Support Recovery with Sparsely Sampled Free Random Matrices

Antonia Tulino
Universitá di Napoli, Federico II
Napoli, ITALY 80125
atulino@princeton.edu

Giuseppe Caire
University of Southern California
Los Angeles, CA 90089, USA
caire@usc.edu

Shlomo Shamai
Technion
Haifa, ISRAEL 32000
sshlomo@ee.technion.ac.il

Sergio Verdú
Princeton University
Princeton, NJ 08544, USA
verdu@princeton.edu

***Abstract*—Consider a Bernoulli-Gaussian complex n-vector whose components are X_iB_i, with B_i ∼Bernoulli-q and $X_i \sim \mathcal{CN}(0, \sigma^2)$, iid across i and mutually independent. This random q-sparse vector is multiplied by a random matrix U, and a randomly chosen subset of the components of average size np, $p \in [0, 1]$, of the resulting vector is then observed in additive Gaussian noise. We extend the scope of conventional noisy compressive sampling models where U is typically the identity or a matrix with iid components, to allow U that satisfies a certain freeness condition, which encompasses Haar matrices and other unitarily invariant matrices. We use the replica method and the decoupling principle of Guo and Verdú, as well as a number of information theoretic bounds, to study the input-output mutual information and the support recovery error rate as $n \to \infty$.**

***Index Terms*—Compressed Sensing, Random Matrices, Rate-Distortion Theory, Sparse Models, Support Recovery**

I. Introduction

A. Model Setup

Consider the n-dimensional complex-valued observation model:

$$\mathbf{y} = \mathbf{AUXb} + \mathbf{z} \tag{1}$$

where:

- $\mathbf{X} = \mathrm{diag}(\mathbf{x})$, and $\mathbf{x}$ is an iid complex Gaussian n-vector with components $X_i \sim \mathcal{CN}(0, \sigma^2)$;
- $\mathbf{b}$ is an iid n-vector with components B_i ∼Bernoulli-q, i.e., $\mathbb{P}[B_i = 1] = q = 1 - \mathbb{P}[B_i = 0]$;
- $\mathbf{A}$ is an $n \times n$ diagonal matrix with iid diagonal elements A_i ∼Bernoulli-p, i.e., $\mathbb{P}[A_i = 1] = p = 1 - \mathbb{P}[A_i = 0]$;
- $\mathbf{U}$ is an $n \times n$ random matrix such that
$$\mathbf{R} = \mathbf{U}^\dagger \mathbf{A}^\dagger \mathbf{A} \mathbf{U} \tag{2}$$
is free from any deterministic Hermitian matrix (see [1] and references therein).
- $\mathbf{z}$ is an iid complex Gaussian vector with components $Z_i \sim \mathcal{CN}(0, 1)$;
- $\mathbf{A}$, $\mathbf{U}$, $\mathbf{X}$, $\mathbf{b}$ and $\mathbf{z}$ are mutually independent.

The non-zero elements of $\mathbf{b}$ define the support of the Bernoulli-Gaussian vector $\mathbf{v} = \mathbf{Xb}$, which is sparse with "sparsity" (average fraction of non-zero elements) equal to q. The non-zero diagonal elements of $\mathbf{A}$ define the components of the product $\mathbf{UXb}$ for which a noisy observation is actually provided. The "sampling rate" (average fraction of observed components) is equal to p. The sensing matrix $\mathbf{AU}$ is known to the signal processor the goal of which, in this paper, is to detect the support of $\mathbf{v}$, i.e., to find the position of the non-zero components of $\mathbf{b}$. For later use, we also define $\mathbf{B} = \mathrm{diag}(\mathbf{b})$, such that $\mathbf{v}$ can be written equivalently as $\mathbf{Bx}$. When $\mathbf{U}$ consists of iid coefficients, (1) is the standard compressive sensing model [2], [3] in which the sensing space and the signal space are incoherent. Our more general setup, which, for example, encompasses Haar-distributed $\mathbf{U}$, retains that desirable property.

B. Overview

In this paper, we examine the mutual information $I(\mathbf{b}; \mathbf{y}|\mathbf{A}, \mathbf{U})$ of (1), referred to in the following as the *input-output mutual information*, and use it to analyze the support recovery given $\mathbf{y}$ and $\mathbf{A}, \mathbf{U}$. Denoting the recovered support by $\widehat{\mathbf{b}} = (\widehat{B}_1, \ldots, \widehat{B}_n)^\mathsf{T}$, with $\widehat{B}_i \in \{0, 1\}$, the objective is to minimize the support recovery error rate, defined by

$$D(p, q, \sigma^2) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \mathbb{P}[B_i \neq \widehat{B}_i] \tag{3}$$

where the expectation is with respect to $\mathbf{A}$, $\mathbf{U}$, $\mathbf{X}$, $\mathbf{b}$, and $\mathbf{z}$. We refer to the performance criterion in (3) as *support estimation Hamming distortion*.

