INSIGHTS INTO THE STABLE RECOVERY OF SPARSE SOLUTIONS IN OVERCOMPLETE REPRESENTATIONS USING NETWORK INFORMATION THEORY

Yuzhe Jin and Bhaskar D. Rao

Department of Electrical and Computer Engineering,
University of California at San Diego, La Jolla, CA 92093-0407, USA
{yujin, brao}@ucsd.edu

ABSTRACT

In this paper, we examine the problem of overcomplete representations and provide new insights into the problem of stable recovery of sparse solutions in noisy environments. We establish an important connection between the inverse problem that arises in overcomplete representations and wireless communication models in network information theory. We show that the stable recovery of a sparse solution with a single measurement vector (SMV) can be viewed as decoding competing users simultaneously transmitting messages through a Multiple Access Channel (MAC) at the same rate. With multiple measurement vectors (MMV), we relate the inverse problem to the wireless communication scenario with a Multiple-Input Multiple-Output (MIMO) channel. In each case, based on the connection established between the two domains, we leverage channel capacity results with outage analysis to shed light on the fundamental limits of any algorithm to stably recover sparse solutions in the presence of noise. Our results explicitly indicate the conditions on the key model parameters, e.g. degree of overcompleteness, degree of sparsity, and the signal-to-noise ratio, to guarantee the existence of asymptotically stable reconstruction of the sparse source.

Index Terms— Overcomplete representations, inverse problems, sparsity, multiple access channel, channel capacity.

1. INTRODUCTION

The problem of overcomplete representations and the computation of a sparse solution to the associated underdetermined inverse problem arises in many application domains, such as biomagnetic inverse problems, image restoration, bandlimited extrapolation and spectral estimation, channel equalization, sensor networks [1, 2], etc. The underlying sparse recovery problem is to represent a signal of interest by using the minimum number of vectors from the overcomplete dictionary. Although the original problem is NP-hard, several methods have been proposed to recover the sparse solution. In the noiseless case [3] it is shown that the sparse solution can be efficiently recovered via convex relaxation, and the method has also proven to be effective in noisy settings [4]. Greedy forward sequential selection methods have also shown to be effective in solving the same problem [5]. In addition, the presence of multiple measurements have shown to greatly improve the ability to recover the sparse solution [1, 6].

Amid the encouraging discoveries, some recent research effort has been focused on understanding the fundamental limitations associated with the stable recovery of sparse solutions. An important question, and the subject of this paper, is the fundamental limits on the ability of any algorithm to stably recover sparse solutions in the presence of noise. Among recent attempts, information-theoretic tools have begun to be employed to explore insights into this problem. For instance, by modeling the inverse problem as a measurement channel, [2] combines channel capacity and rate-distortion theory to reach design constraints. In [7], the authors derive performance bounds based on an analysis of the error probabilities of optimal decoding schemes. [8] shows how fast decoding error probability decays with model parameters for signals from binary constellation.

In this work, we show that computing the sparse solution to the inverse problem arising in overcomplete representations in the presence of measurement noise can be modeled as the problem of decoding competing users simultaneously transmitting information through a multiple access channel (MAC). This connection is then used to shed insight into the stable recovery problem, in particular the dependence of the successful decoding of one activation site (non-zero entry) not only on the noise level but also on the behavior of all other active sites. In addition, when multiple measurement vectors (MMV) are available we show that the problem can be modeled as decoding information transmitted through a MIMO channel. The capacity results from the MIMO literature shed insight into the improved performance enabled by MMV. In both cases, we utilize channel capacity results with outage analysis to expose fundamental limits on the asymptotically stable recovery of sparse solutions.

2. PRELIMINARY BACKGROUND

2.1. Inverse problem with sparsity requirement

We consider the signal model with measurement noise as following,

$$B = AX + W \tag{1}$$

where $A \in \mathcal{R}^{M\times N}$ with $M \leq N$, and, usually, $M \ll N$. A is often referred to as the overcomplete dictionary. $X \in \mathcal{R}^{N\times L}$, is the source signal to be recovered. We can partition X into columns, i.e. $X = [\mathbf{x}_1, ..., \mathbf{x}_L]$, representing L vectors. $W \in \mathcal{R}^{M\times L}$, is measurement noise corrupting the measurements, and it can be also partitioned as $W = [\mathbf{w}_1, ..., \mathbf{w}_L]$. In this work, each element of W is assumed to be i.i.d. Gaussian, i.e. $N(0, \sigma_n^2)$. $B \in \mathcal{R}^{M\times L}$, and $B = [\mathbf{b}_1, ..., \mathbf{b}_L]$ are the L measurement vectors or signals to be represented. To simplify notation, define integer set $\mathbb{N}_Q \triangleq \{1, 2, ..., Q\}$.

