
DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD
DASH.HARVARD.EDU


HARVARD LIBRARY
Office for Scholarly Communication

Asymptotic Achievability of the Cramér–Rao Bound For Noisy Compressive Sampling

Citation
Babadi, Behtash, Nicholas Kalouptsidis, and Vahid Tarokh. 2009. Asymptotic achievability of the Cramér–Rao Bound for noisy compressive sampling. IEEE Transactions on Signal Processing 57(3): 1233-1236.

Published Version
http://dx.doi.org/10.1109/TSP.2008.2010379

Permanent link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:2757546

Terms of Use
This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Open Access Policy Articles, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#OAP

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. Submit a story .

Accessibility

of the first summand. Taking (1) into account, we can distinguish two cases.

1) If H is a RKHS with reproducing kernel R, i.e., $H = H(R)$, then $H \otimes H$ is also a RKHS with reproducing kernel $R \otimes R$ and thus, (1) implies $K(s,t) = \sum_{i=1}^{\infty} \lambda_i \phi_i(s)\phi_i(t), \forall s,t \in I$. Moreover, the convergence of the preceding series is absolute in $s,t \in I$. Hence, (11) holds.
2) If H is a L_2-space, then (1) does not necessarily imply the pointwise convergence of the series $\sum_{i=1}^{\infty} \lambda_i \phi_i(s)\phi_i(t)$ to $K(s,t)$. For this assertion to hold it is necessary to impose additional conditions on the process $X(t)$. For example, if $\{X(t); t \in I\}$ is a smooth process[5] then, there exists a finite, non-negative measure μ on $(I, \mathcal{B}_I)$ ($\mathcal{B}_I$ is the σ-algebra of Lebesgue measurable subsets of I) which is equivalent to the Lebesgue measure and satisfies $\int_I K(t,t)d\mu(t) < \infty$ [12]. In this case, we set $H = L_2(I, \mathcal{B}_I, \mu)$. Moreover, it follows that $K(s,t) = \sum_{i=1}^{\infty} \lambda_i \phi_i(s)\phi_i(t), s,t \in I$, where the series converges absolutely in $s,t \in I$. Therefore, (11) follows.

ACKNOWLEDGMENT

The authors would like to thank Dr. D. Cárdenas for his helpful comments.

REFERENCES

[1] E. Parzen, "Extraction and detection problems and RKHSs," *J. SIAM Control*, vol. 1, no. 1, pp. 35–62, 1962.
[2] T. Kailath, "RKHS approach to detection and estimation problems—Part I: Deterministic signals in Gaussian noise," *IEEE, Trans. Inf. Theory*, vol. IT-17, no. 5, pp. 530–549, 1971.
[3] T. Kailath and H. L. Weinert, "An RKHS approach to detection and estimation problems—Part II: Gaussian signal detection," *IEEE, Trans. Inf. Theory*, vol. IT-21, no. 1, pp. 15–23, 1975.
[4] T. Kailath and D. Duttweiler, "An RKHS approach to detection and estimation problems—Part III: Generalized innovations representations and a likelihood-ratio formula," *IEEE, Trans. Inf. Theory*, vol. IT-18, no. 6, pp. 730–745, 1972.
[5] T. Kailath and V. Poor, "Detection of stochastic processes," *IEEE, Trans. Inf. Theory*, vol. 44, no. 6, pp. 2230–2259, 1998.
[6] N. Aronszajn, "Theory of reproducing kernels," *Amer. Math. Soc. Trans.*, vol. 68, pp. 337–404, May 1950.
[7] A. Berlinet and C. Thomas-Agnan, *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Norwell, MA: Kluwer, 2004.
[8] T. Kailath, R. T. Geesey, and H. L. Weinert, "Some relations among RKHS norms, Fredholm equations, and innovations representations," *IEEE, Trans. Inf. Theory*, vol. IT-18, no. 3, pp. 341–348, 1972.
[9] H. J. Weiner, "The gradient iteration in time series analysis," *Soc. Ind. Appl. Math. J.*, vol. 13, no. 4, pp. 1096–1101, 1965.
[10] M. Chen, Z. Chen, and G. Chen, *Approximate Solutions of Operator Equations*. Singapore: World Science, 1997.
[11] M. N. Lukic and J. H. Beder, "Stochastic processes with sample paths in reproducing kernel hilbert spaces," *Trans. Amer. Math. Soc.*, vol. 353, no. 10, pp. 3945–3969, 2001.
[12] S. Cambanis, "Representation of stochastic processes of second order and linear operations," *J. Math. Anal. Appl.*, vol. 41, no. 3, pp. 603–620, 1973.
[13] H. Ren, "Functional inverse regression and reproducing kernel Hilbert spaces," Ph.D. dissertation, Texas A&M Univ., College Station, TX, 2005.
[14] H. V. Poor, *An Introduction to Signal Detection and Estimation*, 2nd ed. New York: Springer-Verlag, 1994.
[15] K. K. Phoon, S. P. Huang, and S. T. Quek, "Implementation of Karhunen-Loève expansion for simulation using a wavelet-Galerkin scheme," *Probabilist. Eng. Mechan.*, vol. 17, no. 3, pp. 293–303, 2002.
[16] J. F. Traub and A. G. Werschulz, *Complexity and Information*. Oxford, U.K.: Oxford Univ. Press, 1998.
[17] R. Cools, "Advances in multidimensional integration," *J. Comput. Appl. Math.*, vol. 149, no. 1, pp. 1–12, 2002.
[18] D. Bini and L. Gemignani, "Iteration schemes for the divide-and-conquer eigenvalue solver," *Numer. Math.*, vol. 67, no. 4, pp. 403–425, 1994.
[19] R. J. Barton and H. V. Poor, "On generalized signal-to-noise ratios in quadratic detection," *Math. Control Signals Syst.*, vol. 5, no. 1, pp. 81–91, 1992.
[20] E. Parzen, "Probability density functionals and reproducing kernel Hilbert spaces," in *Proceedings of the Symposium on Time Series Analysis*, M. Rosenblatt, Ed. New York: Wiley, 1963, pp. 155–169.