The information theoretic analysis of the fundamental limits of *exact support recovery* (one misclassified component is equivalent to many) was initiated in [4], which gave a lower bound on the number of required observations by drawing a parallel with reliable information transmission through a Gaussian channel. Fano's inequality was then used in [5] to bound the probability of exact support recovery, This requires imposing a lower bound on the magnitude of the nonzero coefficients of the sparse vector $\mathbf{v} = \mathbf{Xb}$, and assumes that the cardinality of the support is known. Such assumptions are not needed in our case since the performance metric (3) allows for partial recovery under the Hamming distortion (see also [6]). An information-theoretic approach to finding bounds on the sampling rate required for recovery of the sparse signal under various performance metrics is given in [7].

We provide both bounds and exact expressions for the input-output mutual information rate, as well as bounds and an expression for the *optimal* support estimation Hamming distortion. Our exact expressions (as opposed to bounds) rely on replica-method techniques, previously applied in various

978-1-4577-0594-6/11/$26.00 ©2011 IEEE

problems involving iid matrices, e.g. [8], [10], [9], [11]. Of particular relevance to our work is the *decoupling principle*, pioneered in [8] in the setting of iid matrices. Unitarily invariant sensor matrices have also been addressed in [12], with focus on noiseless L_p reconstructions of the signal, rather than the estimation of its support. The recent work in [13] gives results for iid Gaussian sensor matrices, based on the analysis of a message passing algorithm rather than the replica method. A full rigorization of the decoupling principle for compressive sensing applications with iid $\mathbf{U}$ has been recently announced in [14].

Section II states the main results on the input-output mutual information. Upper bounds on the mutual information are developed in Section III. Section IV extends the decoupling principle [8] to the model in (1) and provides the analysis of the optimal symbol-by-symbol MAP estimator for the support. Proofs and technical details are omitted because of space limitation, and can be found in [15].

II. Main Results

Our main results concern the mutual information rate

$$\mathcal{I} \triangleq \lim_{n\to\infty} \frac{1}{n} I(\mathbf{b};\mathbf{y}|\mathbf{A},\mathbf{U}) = \mathcal{I}_1 - \mathcal{I}_2 \tag{4}$$

where

$$\mathcal{I}_1 \triangleq \lim_{n\to\infty} \frac{1}{n} I(\mathbf{v};\mathbf{y}|\mathbf{A},\mathbf{U}) \tag{5}$$

$$\mathcal{I}_2 \triangleq \lim_{n\to\infty} \frac{1}{n} I(\mathbf{x};\mathbf{y}|\mathbf{A},\mathbf{U},\mathbf{b}). \tag{6}$$

and right-most equality in 4) follows from

$$I(\mathbf{b};\mathbf{y}|\mathbf{A},\mathbf{U}) = I(\mathbf{x},\mathbf{b};\mathbf{y}|\mathbf{A},\mathbf{U}) - I(\mathbf{x};\mathbf{y}|\mathbf{A},\mathbf{U},\mathbf{b}), \tag{7}$$

using the fact that, from model (1), $I(\mathbf{x},\mathbf{b};\mathbf{y}|\mathbf{A},\mathbf{U}) = I(\mathbf{X}\mathbf{b};\mathbf{y}|\mathbf{A},\mathbf{U}) = I(\mathbf{v};\mathbf{y}|\mathbf{A},\mathbf{U})$.

A. Error rate lower bound via mutual information

We can bound the minimal support estimation Hamming distortion $D(p,q,\sigma^2)$ defined in (3) in terms of $\mathcal{I}$ using the following simple result.

Theorem 1 *Given a joint distribution $P_{XY} = P_X P_{Y|X}$ on $\mathcal{X}\times\mathcal{Y}$, a reconstruction alphabet $\widehat{\mathcal{X}}$ and a distortion measure $\mathsf{d}:\mathcal{X}\times\widehat{\mathcal{X}}\mapsto[0,\infty)$, let*

$$R(d) \triangleq \inf_{P_{\widehat{X}|X}:\mathbb{E}[\mathsf{d}(X,\widehat{X})]\le d} I(X;\widehat{X}) \tag{8}$$

Then

$$R(\inf \mathbb{E}[\mathsf{d}(X,\widehat{X})]) \le I(X;Y) \tag{9}$$

where the infimum is over all $P_{\widehat{X}|Y}:\mathcal{Y}\to\widehat{\mathcal{X}}$ and $X \leftrightarrow Y \leftrightarrow \widehat{X}$.

Since $R(d)$ is a monotonically decreasing function, (9) gives an information theoretic lower bound on the non-information-theoretic quantity $\inf \mathbb{E}[\mathsf{d}(X,\widehat{X})]$. In our case, using the rate-distortion function of a Bernoulli-q source with Hamming distortion, given by $R(d) = \max\{h(q)-h(d),0\}$, Theorem 1 results in

$$D(p,q,\sigma^2) \ge h^{-1}(h(q)-\mathcal{I}) \tag{10}$$

where $h(x) = x\log\frac{1}{x} + (1-x)\log\frac{1}{1-x},\ x\in[0,1]$ denotes the binary entropy function and where we assume $q\le\frac{1}{2}$ (notice that $\mathcal{I}\le h(q)$ by definition (4)).