Model (1) can be recognized as an inverse problem either with single measurement vector (SMV) when $L = 1$, or with multiple measurement vectors (MMV) when $L > 1$. In this paper, we will consider both scenarios and make the following distinct and important assumptions about the desired solutions [1].

(i) The solution vectors $\mathbf{x}_l, l \in \mathbb{N}_L$, are sparse, i.e. most of the entries are zero. This requirement is the same for SMV and MMV.

(ii) The solution vectors $\mathbf{x}_l, l \in \mathbb{N}_L$, are required to have the same sparsity profile so that the indices of the non-zero entries are independent of l. This requirement reflects the consistency of the underlying sources' activities, and provides informative coupling between the vectors.

We denote the number of non-zero rows in X by K, which for SMV will reduce to the number of non-zero entries in the column vector. Note that $K \ll N$ in the sparse source recovery problem.

This work was supported by NSF grant DGE-0333451 and the RESCUE project, NSF grant 0331690.

2.2. Information channels

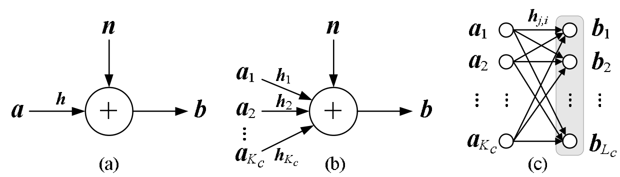

Fig. 1. Information channels.

We introduce several important information channels that will be employed in this work. Fig.1-(a) depicts a typical fading Gaussian channel, where $\boldsymbol{a}$ is the channel input, $\boldsymbol{h}$ is the fading channel gain, and $\boldsymbol{b}$ is the channel output corrupted by white Gaussian noise $\boldsymbol{n}$. In Fig.1-(b), K_c users, $\boldsymbol{a}_i$, with corresponding fading channel gains $\boldsymbol{h}_i, i \in \mathbb{N}_{K_c}$, access a Gaussian channel simultaneously with only one receiver observing the noisy channel output $\boldsymbol{b}$, forming a fading multiple access channel (MAC). Fig.1-(c) illustrates a fading multiple-input multiple-output (MIMO) channel, where L_c receivers, $\boldsymbol{b}_i, i \in \mathbb{N}_{L_c}$, are added into channel (b) with information paths linking all transmitter-receiver pairs. $\boldsymbol{h}_{j,i}$ denotes the fading channel gain between user i and receiver j. The shaded box emphasizes the cooperation among all receivers.

For each channel, user j, $j \in \mathbb{N}_{K_c}$, will have access to a codebook $\mathcal{C}^{(j)} = \{\mathbf{c}_1^{(j)}, ...\mathbf{c}_{N_c^{(j)}}^{(j)}\}$, where $\mathbf{c}_i^{(j)}, i \in \mathbb{N}_{N_c^{(j)}}$, are codewords of length M_c. The Rate R_j of user j's codebook is defined as

$$R_j = (\log N_c^{(j)})/M_c \tag{2}$$

To transmit information, a user will select a codeword from its codebook and transmit it through the channel making M_c uses of the channel. Channel capacity C determines the supremum of amount of information that can be conveyed through the channel with diminishing error probability of decoding.

3. MAC VS. SPARSE RECOVERY PROBLEM WITH SMV

3.1. $K = 1$: Only one non-zero entry in X

We motivate the connection between MAC and inverse problem with SMV, i.e. $L = 1$, by first studying the simplest case where only one non-zero entry exists in the source vector X.

We first rewrite Eq. (1) for $L = 1$ as following,

$$\mathbf{b}_1 = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_N]\,[x_1, x_2, \ldots, x_N]^\intercal + \mathbf{w}_1 \tag{3}$$

where $\mathbf{a}_j, j \in \mathbb{N}_N$, are column vectors of A. Without loss of generality, suppose the non-zero entry is at index i. By peeling off columns due to the zero entries in X, the *effective* form of Eq. (3) is

$$\mathbf{b}_1 = x_i \mathbf{a}_i + \mathbf{w}_1 \tag{4}$$

Clearly, the only column which survives the matrix multiplication in (3) is $\mathbf{a}_i$, and it is scaled by x_i and later contaminated by noise $\mathbf{w}_1$. Eq. (4) is to be contrasted with the scalar fading Gaussian channel which is mathematically described as

$$\boldsymbol{b} = \boldsymbol{h}\boldsymbol{a} + \boldsymbol{n} \tag{5}$$

where $\boldsymbol{n} \sim N(0, \sigma_n^2)$, and pictorially depicted in Fig.1-(a).