[5]In the sense defined in [12]. For instance, a mean-square continuous process is smooth.

Asymptotic Achievability of the Cramér–Rao Bound for Noisy Compressive Sampling

Behtash Babadi, Nicholas Kalouptsidis, and Vahid Tarokh, *Fellow, IEEE*

***Abstract*—We consider a model of the form $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}$, where $\mathbf{x} \in \mathbb{C}^M$ is sparse with at most L nonzero coefficients in unknown locations, $\mathbf{y} \in \mathbb{C}^N$ is the observation vector, $\mathbf{A} \in \mathbb{C}^{N \times M}$ is the measurement matrix and $\mathbf{n} \in \mathbb{C}^N$ is the Gaussian noise. We develop a Cramér–Rao bound on the mean squared estimation error of the nonzero elements of $\mathbf{x}$, corresponding to the genie-aided estimator (GAE) which is provided with the locations of the nonzero elements of $\mathbf{x}$. Intuitively, the mean squared estimation error of any estimator without the knowledge of the locations of the nonzero elements of $\mathbf{x}$ is no less than that of the GAE. Assuming that L/N is fixed, we establish the existence of an estimator that asymptotically achieves the Cramér–Rao bound without any knowledge of the locations of the nonzero elements of $\mathbf{x}$ as $N \to \infty$, for $\mathbf{A}$ a random Gaussian matrix whose elements are drawn i.i.d. according to $\mathcal{N}(0,1)$.**

***Index Terms*—Compressive sampling, information theory, parameter estimation.**

I. INTRODUCTION

We consider the problem of estimating a sparse vector based on noisy observations. Suppose that we have a compressive sampling (Please see [2] and [5]) model of the form

$$\mathbf{y} = \mathbf{A}\mathbf{x}_{\text{true}} + \mathbf{n} \tag{1}$$

where $\mathbf{x}_{\text{true}} \in \mathbb{C}^M$ is the unknown sparse vector to be estimated, $\mathbf{y} \in \mathbb{C}^N$ is the observation vector, $\mathbf{n} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_N) \in \mathbb{C}^N$ is the Gaussian noise and $\mathbf{A} \in \mathbb{C}^{N \times M}$ is the measurement matrix. Suppose that $\mathbf{x}_{\text{true}}$ is sparse, i.e., $||\mathbf{x}_{\text{true}}||_0 = L < M$. Let $\mathcal{I} \triangleq \text{supp}(\mathbf{x}_{\text{true}})$ and

$$\alpha \triangleq L/N$$
$$\beta \triangleq M/L > 2$$

be fixed numbers.

The estimator must both estimate the locations and the values of the non-zero elements of $\mathbf{x}_{\text{true}}$. If a Genie provides us with $\mathcal{I}$, the problem

Manuscript received May 19, 2008; revised October 18, 2008. First published December 02, 2008; current version published February 13, 2009. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Athanasios P. Liavas.