B. Mutual information rate $\mathcal{I}_1$ via replica method

For any $(X,Y)\sim P_{XY}$, we denote the minimum mean-square error for estimating X from Y as

$$\mathsf{mmse}(X|Y) \triangleq \mathbb{E}[|X-\mathbb{E}[X|Y]|^2]. \tag{11}$$

With this definition, we have:

Theorem 2 *Let B_0, X_0, Z be mutually independent random variables, with $\mathbb{P}[B_0=1] = q = 1-\mathbb{P}[B_0=0]$, $X_0\sim\mathcal{CN}(0,\sigma^2)$, and $Z\sim\mathcal{CN}(0,1)$, and define $V_0 = B_0X_0$. Let $\mathcal{R}_{\mathbf{R}}(\cdot)$ denote the R-transform [1] of the random matrix $\mathbf{R}$ defined in (2). Then,*

$$\mathcal{I}_1 = I\left(V_0; V_0 + \frac{1}{\sqrt{\eta}}Z\right) + \log(e)\int_0^{\chi}(\mathcal{R}_{\mathbf{R}}(-w)-\eta)\,dw \tag{12}$$

where η and χ are the non-negative solutions of the system of equations:

$$\eta = \mathcal{R}_{\mathbf{R}}(-\chi) \tag{13a}$$

$$\chi = \mathsf{mmse}\left(V_0|V_0+\frac{1}{\sqrt{\eta}}Z\right) \tag{13b}$$

C. Mutual information rate $\mathcal{I}_2$ via freeness

Theorem 3 *Let $\mathcal{V}_{\mathbf{R}}(\cdot)$ and $\eta_{\mathbf{R}}(\cdot)$ denote the Shannon transform and η-transform [1] of $\mathbf{R}$ defined in (2). Then,*

$$\mathcal{I}_2 = \mathcal{V}_{\mathbf{R}}(\alpha\sigma^2) + q\log\left(1+\nu\sigma^2\right) - \log(1+\alpha\nu\sigma^2) \tag{14}$$

where α,ν are the unique non-negative solutions of the system of equations

$$\eta_{\mathbf{R}}(\alpha\sigma^2) = \frac{1}{1+\alpha\nu\sigma^2} = \frac{q}{1+\nu\sigma^2} + (1-q) \tag{15}$$

D. Special Cases

1) $\mathbf{U}$ is an iid random matrix: Assuming $\mathbf{U}$ with iid entries with mean zero and variance $\frac{1}{n}$, we can replace $\mathbf{AU}$ with $\sqrt{p}\mathbf{H}$, where $\mathbf{H}$ is $np\times n$ iid elements with mean zero and variance $1/(np)$. This reduces $\mathcal{I}_1$ to the case studied in [8], with scaled Bernoulli-Gaussian input $\sqrt{p}\mathbf{v}$.

The R-transform of of $\mathbf{R} = \mathbf{H}^\dagger\mathbf{H}$ in Theorem 2 is given by:

$$\mathcal{R}_{\mathbf{R}}(z) = \frac{p}{p-z}. \tag{16}$$

The fixed point equations (13a) and (13b) reduce to

$$\frac{1}{\eta} = 1 + \mathsf{mmse}\left(V_0|V_0+\frac{1}{\sqrt{p\eta}}Z\right), \tag{17}$$

and (12) takes on the form

$$\mathcal{I}_1 = I\left(V_0; V_0 + \frac{1}{\sqrt{p\eta}}Z\right) + p\log(e)\left(\log_e\left(\frac{1}{\eta}\right) + \eta - 1\right) \tag{18}$$

which coincides with the form found in [8]. Furthermore, $\mathcal{I}_2$ can be immediately calculated by noticing that $\mathbf{y} = \sqrt{p}\mathbf{HBx}+\mathbf{z}$, after eliminating the zero columns of the product $\mathbf{HB}$, is asymptotically equivalent to $\mathbf{y} = \sqrt{p}\widetilde{\mathbf{H}}\widetilde{\mathbf{x}} + \mathbf{z}$, where $\widetilde{\mathbf{H}}$ has dimension $np \times nq$, iid elements with mean zero and variance $1/(np)$, and $\widetilde{\mathbf{x}}$ is an iid nq-vector with elements $\sim \mathcal{CN}(0, \sigma^2)$. Then, $\mathcal{I}_2$ takes on the well-known form [1] of the Shannon transform of the Marčenko-Pastur law, scaled by p (since in the definition of $\mathcal{I}_2$ we normalize by n):

$$\begin{aligned}\mathcal{I}_2 \;=\; & p\mathcal{V}_{\widetilde{\mathbf{H}}\widetilde{\mathbf{H}}^\dagger}(p\sigma^2) = q\log\left(1 + p\sigma^2 - \frac{\mathcal{F}}{4}\right) + \\ & +p\log\left(1 + q\sigma^2 - \frac{\mathcal{F}}{4}\right) - \frac{\mathcal{F}}{4\sigma^2}\log(e)\end{aligned} \tag{19}$$

with

$$\mathcal{F} = \left(\sqrt{\sigma^2\left(\sqrt{p}+\sqrt{q}\right)^2 + 1} - \sqrt{\sigma^2\left(\sqrt{p}-\sqrt{q}\right)^2 + 1}\right)^2.$$

