [2] which was the main building block of the interesting results obtained in [1]. After reproving a new form of the above mentioned lemma using our new assumptions, we showed that the main theorem of [1] is still valid under these assumptions. So, the Cramér-Rao bound of the GAE and the Cramér-Rao bound for estimation with no prior knowledge about the original vector except for its degree of sparsity, are indeed asymptotically equal and the Jointly Typical Estimator first introduced in [1] can achieve this bound. Unfortunately, this method of estimation is impractical and to the best knowledge of the authors, the problem of finding a practical estimator that can achieve the Cramér-Rao bound is still open.

Appendix

Lemma A.1: The function $g(\nu)$ defined in (28) will reach its minimum at ν^* given in (29). Moreover, $\nu^* < 0$ and $1 - 2\sigma_n^2 \nu^* > 0$.

Proof: By taking the derivative of $g(\nu)$ with respect to ν and setting it to zero, we will have the following equation for finding the roots of $\frac{\partial g}{\partial \nu}$:

$$(-\bar{\epsilon} + \sigma_n'^2 + \gamma^2)X^2 - \sigma_n'^2 X - \gamma^2 = 0 \ ,$$

in which $X \triangleq (1 - 2\nu^* \sigma_n^2)$. By solving this equation with respect to X, we will have two solutions for X:

$$X_1 = \frac{\sigma_n'^2 + \sqrt{(\sigma_n'^2 + 2\gamma^2)^2 - 4\gamma^2\bar{\epsilon}}}{2(-\bar{\epsilon} + \sigma_n'^2 + \gamma^2)} \ ,$$

and

$$X_2 = \frac{\sigma_n'^2 - \sqrt{(\sigma_n'^2 + 2\gamma^2)^2 - 4\gamma^2\bar{\epsilon}}}{2(-\bar{\epsilon} + \sigma_n'^2 + \gamma^2)} \ .$$

First, it is important to note that both of these solutions are real, as by substituting $\gamma^2 - \bar{\epsilon}$ with ϵ we will have that $(\sigma_n'^2 + 2\gamma^2)^2 - 4\gamma^2\bar{\epsilon} = \sigma_n'^4 + 4\gamma^2\sigma_n'^2 + 4\gamma^2\epsilon > 0$. Furthermore, it also shows that $X_1 > 0$ and $X_2 < 0$. As we are looking for a ν^* that satisfies the constraint $1 - 2\nu^* \sigma_n^2 > 0$, the latter solution X_2 is not acceptable, and so we have $X = (1 - 2\nu^* \sigma_n^2) = X_1$. By taking the second derivative of $g(\nu)$ with respect to ν, it is easy to show that $\frac{\partial^2}{\partial \nu^2} g(\nu^*) \geq 0$ and so $g(\nu)$ will reach its minimum value at ν^*. It is important to note that the following expressions are also valid and can be extracted from the expression for X:

$$1 - 2\nu^* \sigma_n^2 = \frac{\sigma_n'^2 + \sqrt{(\sigma_n'^2 + 2\gamma^2)^2 - 4\gamma^2\bar{\epsilon}}}{-2(\bar{\epsilon} + \sigma_n'^2 + \gamma^2)} \tag{68}$$

$$\frac{1}{1 - 2\nu^* \sigma_n^2} = \frac{\sqrt{(\sigma_n'^2 + 2\gamma^2)^2 - 4\gamma^2\bar{\epsilon}} - \sigma_n'^2}{2\gamma^2} \tag{69}$$

$$\nu^* = \frac{2\gamma^2 - 2\bar{\epsilon} + \sigma_n'^2 - \sqrt{(\sigma_n'^2 + 2\gamma^2)^2 - 4\gamma^2\bar{\epsilon}}}{4\sigma_n^2(\gamma^2 - \bar{\epsilon} + \sigma_n'^2)} \tag{70}$$

$$\frac{\gamma^2 \nu^*}{1 - 2\nu^* \sigma_n^2} = \frac{\sqrt{(\sigma_n'^2 + 2\gamma^2)^2 - 4\gamma^2\bar{\epsilon}} - \sigma_n'^2 - 2\gamma^2}{4\sigma_n^2} \ . \tag{71}$$

By looking at (71), it is obvious that the nominator of the right hand side of this equation is a negative term (because $\sqrt{(\sigma_n'^2 + 2\gamma^2)^2 - 4\gamma^2\bar{\epsilon}} < \sigma_n'^2 + 2\gamma^2$), while the denominator, i.e. $4\sigma_n^2$, is a positive term. So, by using the fact that $1 - 2\sigma_n^2 \nu^* > 0$ (as we have proven before), one can concludes that $\nu^* < 0$, which completes the proof. □

References

[1] B. Babadi, N. Kalouptsidis, and V. Tarokh, "Asymptotic achievability of the Cramér–Rao bound for noisy compressive sampling," *IEEE Transactions on Signal Processing*, vol. 57, no. 3, pp. 1233–1236, 2009.