B. Babadi and V. Tarokh are with the School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138 USA (e-mail: behtash@seas.harvard.edu; vahid@seas.harvard.edu).

N. Kalouptsidis is with the Department of Informatics and Telecommunications, National and Kapodistrian University of Athens, Athens 157 73, Greece (e-mail: kalou@di.uoa.gr).

Digital Object Identifier 10.1109/TSP.2008.2010379

1053-587X/$25.00 © 2009 IEEE

Authorized licensed use limited to: Harvard University SEAS. Downloaded on March 1, 2009 at 15:27 from IEEE Xplore. Restrictions apply.

reduces to estimating the values of the non-zero elements of $\mathbf{x}_{\text{true}}$. We denote the estimator to this reduced problem by genie-aided estimator (GAE).

Clearly, the mean-square estimation error (MSE) of any estimator is no less than that of the GAE (see [3]), since the GAE does not need to estimate the locations of the nonzero elements of $\mathbf{x}_{\text{true}}$ ($\log_2 \binom{M}{L} \doteq MH(1/\beta)$ bits, where $H(\cdot)$ is the binary entropy function).

Recently, Haupt and Nowak [6] and Candès and Tao [3] have proposed estimators which achieve the estimation error of the GAE up to a factor of $\log M$. In [6], a measurement matrix based on Rademacher projections is constructed and an iterative bound-optimization recovery procedure is proposed. Each step of the procedure requires $O(MN)$ operations and the iterations are repeated until convergence is achieved. It has been shown that the estimator achieves the estimation error of the GAE up to a factor of $\log M$.

Candès and Tao [3] have proposed an estimator based on linear programming, namely the Dantzig Selector, which achieves the estimation error of the GAE up to a factor of $\log M$, for Gaussian measurement matrices. The Dantzig Selector can be recast as a linear program and can be efficiently solved by the well-known primal-dual interior point methods, as suggested in [3]. Each iteration requires solving an $M \times M$ system of linear equations and the iterations are repeated until convergence is attained.

In this correspondence, we construct an estimator based on Shannon theory and the notion of typicality [4] that asymptotically achieves the Cramér–Rao bound on the estimation error of the GAE without the knowledge of the locations of the nonzero elements of $\mathbf{x}_{\text{true}}$, for Gaussian measurement matrices. Although the estimator presented in this correspondence has higher complexity (exponential) compared to the estimators in [3] and [6], to the best of our knowledge it is the first result establishing the achievability of the Cramér–Rao bound for noisy compressive sampling. The problem of finding efficient and low-complexity estimators that achieve the Cramér–Rao bound for noisy compressive sampling still remains open.

The outline of this correspondence follows next. In Section II, we state the main result of this correspondence and present its proof in Section III. We then discuss the implications of our results in Section IV.

II. Main Result

The main result of this correspondence is the following:

Main Theorem: In the compressive sampling model of $\mathbf{y} = \mathbf{A}\mathbf{x}_{\text{true}} + \mathbf{n}$, let $\mathbf{A}$ be a measurement matrix whose elements are i.i.d. and distributed according to $\mathcal{N}(0,1)$. Let $e_G(N)$ be the Cramér–Rao bound on the mean squared error of the GAE (averaged over $\mathbf{n}$ and $\mathbf{A}$), $\mu(\mathbf{x}) \triangleq \min_{i \in \mathcal{I}} |x_i|$ and α be a fixed number. If

- $L\mu^4(\mathbf{x}_{\text{true}})/\log L \to \infty$ as $N \to \infty$
- $\alpha < 1/(9 + 4\log(\beta - 1))$
- $\|\mathbf{x}_{\text{true}}\|^2$ grows polynomially in N,

assuming that the locations of the nonzero elements of $\mathbf{x}_{\text{true}}$ are not known, there exists an estimator (namely, the joint typicality estimator, given explicitly in this correspondence) for the nonzero elements of $\mathbf{x}_{\text{true}}$ with mean-square error $e_\delta(N)$, such that

$$\lim_{N \to \infty} |e_\delta(N) - e_G(N)| . = 0$$

III. Proof of the Main Theorem

In order to establish the Main Result, we need to define the Joint Typicality Estimator. We first state the following Lemma for random matrices:

Lemma 3.1: Let $\mathbf{A}$ be a measurement matrix whose elements are i.i.d. and distributed according to $\mathcal{N}(0,1)$, $\mathcal{J} \subset \{1, 2, \cdots, M\}$ such that $|\mathcal{J}| = L$ and $\mathbf{A}_{\mathcal{J}}$ be the submatrix of $\mathbf{A}$ with columns corresponding to the index set $\mathcal{J}$. Then, $\text{rank}(\mathbf{A}_{\mathcal{J}}) = L$ with probability 1.