2) $\mathbf{U}$ is Haar-distributed: If $\mathbf{U}$ is uniformly distributed on the manifold of $n \times n$ unitary matrices, the eigenvalue distribution of $\mathbf{R}$ coincides with that of $\mathbf{AA}^\dagger = \mathbf{A}$, so that $|\mathsf{R}|^2$ ∼Bernoulli-q, and

$$\mathcal{R}_{\mathbf{R}}(z) = \mathcal{R}_{\mathbf{A}}(z) = \frac{-1 + z + \sqrt{4zp + (1-z)^2}}{2z} \tag{20}$$

This allows for an easy calculation of (12) with the corresponding fixed point equations (13a) and (13b). Furthermore, the following closed-form expression holds [16]

$$\mathcal{I}_2 = q\log\left(1 + \hat{p}\sigma^2\right) + d(p||\hat{p}) \tag{21}$$

with

$$\begin{aligned}\hat{p} \;=\; & \frac{(p-q)\sigma^2 - 1}{2(1-q)\sigma^2} \\ + \; & \frac{\sqrt{((2-p-q)\sigma^2+1)^2 - 4(1-p)(1-q)(1+\sigma^2)\sigma^2}}{2(1-q)\sigma^2},\end{aligned} \tag{22}$$

and

$$d(a||b) = a\log\frac{a}{b} + (1-a)\log\frac{1-a}{1-b}. \tag{23}$$

3) $\mathbf{A} = \mathbf{I}$, unitary $\mathbf{U}$: In this case, $\mathbf{R} = \mathbf{U}^\dagger\mathbf{A}^\dagger\mathbf{AU} = \mathbf{I}$ and $\mathcal{R}_{\mathbf{R}}(z) = 1$. Hence, (13a) and (13b) become

$$\eta = 1 \tag{24}$$

$$\chi = \frac{\sigma^2}{\sigma^2+1}. \tag{25}$$

Since $\mathbf{A} = \mathbf{I}$ implies $p = 1$, (22) yields $\hat{p} = 1$ and (recalling (21))

$$\begin{aligned}\mathcal{I} &= I(V_0; V_0 + Z) - q\log\left(1+\sigma^2\right) && (26)\\ &= I(V_0; V_0 + Z) - I(V_0; V_0 + Z|B_0) && (27)\\ &= I(B_0; V_0 + Z) && (28)\\ &= h(q) - H(B_0|V_0 + Z) && (29)\end{aligned}$$

where (28) follows from the fact, in the single-letter model $Y = V_0 + Z/\sqrt{\eta}$, $B_0 \leftrightarrow V_0 \leftrightarrow Y$ with $V_0 = X_0B_0$ as defined in Theorem 2. In fact, in this case, $\mathcal{I} = \frac{1}{n}I(\mathbf{B}; \mathbf{y}|\mathbf{A} = \mathbf{I}, \mathbf{U})$ for all n, not only in the limit of $n \to \infty$.

III. Upper Bounds on the Mutual Information Rate

As a direct consequence of the data processing theorem, we can bound the mutual information rate in the special case in which $\mathbf{U}$ is unitary, by

$$\mathcal{I} \le I(V_0; V_0 + Z) - q\log\left(1 + \sigma^2\right) \tag{30}$$

which holds with equality for $\mathbf{A} = \mathbf{I}$ (see (29)).

In the general case, we have the following upper bounds

Theorem 4

$$\mathcal{I}_1 \;\le\; \mathcal{V}_{\mathbf{R}}(q\sigma^2) \tag{31}$$

$$\mathcal{I}_1 \;\le\; I(V_0; \mathbb{E}\left[|\mathsf{R}|\right]V_0 + Z) \tag{32}$$

where Z and V_0 are as defined in Theorem 2, and where $|\mathsf{R}|^2$ is a random variable distributed as the limiting eigenvalues distribution of of $\mathbf{R}$.

The Gaussian relative entropy upper bound resulting in (31) (see [15]) leads to a high-SNR regime given by $p\log\sigma^2$. We know from [16] that the high-SNR regime of the term $\mathcal{I}_2$ behaves like $\min\{p,q\}\log\sigma^2$. Hence, for $p > q$ (i.e., the rank of the sensing matrix is larger than the sparsity of the support), the bound grows with SNR like $(p-q)\log\sigma^2$, while for $p < q$ the bound has a zero pre-log factor. It follows that for high SNR (large σ^2) the bound (31) grows unbounded for $q < p$, while it converges to a finite value for $q > p$.