To bridge model (4) and (5), we now discuss the key connections between the two domains.

1. $\mathbf{a}_i$ as codeword. We can treat the overcomplete dictionary A as a codebook with each column vector $\mathbf{a}_j$, $j \in \mathbb{N}_N$, as a codeword. Each element of $\mathbf{a}_i$ is fed one by one to the channel (5) as input $\boldsymbol{a}$, resulting in M uses of the channel. Also, $\mathbf{a}_i$ can be viewed as stacking M transmissions of channel input $\boldsymbol{a}$. Noise $\mathbf{w}_1$ and observation $\mathbf{b}_1$ can be related to channel noise $\boldsymbol{n}$ and channel output $\boldsymbol{b}$, respectively, in the same fashion.

2. Randomness in index i. For the inverse problem, we assume that the index i of the non-zero entry can take any value equal-likely over the integer set $\mathbb{N}_N$. This, in the communication context, is akin to the user selecting randomly a codeword from the codebook $\mathcal{C}$ of size N, with each codeword of length M, for transmission.

3. x_i as fading channel gain. We model the value taken by the source, x_i, as a random quantity. The fading channel gain $\boldsymbol{h}$ plays the same role as the random source x_i. It's realized once and then kept fixed during the entire channel use. Essentially, we can interpret the vector representation of (4) as M consecutive uses of the underlying slow fading Gaussian channel (5) with appropriate stacking of the inputs/outputs into vectors.

4. Similarity of objectives. To complete the analogy, we examine the goals in the two domains. For the sparse recovery problem, we aim at identifying the non-zero index i and estimating the value x_i. In the communication context, the goal is successful decoding. To make the problem simpler without significant loss in insight[1] and to make the connection exact, we assume that the gain of the non-zero entries are known similar to the knowledge of channel gain in communications. The important issue of identifying the location of the non-zero entry then becomes the main focus. Based on the above-mentioned connections, the stable recovery problem is tantamount to identifying the correct codeword, i.e. successful decoding, in the communication context.

In summary, the problem of computing a sparse solution to the inverse problem can be interpreted as a channel decoding problem. Consequently, results from channel coding theory can be used to shed light on the sparse recovery problem. In particular, channel capacity results can help shed light on the fundamental limits on the stable recovery problem in the presence of noise. More explicitly, we are now concerned with the stable recovery of a sparse solution with one non-zero entry in the presence of noise, i.e. the recovery of the location (index) of the non-zero entry.

To proceed, we assume input power constraint var$(\boldsymbol{a}) \leq \sigma_a^2$. The instantaneous channel capacity of (5) is given by [9],

$$C(\boldsymbol{h}) = 0.5 \log\left[1 + (\sigma_a^2/\sigma_n^2)\boldsymbol{h}^2\right] \tag{6}$$

and (6) can be achieved when the channel input is Gaussian, i.e. $\boldsymbol{a} \sim N(0, \sigma_a^2)$. Correspondingly, to achieve best performance, elements of A should be independently generated according to $N(0, \sigma_a^2)$.

According to (2), the rate R_A of codebook A is defined as

$$R_A = (\log N)/M \tag{7}$$

Based on channel coding theorem [9], to guarantee successful transmission, it's required that

$$R_A < C(\boldsymbol{h}) \tag{8}$$

Substitute (6) and (7) into (8) immediately yields the condition for any decoding scheme to theoretically succeed,

$$(\log N)/M < 0.5 \log\left[1 + \boldsymbol{h}^2 \mathsf{SNR}\right] \tag{9}$$

where $\mathsf{SNR} \triangleq \sigma_a^2/\sigma_n^2$. Correspondingly, (9) can be also viewed as the condition for any possible sparsity recovery method to succeed. To proceed, we introduce outage probability to account for the random nature of the channel gain $\boldsymbol{h}$ (which in turn is the sparse source x_i). An outage event is defined as the realization of channel gain $\boldsymbol{h}$ fails to support the target rate, i.e. $R_A > C(\boldsymbol{h})$. It will definitely lead to decoding failure, which equivalently means we cannot recover the sparse solution. In practice, we need to set a threshold for the probability that an outage event occurs, in order to evaluate