Proof: Let $\mathbf{A}_i$ and $\mathbf{A}_j$ be any two columns of $\mathbf{A}_{\mathcal{J}}$. The law of large numbers implies

$$\mathbf{A}_i^* \mathbf{A}_j = \sum_k \mathbf{A}_{ik}^* \mathbf{A}_{jk} \to N\delta_{ij} \tag{2}$$

as $N \to \infty$. Thus, the columns of $\mathbf{A}_{\mathcal{J}}$ are mutually orthogonal with probability 1, which proves the statement of the Lemma. ∎

We can now define the notion of Joint Typicality. We adopt the definition from [1]:

Definition 3.2: We say an $N \times 1$ noisy observation vector, $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}$ and a set of indices $\mathcal{J} \subset \{1, 2, \cdots, M\}$, with $|\mathcal{J}| = L$, are δ-jointly typical if $\text{rank}(\mathbf{A}_{\mathcal{J}}) = L$ and

$$\left| \frac{1}{N} \left\| \Pi_{\mathbf{A}_{\mathcal{J}}}^{\perp} \mathbf{y} \right\|^2 - \frac{N - L}{N} \sigma^2 \right| < \delta \tag{3}$$

where $\mathbf{A}_{\mathcal{J}}$ is the submatrix of $\mathbf{A}$ with columns corresponding to the index set $\mathcal{J}$ and $\Pi_{\mathbf{A}_{\mathcal{J}}}^{\perp} \triangleq \mathbf{I} - \mathbf{A}_{\mathcal{J}}(\mathbf{A}_{\mathcal{J}}^* \mathbf{A}_{\mathcal{J}})^{-1} \mathbf{A}_{\mathcal{J}}^*$. We denote δ-jointly typicality of $\mathcal{J}$ with $\mathbf{y}$ by $\mathcal{J} \sim \mathbf{y}$.

Note that we can make the assumption of $\text{rank}(\mathbf{A}_{\mathcal{J}}) = L$ without loss of generality, based on Lemma 3.1.

Definition 3.3 (Joint Typicality Estimator): The Joint Typicality Estimator finds a set of indices $\hat{\mathcal{I}}$ which is δ-jointly typical with $\mathbf{y}$, by projecting $\mathbf{y}$ onto all the possible $\binom{M}{L}$ L-dimensional subspaces spanned by the columns of $\mathbf{A}$ and choosing the one satisfying (3). It then produces the estimate $\hat{\mathbf{x}}$ by projecting $\mathbf{y}$ onto the subspace spanned by $\mathbf{A}_{\hat{\mathcal{I}}}$:

$$\hat{\mathbf{x}} = \left(\mathbf{A}_{\hat{\mathcal{I}}}^* \mathbf{A}_{\hat{\mathcal{I}}}\right)^{-1} \mathbf{A}_{\hat{\mathcal{I}}}^* \mathbf{y}. \tag{4}$$

If the estimator does not find any set δ-typical to $\mathbf{y}$, it will output the zero vector as the estimate. We denote this event by E_0.

We show that the Joint Typicality Estimator has the property stated in the Main Theorem. In order to prove this claim, we first establish the following Lemmas:

Lemma 3.4: For any unbiased estimator $\hat{\mathbf{x}}$ of $\mathbf{x}_{\text{true}}$,

$$\mathbb{E}\left\{\|\hat{\mathbf{x}} - \mathbf{x}_{\text{true}}\|^2\right\} \geq e_G(N) \triangleq \sigma^2 \text{Tr}\left((\mathbf{A}_{\mathcal{I}}^* \mathbf{A}_{\mathcal{I}})^{-1}\right) \tag{5}$$

where $\mathbf{A}_{\mathcal{I}}$ is the submatrix of $\mathbf{A}$ with columns corresponding to the index set $\mathcal{I}$.