Regarding (30), the term $I(V_0; V_0+Z)$ behaves at high SNR like $q\log\sigma^2$. Therefore, the right side of (30) converges to a finite value (independent of p and generally dependent on q), for large σ^2.

Finally, the term $I(V_0; \mathbb{E}\left[|\mathsf{R}|\right]V_0 + Z)$ also behaves like $q\log\sigma^2$ for high SNR. Therefore, the bound (32) has zero pre-log factor and converges to a finite value for $q < p$ and grows like $(q-p)\log\sigma^2$ for $q > p$.

We proceed to compare the mutual information rate $\mathcal{I}$ obtained for Haar-distributed $\mathbf{U}$ in Section II-D2, with the bounds (31) and (32), which particularized to that case, become

$$\mathcal{I} \;\le\; p\log(1+q\sigma^2) - q\log\left(1+\hat{p}\sigma^2\right) - d(p||\hat{p}) \tag{33}$$

$$\mathcal{I} \;\le\; I(V_0; pV_0 + Z) - q\log\left(1+\hat{p}\sigma^2\right) - d(p||\hat{p}) \tag{34}$$

where $\hat{p}$ is given in (22). Information (bits) versus p is shown in Figures 1-2, $\mathcal{I}$ is the lower curve; (33) is denoted by circles; (34) is denoted by the dashed-dotted line; and (30) (which does not depend on p) is denoted by a solid horizontal line.

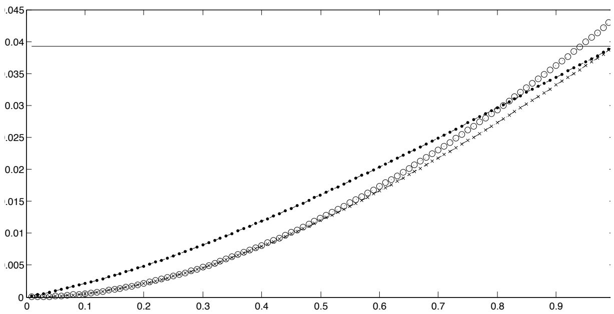

Fig. 1. $\mathcal{I}$ and bounds in Theorem 4 as a function of p; $q = 0.2$; $\sigma^2 = 1$.

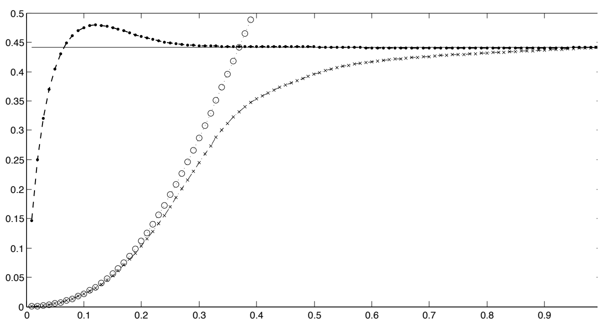

Fig. 2. $\mathcal{I}$ and bounds in Theorem 4 as a function of p; $q = 0.2$; $\sigma^2 = 100$.

IV. Decoupling Principle

By virtue of the decoupling principle introduced in the iid-matrix case by Guo and Verdú [8], the marginal joint distribution of each input coordinate and the output of a generic estimator converges (as the dimension grows) to a fixed input-output joint distribution that can be seen as induced by a "decoupled" (i.e., scalar) Gaussian channel with the same input. In this section, we apply the decoupling principle to the more general setting considered in this paper, whose validity is justified in [15]. Consider the observation model (1). Let

$$q_{\mathbf{y}|\mathbf{b},\mathbf{A},\mathbf{U},\mathbf{X}}(\mathbf{y}|\mathbf{b},\mathbf{A},\mathbf{U}\mathbf{X}) = \left(\frac{\gamma}{\pi}\right)^n \exp\left(-\gamma \left\|\mathbf{y} - \mathbf{A}\mathbf{U}\mathbf{X}\mathbf{b}\right\|^2\right) \tag{35}$$

denote a *postulated* transition probability distribution, and let $q_B(\mathbf{b}) = \prod_{i=1}^n q_B(b_i)$ denote a *postulated* a-priori distribution for $\mathbf{b}$, not necessarily Bernoulli-q. We consider a class of estimators for $\mathbf{b}$ given $\mathbf{y}, \mathbf{A}, \mathbf{U}$ parameterized by γ and $q_B(\cdot)$, defined as follows. Letting $p_X(\mathbf{x}) = \frac{1}{(\pi\sigma^2)^n} e^{-\|\mathbf{x}\|^2/\sigma^2}$ and applying Bayes rule, the assumed a-posteriori distribution for $\mathbf{b}$ given $\mathbf{y}, \mathbf{A}, \mathbf{U}$ is

$$\begin{aligned} & q_{\mathbf{b}|\mathbf{y},\mathbf{A},\mathbf{U}}(\mathbf{b}|\mathbf{y},\mathbf{A},\mathbf{U}) \\ &= \frac{\int q_{\mathbf{y}|\mathbf{b},\mathbf{A},\mathbf{U},\mathbf{X}}(\mathbf{y}|\mathbf{b},\mathbf{A},\mathbf{U}\mathbf{X}) q_B(\mathbf{b}) p_X(\mathbf{x}) d\mathbf{x}}{\sum_{\mathbf{b}'} \int q_{\mathbf{y}|\mathbf{b},\mathbf{A},\mathbf{U},\mathbf{X}}(\mathbf{y}|\mathbf{b}',\mathbf{A},\mathbf{U}\mathbf{X}) q_B(\mathbf{b}') p_X(\mathbf{x}) d\mathbf{x}} \end{aligned} \tag{36}$$