[1] Once the positions are accurately determined the exact value of x_i can be readily computed using a least squares approach.

the performance of a system. Suppose we upper-bound the outage probability by ϵ, which translates into

$$P\left(0.5\log\left[1+\boldsymbol{h}^2\mathsf{SNR}\right] < (\log N)/M\right) \leq \epsilon \tag{10}$$

To solve for an explicit relation between key parameters, we assume $\boldsymbol{h} \sim N(0, \sigma_h^2)$, which in turns implies $x_i \sim N(0, \sigma_h^2)$. It can be shown that the following inequality must hold for (10) to be valid,

$$N \leq \left(1+2\delta^2\sigma_h^2\mathsf{SNR}\right)^{M/2} \tag{11}$$

where $\mathrm{erf}(\delta)=\epsilon$ and $\mathrm{erf}(\cdot)$ is the error function. (11) clearly unveils the fundamental tradeoff between degree of overcompleteness, N vs M, and other model parameters, such as source signal strength σ_h^2, environment noise level σ_n^2. Especially, the exponential growth of N with increasing M indicates the degree of overcompleteness possible and still have stable recovery.

3.2. $K > 1$: Analogy with Gaussian MAC

To generalize our discussion, let's suppose throughout the paper that $s_i, i \in \mathbb{N}_K$, are the indices of non-zero rows in X. Now the effective form of Eq. (4) is given by

$$\mathbf{b}_1 = x_{s_1}\mathbf{a}_{s_1} + x_{s_2}\mathbf{a}_{s_2} + \ldots + x_{s_K}\mathbf{a}_{s_K} + \mathbf{w}_1 \tag{12}$$

The corresponding scalar Gaussian channel is as follows,

$$\boldsymbol{b} = \boldsymbol{h}_1\boldsymbol{a}_1 + \boldsymbol{h}_2\boldsymbol{a}_2 + \ldots + \boldsymbol{h}_K\boldsymbol{a}_K + \boldsymbol{n} \tag{13}$$

where $\boldsymbol{n} \sim N(0, \sigma_n^2)$. To bridge Eq. (12) and (13), we notice that for $i \in \mathbb{N}_K$, fading channel gain $\boldsymbol{h}_i$ correspond to the value of the random source signals x_{s_i}. They are realized once and held fixed. Next, we can view the existence of each non-zero entry as a user with $\mathbf{a}_{s_i}$ corresponding to the codeword selected by user i for transmission. Equation (13) corresponds to one use of the channel and (12) can be interpreted as the vector version obtained after M uses. Hence, recovering the sparse solution and decoding multiple users' messages are just different interpretation of the same problem. We again employ information-theoretic results based on (13) to obtain performance limits of the sparse recovery problem.

Note that channel (13), pictorially demonstrated in Fig.1-(b), is recognized as a slow fading Gaussian MAC with K users. To characterize the limited channel resource shared by competing users and the potential trade-offs, a *capacity region* is usually computed to represent all possibilities of admissible user-rate allocations for successful transmission. Assume all users hold the same power constraint, i.e. $\mathrm{var}(\boldsymbol{a}_i) \leq \sigma_a^2, i \in \mathbb{N}_K$. The capacity region of channel (13) is then described by [9],

$$\sum_{i\in T} R_i \leq 0.5\log\left(1+\mathsf{SNR}\sum_{j\in T}\boldsymbol{h}_j^2\right),\ \forall\, T \subseteq \mathbb{N}_K \tag{14}$$

where R_i is the rate for user i. Note that Eq. (14) consists of (2^K-1) inequalities. Equalities hold when all users are i.i.d. Gaussian, i.e. $\boldsymbol{a}_i \sim N(0, \sigma_a^2), i \in \mathbb{N}_K$. This requirement also means all elements in codebook A should be independently generated according to the same Gaussian distribution to ensure the best performance.