Proof: Assuming that a Genie provides us with $\mathcal{I}$, we have

$$p_{\mathbf{y}|\mathbf{x}}(\mathbf{y}; \mathbf{x}) = \frac{1}{(2\pi)^{N/2}\sigma^N} \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{y} - \mathbf{A}_{\mathcal{I}}\mathbf{x}_{\mathcal{I}}\|^2\right). \tag{6}$$

where $\mathbf{x}_{\mathcal{I}}$ is the subvector of $\mathbf{x}$ with elements corresponding to the index set $\mathcal{I}$. The Fisher information matrix is then given by

$$\mathbf{J}_{ij} \triangleq -\mathbb{E}\left\{\frac{\partial^2 \ln p_{\mathbf{y}|\mathbf{x}}(\mathbf{y}; \mathbf{x})}{\partial x_i \partial x_j}\right\} = \frac{1}{\sigma^2}(\mathbf{A}_{\mathcal{I}}^* \mathbf{A}_{\mathcal{I}})_{ij}. \tag{7}$$

Therefore, for any unbiased estimator $\hat{\mathbf{x}}$ by the Cramér–Rao bound [7],

$$\mathbb{E}\left\{\|\hat{\mathbf{x}} - \mathbf{x}_{\text{true}}\|^2\right\} \geq \text{Tr}(\mathbf{J}^{-1}) = \sigma^2 \text{Tr}\left((\mathbf{A}_{\mathcal{I}}^* \mathbf{A}_{\mathcal{I}})^{-1}\right). \tag{8}$$

∎

We note that Lemma 3.4 is also presented in [3].

Lemma 3.5: Let $\mathbf{A}$ be an $N \times L$ matrix whose elements are i.i.d. and distributed according to $\mathcal{N}(0,1)$. Then

$$\text{Tr}\left((\mathbf{A}^* \mathbf{A})^{-1}\right) \to \alpha. \tag{9}$$

Authorized licensed use limited to: Harvard University SEAS. Downloaded on March 1, 2009 at 15:27 from IEEE Xplore. Restrictions apply.

as $N \to \infty$ with $\alpha = L/N$ fixed.

Proof: Let $\mathbf{B} \triangleq \mathbf{A}^*\mathbf{A}$. We have

$$\mathbf{B}_{ij} = \sum_{k=1}^{N} \mathbf{A}_{ki}^* \mathbf{A}_{kj}. \tag{10}$$

Since the elements of $\mathbf{A}$ are i.i.d. and normally distributed, we can invoke the law of large numbers as follows:

$$\frac{1}{N}\mathbf{B}_{ij} = \frac{1}{N}\sum_{k=1}^{N} \mathbf{A}_{ki}^* \mathbf{A}_{ji} \to \mathbb{E}\left\{\mathbf{A}_{ki}^*\right\}\mathbb{E}\{\mathbf{A}_{kj}\} = 0 \tag{11}$$

if $i \neq j$ and

$$\frac{1}{N}\mathbf{B}_{ii} = \frac{1}{N}\sum_{k=1}^{N} \mathbf{A}_{ki}^* \mathbf{A}_{ki} \to \mathbb{E}\left\{|\mathbf{A}_{ki}|^2\right\} = 1 \tag{12}$$

as $N \to \infty$. Therefore

$$\frac{1}{N}\mathbf{B} = \frac{1}{N}\mathbf{A}^*\mathbf{A} \to \mathbf{I}_L. \tag{13}$$

We note that $1/N\mathbf{B}$ is non-singular with probability 1, since it converges to $\mathbf{I}_L$ element-wise with probability 1. Thus, the limit $\lim_{N\to\infty}(1/N\mathbf{B})^{-1}$ exists with probability 1. Hence

$$\left(\frac{1}{N}\mathbf{B}\right)^{-1} = \left(\frac{1}{N}\mathbf{A}^*\mathbf{A}\right)^{-1} \to \mathbf{I}_L \tag{14}$$

as $N \to \infty$. Taking the trace of both sides of the above equation, along with the linearity of the trace operator, proves the statement of the Lemma. ∎

Lemma 3.6 (Lemma 3.3 of [1]): For any $\delta > 0$,

$$\mathbb{P}\left(\left|\frac{1}{N}\left\|\Pi_{\mathbf{A}_\mathcal{I}}^{\perp}\mathbf{y}\right\|^2 - \frac{N-L}{N}\sigma^2\right| > \delta\right) \leq 2\exp\left(-\frac{\delta^2}{4\sigma^4}\frac{N^2}{N-L+\frac{2\delta}{\sigma^2}N}\right) \tag{15}$$

and

$$\mathbb{P}\left(\left|\frac{1}{N}\left\|\Pi_{\mathbf{A}_\mathcal{J}}^{\perp}\mathbf{y}\right\|^2 - \frac{N-L}{N}\sigma^2\right| < \delta\right) \leq \exp\left(-\frac{N-L}{4}\left(\frac{\sum_{k\in\mathcal{I}\setminus\mathcal{J}}|x_k|^2 - \delta'}{\sum_{k\in\mathcal{I}\setminus\mathcal{J}}|x_k|^2 + \sigma^2}\right)^2\right) \tag{16}$$