Then, we are concerned with *Posterior-Mean Estimators* of the form

$$\widehat{\mathbf{b}}(\mathbf{y},\mathbf{A},\mathbf{U}) = \sum_{\mathbf{b}} \mathbf{b}\, q_{\mathbf{b}|\mathbf{y},\mathbf{A},\mathbf{U}}(\mathbf{b}|\mathbf{y},\mathbf{A},\mathbf{U}) \tag{37}$$

Formally, $\widehat{\mathbf{b}}(\mathbf{y},\mathbf{A},\mathbf{U})$ is the conditional mean of $\mathbf{b}$ with respect to the postulated joint distribution induced by (35) and $q_B(\cdot)$. If $\gamma = 1$ and $q_B(\cdot)$ is Bernoulli-q, then (37) coincides with the (non-linear) MMSE estimator. However, considering general γ and $q_B(\cdot)$ allows to study a whole family of estimators through the same unified framework [10], [8]. For the purpose of analysis, it is convenient to define the conditional distribution of $\mathbf{y}$ given $\mathbf{A}, \mathbf{B}$, is induced by (1), by letting $\mathbf{b}_0 \sim p_{B_0}(\cdot)$, Bernoulli-$q$, and defining the conditional Gaussian distribution

$$p_{\mathbf{y}|\mathbf{b}_0,\mathbf{A},\mathbf{U},\mathbf{X}}(\mathbf{y}|\mathbf{b}_0,\mathbf{A},\mathbf{U}\mathbf{X}) = \frac{1}{\pi^n} \exp\left(-\left\|\mathbf{y} - \mathbf{A}\mathbf{U}\mathbf{X}\mathbf{b}_0\right\|^2\right) \tag{38}$$

Then,

$$\begin{aligned} & p_{\mathbf{y}|\mathbf{b}_0,\mathbf{A},\mathbf{U}}(\mathbf{y}|\mathbf{b}_0,\mathbf{A},\mathbf{U}) \\ &= \int p_{\mathbf{y}|\mathbf{b}_0,\mathbf{A},\mathbf{U},\mathbf{X}}(\mathbf{y}|\mathbf{b}_0,\mathbf{A},\mathbf{U}\mathbf{X}) p_X(\mathbf{x}) d\mathbf{x}. \end{aligned} \tag{39}$$

The decoupling principle can be stated as follows. Let $(B_{0i}, B_i, \widehat{B}_i)$ denote the i-th components of the random vectors $\mathbf{b}_0, \mathbf{b}, \widehat{\mathbf{b}}$, obeying the joint n-variate measure

$$p_{B_0}(\mathbf{b}_0) p_{\mathbf{y}|\mathbf{b}_0,\mathbf{A},\mathbf{U}}(\mathbf{y}|\mathbf{b}_0,\mathbf{A},\mathbf{U}) q_{\mathbf{b}|\mathbf{y},\mathbf{A},\mathbf{U}}(\mathbf{b}|\mathbf{y},\mathbf{A},\mathbf{U}) \tag{40}$$

with $\widehat{\mathbf{b}} = \widehat{\mathbf{b}}(\mathbf{y},\mathbf{A},\mathbf{U})$ given by (37). In passing, notice that (40) satisfies the conditional Markov chain $\mathbf{b}_0 \to \mathbf{y} \to \mathbf{b}$, given $\mathbf{A}, \mathbf{U}$. Then, in the limit of $n \to \infty$, under the assumption that the replica-symmetric analysis holds (see [15]), the joint distribution of $(B_{0i}, B_i, \widehat{B}_i)$ converges to the joint distribution of the triple $(B_0, B, \widehat{B})$ induced by

$$p_{B_0}(b_0) p_{Y|b_0;\eta}(y|b_0) q_{B|Y;\xi}(b|y) \tag{41}$$

with $\widehat{B} = \sum_b b q_{B|Y;\xi}(b|y)$, where $p_{B_0}(\cdot)$ is the original Bernoulli-q distribution, where we define

$$p_{Y|V_0;\eta}(y|v_0) = \frac{\eta}{\pi} \exp\left(-\eta \left|y - v_0\right|^2\right) \tag{42}$$

$$p_{Y|B_0;\eta}(y|b_0) = \int p_{Y|V_0;\eta}(y|x_0 b_0) p_X(x_0) dx_0 \tag{43}$$