[2] M. Akçakaya and V. Tarokh, "Shannon theoretic limits on noisy compressive sampling," *IEEE Transactions on Information Theory*, vol. 56, no. 1, pp. 492–504, 2010.

[3] D. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[4] E. Candès, "Compressive sampling," in *Proceedings of the International Congress of Mathematicians*, vol. 3. Citeseer, 2006, p. 14331452.

[5] Y. Tsaig and D. Donoho, "Extensions of compressed sensing," *Signal processing*, vol. 86, no. 3, pp. 549–571, 2006.

[6] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on information theory*, vol. 52, no. 2, p. 489, 2006.

[7] R. Baraniuk, E. Candès, R. Nowak, and M. Vetterli, "Compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 12–13, 2008.

[8] R. Niazadeh, M. Babaie-Zadeh, and C. Jutten, "An alternating minimization method for sparse channel estimation," in *Ninth International Conference on Latent Variable Analysis and Signal Seperation (LVA-ICA, formerly known as ICA)*. Springer-Verlag, 2010, pp. 319–327.

[9] E. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006.

[10] J. Tropp and A. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 53, no. 12, p. 4655, 2007.

[11] D. Donoho, "For most large underdetermined systems of linear equations the minimal l1-norm solution is also the sparsest solution," *Communications on Pure and Applied Mathematics*, vol. 59, no. 6, pp. 797–829, 2006.

[12] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM review*, vol. 43, no. 1, pp. 129–159, 2001.

[13] T. Blumensath and M. Davies, "Gradient pursuits," *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2370–2382, 2008.

[14] I. Gorodnitsky and B. Rao, "Sparse signal reconstruction from limited data using FOCUSS: are-weighted minimum norm algorithm," *IEEE Transactions on Signal Processing*, vol. 45, no. 3, pp. 600–616, 1997.

[15] D. Needell and J. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301–321, 2009.

[16] H. Mohimani, M. Babaie-Zadeh, and C. Jutten, "A Fast Approach for Overcomplete Sparse Decomposition Based on Smoothed l0 Norm," *IEEE Transactions on Signal Processing*, vol. 57, no. 1, pp. 289–301, 2009.

[17] S. Kay, "Fundamentals of statistical signal processing: estimation theory," *Prentice-Hall Signal Processing Series*, 1993.

[18] C. Carbonelli, S. Vedantam, and U. Mitra, "Sparse channel estimation with zero tap detection," *IEEE Transactions on Wireless Communications*, vol. 6, no. 5, pp. 1743–1763, 2007.

[19] E. Candès and T. Tao, "The Dantzig selector: statistical estimation when p is much larger than n," *Annals of Statistics*, vol. 35, no. 6, pp. 2313–2351, 2007.

[20] Z. Ben-Haim and Y. Eldar, "The Cramér-Rao bound for estimating a sparse parameter vector," *Signal Processing, IEEE Transactions on*, vol. 58, no. 6, pp. 3384–3389, 2010.

[21] J. Haupt and R. Nowak, "Signal reconstruction from noisy random projections," *IEEE Transactions on Information Theory*, vol. 52, no. 9, pp. 4036–4048, 2006.

[22] T. Cover and J. Thomas, *Elements of information theory*. Wiley, 2006.

[23] C. Shannon, "A mathematical theory of communication," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 5, no. 1, p. 55, 2001.

[24] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, no. 3, pp. 253–263, 2008.

[25] D. Achlioptas, "Database-friendly random projections: Johnson-Lindenstrauss with binary coins," *Journal of Computer and System Sciences*, vol. 66, no. 4, pp. 671–687, 2003.

[26] T. Hagerup and C. Rub, "A guided tour of Chernoff bounds," *Proc. Lett*, vol. 33, pp. 305–308, 1989.

[27] A. Papoulis, S. Pillai, P. A, and P. SU, *Probability, random variables, and stochastic processes*. McGraw-Hill New York, 1965.

[28] J. Haupt, W. Bajwa, G. Raz, and R. Nowak, "Toeplitz compressed sensing matrices with applications to sparse channel estimation," *Information Theory, IEEE Transactions on*, vol. 56, no. 11, pp. 5862–5875, 2010.