where $\mathcal{J}$ is an index set such that $|\mathcal{J}| = L$ and $|\mathcal{I}\cap\mathcal{J}| = K < L$ and $\operatorname{rank}(\mathbf{A}_\mathcal{J}) = L$ and

$$\delta' = \delta\frac{N}{N-L}.$$

Proof: Please refer to [1] for the proof. ∎

Proof of the Main Theorem: We can upper-bound the MSE of the Joint Typicality Estimator, averaged over all possible Gaussian measurement matrices, as follows:

$$\begin{aligned} e_\delta(N) &\triangleq \mathbb{E}_{\mathbf{n},\mathbf{A}}\left\{\|\hat{\mathbf{x}} - \mathbf{x}_{\text{true}}\|^2\right\} \\ &\leq \int_{\mathbf{A}} \|\mathbf{x}_{\text{true}}\|^2 \mathbb{P}(E_0)dP(\mathbf{A}) \\ &\quad + \int_{\mathbf{A}} \sum_{\mathcal{J}} \mathbb{E}_{\mathbf{n}|\mathbf{A}}\left\{\left\|\left(\mathbf{A}_\mathcal{J}^*\mathbf{A}_\mathcal{J}\right)^{-1}\mathbf{A}_\mathcal{J}^*\mathbf{y} - \mathbf{x}_{\text{true}}\right\|^2\right\} \\ &\qquad \times \mathbb{P}(\mathcal{J}\sim\mathbf{y})dP(\mathbf{A}) \end{aligned} \tag{17}$$

where $\mathbb{P}(\cdot)$ denotes the event probability defined over the noise density, $dP(\mathbf{A})$ denotes the Gaussian probability measure and the inequality follows from the union bound. Taking the term corresponding to $\mathcal{I}$ out of the summation, we can rewrite (17) as

$$\begin{aligned} e_\delta(N) &\leq \int_{\mathbf{A}} \|\mathbf{x}_{\text{true}}\|^2 \mathbb{P}(E_0)dP(\mathbf{A}) \\ &\quad + \int_{\mathbf{A}} \mathbb{E}_{\mathbf{n}|\mathbf{A}}\left\{\left\|\left(\mathbf{A}_\mathcal{I}^*\mathbf{A}_\mathcal{I}\right)^{-1}\mathbf{A}_\mathcal{I}^*\mathbf{y} - \mathbf{x}_{\text{true}}\right\|^2\right\} \\ &\qquad \times \mathbb{P}(\mathcal{I}\sim\mathbf{y})dP(\mathbf{A}) \\ &\quad + \int_{\mathbf{A}} \sum_{\mathcal{J}\neq\mathcal{I}} \mathbb{E}_{\mathbf{n}|\mathbf{A}}\left\{\left\|\left(\mathbf{A}_\mathcal{J}^*\mathbf{A}_\mathcal{J}\right)^{-1}\mathbf{A}_\mathcal{J}^*\mathbf{y} - \mathbf{x}_{\text{true}}\right\|^2\right\} \\ &\qquad \times \mathbb{P}(\mathcal{J}\sim\mathbf{y})dP(\mathbf{A}) \end{aligned} \tag{18}$$

The first term on the right-hand side of (18) can be upper-bounded as

$$\int_{\mathbf{A}} \|\mathbf{x}_{\text{true}}\|^2 \mathbb{P}(E_0)dP(\mathbf{A}) \leq 2\|\mathbf{x}_{\text{true}}\|^2 \exp\left(-\frac{\delta^2}{4\sigma^4}\frac{N^2}{N-L+\frac{2\delta}{\sigma^2}N}\right) \tag{19}$$

where the inequality follows from Lemma 3.6. This term clearly tends to zero as $N \to \infty$, since $\|\mathbf{x}_{\text{true}}\|^2$ grows polynomially in N by assumption.