and

$$q_{Y|V;\xi}(y|v) = \frac{\xi}{\pi} \exp\left(-\xi \left|y - v\right|^2\right) \tag{44}$$

$$q_{Y|B;\xi}(y|b) = \int q_{Y|V;\xi}(y|xb) p_X(x) dx \tag{45}$$

$$q_{B|Y;\xi}(b|y) = \frac{q_{Y|B;\xi}(y|b) q_B(b)}{\sum_{b'} q_{Y|B;\xi}(y|b') q_B(b')}, \tag{46}$$

where X_0 and X are iid $\sim p_X(\cdot)$, we let $V = XB$ and $V_0 = X_0 B_0$, and where the parameters η and ξ are obtained by solving the system of fixed-point equations

$$\chi = \gamma \mathsf{mmse}(V|Y) \tag{47}$$

$$\delta = \mathbb{E}\left[|V_0 - \mathbb{E}[V|Y]|^2\right] \tag{48}$$

$$\xi = \gamma \mathcal{R}_{\mathbf{R}}(-\chi) \tag{49}$$

$$\eta = \frac{(\xi/\gamma)^2}{\xi/\gamma + \dot{\mathcal{R}}_{\mathbf{R}}(-\chi)(\delta - \chi)} \tag{50}$$

All expectations in (47) – (50) are with respect to the joint distribution of V_0, Y, V

$$p_{V_0}(v_0)p_{Y|v_0;\eta}(y|v_0)q_{V|Y;\xi}(v|y) \quad (51)$$

where $p_{V_0}(\cdot)$ is the Bernoulli-Gaussian induced by $V_0 = X_0B_0$, $p_{Y|V_0;\eta}(y|v_0)$ is given in (42) and where

$$q_{V|Y;\xi}(v|y) = \frac{q_{Y|V;\xi}(y|v)q_V(v)}{\int q_{Y|V;\xi}(y|v')q_V(v')dv'}, \quad (52)$$

with $q_{Y|V;\xi}(y|v)$ given in (44) and where $q_V(\cdot)$ is induced by $V = XB$.

If the solution to (47) – (50) is not unique, then we choose the solution that minimizes the "free energy" (evaluated in nats):

$$\mathcal{E} = \log\frac{e\pi}{\gamma} - \frac{\xi}{\eta} + \gamma + \xi(\delta - \chi) + \int_0^\chi \mathcal{R}_{\mathbf{R}}(-w)dw \\ - \delta\xi + \xi(\xi/\eta - 1)\chi/\gamma - \mathbb{E}\left[\log\left(q_{Y;\xi}(Y)\right)\right] - \log\frac{e\pi}{\xi} \quad (53)$$

with $q_{Y;\xi}(Y) = \int q_{Y|V;\xi}(y|v)q_V(v)dv$.

As an application of this decoupling principle, we can determine the minimum achievable $D(p,q,\sigma^2)$ by particularizing the above formulas for the MAP estimator of B_i given $\mathbf{y}, \mathbf{A}, \mathbf{U}$, operating according to the optimal decision rule

$$\widehat{B}_i^{\text{map}}(\mathbf{y}, \mathbf{A}, \mathbf{U}) = \arg\max_{b\in\{0,1\}} \mathbb{P}[B_i = b|\mathbf{y}, \mathbf{A}, \mathbf{U}] \quad (54)$$

Invoking the decoupling principle, we notice that in the matched case $\gamma = 1$, $q_B(\cdot) \equiv p_{B_0}(\cdot)$, the posterior distribution of B_i given $\mathbf{y}, \mathbf{A}, \mathbf{U}$ reduces to $p_{B_0|Y;\eta}(b_0|y)$ induced by (43), i.e., by the single-letter observation model

$$Y = X_0B_0 + \frac{1}{\sqrt{\eta}}Z \quad (55)$$

where η is obtained from (13a) – (13b) in Theorem 2. Using the fact that Y for $B_0 = 1$ is $\sim \mathcal{CN}\left(0, \sigma^2 + 1/\eta\right)$ and for $B_0 = 0$ is $\sim \mathcal{CN}(0, 1/\eta)$, we obtain

$$\mathbb{P}[B_0 = 1|Y = y] = \frac{1}{1 + \frac{1-q}{q}(1+\eta\sigma^2)\exp\left(-\frac{\sigma^2\eta|y|^2}{\sigma^2+1/\eta}\right)} \quad (56)$$

and $\mathbb{P}[B_0 = 0|Y = y] = 1 - \mathbb{P}[B_0 = 1|Y = y]$. The optimal asymptotic error rate (support estimation Hamming distortion) can be analyzed by considering the MAP rule applied to the decoupled channel (55), $\widehat{B}^{\text{map}}(y) = \arg\max_{b_0\in\{0,1\}} \mathbb{P}[B_0 = b_0|Y = y]$. The MAP rule for the estimation of B_0 from Y is given by

$$\widehat{B}^{\text{map}}(y) = \begin{cases} 1, & \text{for } |y|^2 \geq \tau \\ 0, & \text{elsewhere} \end{cases} \quad (57)$$

with

$$\tau = \frac{\eta\sigma^2 + 1}{\eta^2\sigma^2}\log\frac{(1-q)(\eta\sigma^2+1)}{q} \quad (58)$$