Using Lemma 3.5, the second term on the right-hand side of (18) can be expressed as follows:

$$\begin{aligned} &\int_{\mathbf{A}} \mathbb{E}_{\mathbf{n}|\mathbf{A}}\left\{\left\|\left(\mathbf{A}_\mathcal{I}^*\mathbf{A}_\mathcal{I}\right)^{-1}\mathbf{A}_\mathcal{I}^*\mathbf{y} - \mathbf{x}_{\text{true}}\right\|^2\right\} \mathbb{P}(\mathcal{I}\sim\mathbf{y})dP(\mathbf{A}) \\ &\quad = \mathbb{E}_{\mathbf{n},\mathbf{A}}\left\{\left\|\left(\mathbf{A}_\mathcal{I}^*\mathbf{A}_\mathcal{I}\right)^{-1}\mathbf{A}_\mathcal{I}^*\mathbf{n}\right\|^2\right\} \\ &\quad = \mathbb{E}_{\mathbf{A}}\left\{\sigma^2 tr\left(\mathbf{A}_\mathcal{I}^*\mathbf{A}_\mathcal{I}\right)^{-1}\right\} \to \alpha\sigma^2. \end{aligned} \tag{20}$$

Finally, the third term on the right-hand side of (18) can be upper-bounded as

$$\begin{aligned} &\int_{\mathbf{A}} \sum_{\mathcal{J}\neq\mathcal{I}} \mathbb{E}_{\mathbf{n}|\mathbf{A}}\left\{\left\|\left(\mathbf{A}_\mathcal{J}^*\mathbf{A}_\mathcal{J}\right)^{-1}\mathbf{A}_\mathcal{J}^*\mathbf{y} - \mathbf{x}_{\text{true}}\right\|^2\right\} \\ &\qquad \times \mathbb{P}(\mathcal{J}\sim\mathbf{y})dP(\mathbf{A}) \\ &\quad \leq \left(L\sigma^2 + \|\mathbf{x}_{\text{true}}\|^2\right)\int_{\mathbf{A}} \sum_{\mathcal{J}\neq\mathcal{I}} \mathbb{P}(\mathcal{J}\sim\mathbf{y})dP(\mathbf{A}) \\ &\quad \leq \left(L\sigma^2 + \|\mathbf{x}_{\text{true}}\|^2\right) \\ &\qquad \times \sum_{\mathcal{J}\neq\mathcal{I}} \exp\left(-\frac{N-L}{4}\left(\frac{\sum_{k\in\mathcal{I}\setminus\mathcal{J}}|x_k|^2 - \delta'}{\sum_{k\in\mathcal{I}\setminus\mathcal{J}}|x_k|^2 + \sigma^2}\right)^2\right) \end{aligned} \tag{21}$$

where x_k denotes the kth component of $\mathbf{x}_{\text{true}}$. The first inequality follows from a trivial upper bound of $(L\sigma^2 + \|\mathbf{x}_{\text{true}}\|^2)$ on the error variance, and the second inequality follows from Lemma 3.6. The number of index sets $\mathcal{J}$ such that $|\mathcal{J}\cap\mathcal{I}| = K < L$ is upper-bounded by $\binom{L}{L-K}\binom{M-L}{L-K}$. Also, $\sum_{k\in\mathcal{I}\setminus\mathcal{J}}|x_k|^2 \geq (L-K)\mu^2(\mathbf{x}_{\text{true}})$. Therefore, we can rewrite the right-hand side of (21) as

$$\begin{aligned} &\sum_{K=0}^{L-1}\binom{L}{L-K}\binom{M-L}{L-K} \\ &\qquad \times \exp\left(-\frac{N-L}{4}\left(\frac{(L-K)\mu^2(\mathbf{x}_{\text{true}}) - \delta'}{(L-K)\mu^2(\mathbf{x}_{\text{true}}) + \sigma^2}\right)^2\right) \\ &\quad = \sum_{K'=1}^{L}\binom{L}{K'}\binom{M-L}{K'} \\ &\qquad \times \exp\left(-\frac{N-L}{4}\left(\frac{K'\mu^2(\mathbf{x}_{\text{true}}) - \delta'}{K'\mu^2(\mathbf{x}_{\text{true}}) + \sigma^2}\right)^2\right). \end{aligned} \tag{22}$$

Authorized licensed use limited to: Harvard University SEAS. Downloaded on March 1, 2009 at 15:27 from IEEE Xplore. Restrictions apply.

We use the inequality

$$\binom{L}{K'} \leq \exp\left(K' \log\left(\frac{Le}{K'}\right)\right) \tag{23}$$

in order to upper-bound the K'th term of the summation in (22) by

$$\exp\left(L\frac{K'}{L}\log\left(\frac{e}{\frac{K'}{L}}\right) + L\frac{K'}{L}\log\frac{(\beta-1)e}{\frac{K'}{L}} - C_0 L\left(\frac{L\frac{K'}{L}\mu^2(\mathbf{x}_{\text{true}}) - \delta'}{L\frac{K'}{L}\mu^2(\mathbf{x}_{\text{true}}) + \sigma^2}\right)^2\right) \tag{24}$$

where $C_0 \triangleq (N-L)/4L$. We define

$$f(z) \triangleq Lz\log\frac{e}{z} + Lz\log\frac{(\beta-1)e}{z} - C_0 L\left(\frac{Lz\mu^2(\mathbf{x}_{\text{true}}) - \delta'}{Lz\mu^2(\mathbf{x}_{\text{true}}) + \sigma^2}\right)^2. \tag{25}$$

By Lemmas 3.4, 3.5 and 3.6 of [1], it can be shown that $f(z)$ asymptotically attains its maximum at either $z = 1/L$ or $z = 1$, if $L\mu^4(\mathbf{x}_{\text{true}})/\log L \to \infty$ as $N \to \infty$. Thus, we can upper-bound the right-hand side of (21) by

$$\begin{aligned}&\left(L\sigma^2 + \|\mathbf{x}_{\text{true}}\|^2\right)\sum_{K'=1}^{L}\exp\left(\max\left\{f\left(\frac{1}{L}\right), f(1)\right\}\right)\\ &\quad = \exp\left(\log\left(L^2\sigma^2 + L\|\mathbf{x}_{\text{true}}\|^2\right) + \max\left\{f\left(\frac{1}{L}\right), f(1)\right\}\right).\end{aligned} \tag{26}$$

It is straightforward to obtain

$$f(1/L) = 2\log L + 2 + \log(\beta-1) - C_0 L\left(\frac{\mu^2(\mathbf{x}_{\text{true}}) - \delta'}{\mu^2(\mathbf{x}_{\text{true}}) + \sigma^2}\right)^2 \tag{27}$$

and

$$f(1) = L\left(2 + \log(\beta-1)\right) - C_0 L\left(\frac{L\mu^2(\mathbf{x}_{\text{true}}) - \delta'}{L\mu^2(\mathbf{x}_{\text{true}}) + \sigma^2}\right)^2. \tag{28}$$

Since $C_0 > 2 + \log(\beta-1)$ due to the assumption of $\alpha < 1/(9 + 4\log(\beta-1))$, both $f(1)$ and $f(1/L)$ will grow to $-\infty$ linearly as $N \to \infty$. Hence, the exponent in (26) will grow to $-\infty$ as $N \to \infty$, as long as $\|\mathbf{x}_{\text{true}}\|^2$ grows polynomially in N. Thus, the MSE of the Joint Typicality Estimator is asymptotically upper-bounded by

$$e_\delta(N) \leq \alpha\sigma^2. \tag{29}$$

By Lemma 3.5, we know that this is the asymptotic Cramér–Rao bound on the mean squared estimation error of the GAE, which proves the statement of the main theorem. ■

IV. Discussion of the Main Result

We consider the problem of estimating a sparse vector, $\mathbf{x}_{\text{true}} \in \mathbb{C}^M$, using noisy observations. Let $\|\mathbf{x}_{\text{true}}\|_0 = L < M$. The estimator must both estimate the locations and the values of the non-zero elements of $\mathbf{x}_{\text{true}}$. Clearly, the mean squared estimation error of any estimator is no less than that of the genie-aided estimator, for which the locations of the nonzero elements of $\mathbf{x}_{\text{true}}$ are known. We have constructed a Joint Typicality Estimator that asymptotically achieves the Cramér–Rao bound on the mean squared estimation error of the GAE without any knowledge of the locations of nonzero elements of $\mathbf{x}_{\text{true}}$, as $N \to \infty$ for $\alpha = L/N$ a fixed number.

References

[1] M. Akçakaya and V. Tarokh, "Shannon theoretic limits on noisy compressive sampling," *IEEE Trans. Inf. Theory* [Online]. Available: www.arxiv.org, submitted for publication

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

[3] E. Candès and T. Tao, "The Dantzig selector: Statistical estimation when p is much larger than n," *Ann. Statist.*, pp. 2313–2351, Dec. 2007.

[4] T. M. Cover and J. M. Thomas, *Elements of Information Theory*, 2nd ed. New York: Wiley, 2006.

[5] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[6] J. Haupt and R. Nowak, "Signal reconstruction from noisy random projections," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, Sep. 2006.

[7] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part I*, 1st ed. New York: Wiley, 2001.

Authorized licensed use limited to: Harvard University SEAS. Downloaded on March 1, 2009 at 15:27 from IEEE Xplore. Restrictions apply.