The detector yields $\widehat{B}^{\text{map}}(y) = 1$, regardless of the value of $y \in \mathbb{C}$, if $q > \frac{1+\eta\sigma^2}{2+\eta\sigma^2}$, in which case $D(p,q,\sigma^2) = 1 - q$. Otherwise,

$$D_{\text{map}}(p,q,\sigma^2) = q\left(1 - \exp\left(\frac{-\tau}{\sigma^2+\eta^{-1}}\right)\right) \\ + (1-q)\exp\left(-\eta\tau\right) \quad (59)$$

obtained from the fact that, conditioned on $B_0 = 0$ (resp. $B_0 = 1$), $|Y|^2$ in (55) is central χ-squared with two degrees of freedom with mean η^{-1} (resp. $\sigma^2 + \eta^{-1}$) .

Acknowledgements

S. Shamai, G. Caire and S. Verdú wish to acknowledge the Binational Science Foundation Grant N. 2008269. The work of G. Caire was partially supported by NSF Grant CCF-0729162. The work of S. Verdú was partially supported by the Center for Science of Information (CSoI), an NSF Science and Technology Center, under grant agreement CCF-0939370.

References

[1] A. M. Tulino and S. Verdú, "Random Matrix Theory and Wireless Communications," *Foundations and Trends In Communications and Information Theory*, vol. 1, no. 1, pp. 1–184, 2004.

[2] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory,* vol. 52, no. 4, pp. 1289-1306, Apr. 2006.

[3] E. J. Candés and T. Tao, "Near optimal signal recovery from random projections: Universal encoding strategies" *IEEE Trans. Inform. Theory,* vol. 52, no. 12, pp. 5406-5425, Dec. 2006.

[4] M. J. Wainwright, "Information theoretic limitations on sparsity recovery in the high-dimensional and noisy setting," *IEEE Trans. on Inform. Theory,* Vol. 55, pp. 5728–5741, 2009.

[5] W. Wang, M. J. Wainwright, K. Ramchandran, "Information-theoretic limits on sparse signal recovery: Dense versus sparse measurement matrices," *IEEE Transactions on Information Theory,* Vol. 56, No. 6, pp. 2967-2979, June 2010.

[6] G. Reeves and M. Gastpar, "Fundamental Tradeoffs for Sparsity Pattern Recovery," Arxiv 1006.3128v1, June 2010, Also, Proc. 2010 *IEEE Int. Symp. on Information Theory.*

[7] M. Akcakaya and V. Tarokh, "Shannon-theoretic limits on noisy compressive sampling," Vol. 56, No. 1, pp. 492–504, Jan. 2009.

[8] D. Guo and S. Verdú , "Randomly Spread CDMA: Asymptotics via Statistical Physics," *IEEE Trans. on Inform. Theory,* Vol. 51, No. 6, pp. 1983–2010, June 2005.

[9] S. Rangan, A. K. Fletcher, and V. K Goyal, "Asymptotic Analysis of MAP Estimation via the Replica Method and Applications to Compressed Sensing," arXiv:0906.3234v2 [cs.IT] 26 Aug. 2009.

[10] T. Tanaka, "A statistical mechanics approach to large-system analysis of CDMA multiuser detectors," *IEEE Trans. Inform. Theory,* vol. 48, pp. 2888-2910, Nov. 2002.

[11] D. Guo, D. Baron and S. Shamai (Shitz), "A Single-letter Characterization of Optimal Noisy Compressed Sensing," Forty-Seventh Annual Allerton Conference on Communication, Control, and Computing September 30-October 2, 2009, Allerton Retreat Center, Monticello, Illinois, USA.

[12] Y. Kabashima, T. Wadayama and T. Tananka, "Statistical Mechanical Analysis of a Typical Reconstruction Limit of Compressed Sensing," arXiv:1001.4298v2 [cs.IT] 2 Jun 2010

[13] M. Bayati and A. Montanari, "The dynamics of message passing on dense graphs, with applications to compressed sensing," (http://arxiv.org/abs/1001.3448)

[14] D. Donoho, "Precise Optimality in Compressed Sensing: Rigorous Theory and Ultra Fast Algorithms," Keynote Address, 2011 Workshop on Information Theory and Applications, UCSD, La Jolla, CA, February 2011

[15] A. Tulino, G. Caire, S. Shamai and S. Verdú, "Support Recovery with Sparsely Sampled Free Random Matrices: An Information Theoretic Approach," preprint, 2011.

[16] A. M. Tulino, G. Caire, S. Shamai and S. Verdú, "Capacity of Channels with Frequency-Selective and Time-Selective Fading," *IEEE Trans. Information Theory,* vol. 56, no. 3, pp. 1187–1215, Mar. 2010.