It's worth pointing out that because different non-zero entries mirror different users, we implicitly assume different users won't select the same codeword, or they will collapse into one user otherwise. Meanwhile, all users have access to a common codebook A. Although these factors may have a negative impact on achievable capacity region, it is negligible and convenient to assume all users are operating at *equal* rate given in (7) when $K \ll N$. Hence, for successful transmission, one requires that

$$\sum_{i\in T}\boldsymbol{h}_i^2 \geq \left(N^{2|T|/M}-1\right)\mathsf{SNR}^{-1},\ \forall\, T \subseteq \mathbb{N}_K$$

where $|\cdot|$ computes the cardinality of its argument. Assuming users experience i.i.d. fading, i.e. $\boldsymbol{h}_i \sim N(0, \sigma_h^2), i \in \mathbb{N}_K$. Mathematically, the probability of success (i.e. complement of outage probability) can be lower-bounded by

$$\begin{aligned}
& P\left(\bigcap_{T\subseteq\mathbb{N}_K}\left\{\sum_{i\in T}\boldsymbol{h}_i^2 \geq \left(N^{2|T|/M}-1\right)\mathsf{SNR}^{-1}\right\}\right) \\
\geq\ & P\left(\bigcap_{i\in\mathbb{N}_K}\left\{\boldsymbol{h}_i^2 \geq \left(N^{2K/M}-1\right)\mathsf{SNR}^{-1}K^{-1}\right\}\right) \quad (15)\\
=\ & P\big(\boldsymbol{h}_1^2 \geq (N^{2K/M}-1)\mathsf{SNR}^{-1}K^{-1}\big)^K \quad (16)
\end{aligned}$$

where $\{\cdot\}$ denote the probability event of its argument. (16) follows from the i.i.d. fading gains. The requirement of upper-bounding outage probability by ϵ translates into lower-bounding (16) by $(1-\epsilon)$. Hence, we obtain an explicit relation between model parameters that achieves desired performance,

$$N \leq (1+2K\delta_1^2\sigma_h^2\mathsf{SNR})^{\frac{M}{2K}} \tag{17}$$

where $\mathrm{erf}(\delta_1) = 1-\sqrt[K]{1-\epsilon}$. Inequality (17) reveals the constraint on model parameters in the general sparse recovery problem. Note that the degree of sparsity, K, comes into the fundamental tradeoff with the most dominating effect as a factor in the exponent. It has a negative impact and forces a reduction of the scale of the overcomplete representation in order to stably recover a sparse solution with increasing number of non-zero entries.

3.3. Competition between non-zero entries

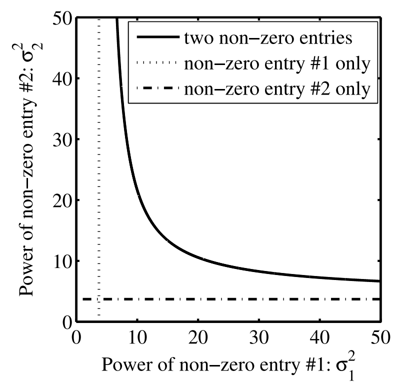

Fig. 2. Admissible regions for desired outage probability.

To show the competition between non-zero entries, we allow fading channel gains to have different variances, i.e. $\boldsymbol{h}_i \sim N(0, \sigma_i^2)$, which also means sparse source may have different signal powers. Fig.2 shows the Admissible Region (AR) of power pairs (σ_1^2, σ_2^2) for two-user MAC (or equivalently two non-zero entries in X) with outage probability $\epsilon \leq 0.1$, for $M=20$, $N=100$, $\mathsf{SNR}=10$dB. The AR is the upper-right side of each curve.

We see that the AR for two non-zero entries is strictly contained in that of each single non-zero case. Clearly, one signal source must have larger power to survive when other sources are active concurrently. Signal sources behave like interferences to each other, leading to a raised effective noise level.

4. MIMO VS. INVERSE PROBLEM WITH MMV

We extend our discussion to the case where multiple measurement vectors are available for recovery, i.e. $L > 1$. Our MIMO channel interpretation will inject novel insight into the performance improvement via MMV, which was originally explained [1] by the information coupling engraved into the common sparsity profile in source X. For MMV, the model of sparsity recovery (1) can be rewritten as

$$\mathbf{b}_j = [\mathbf{a}_1, \dots, \mathbf{a}_N]\,[x_{1,j}, \dots, x_{N,j}]^{\intercal} + \mathbf{w}_j, j \in \mathbb{N}_L \tag{18}$$

The effective form of (18) after removing zero terms is

$$\mathbf{b}_j = x_{s_1,j}\mathbf{a}_{s_1} + ... + x_{s_K,j}\mathbf{a}_{s_K} + \mathbf{w}_j, j \in \mathbb{N}_L \tag{19}$$

The corresponding communication channel for receive antenna j is

$$\boldsymbol{b}_j = \boldsymbol{h}_{j,1}\boldsymbol{a}_1 + ... + \boldsymbol{h}_{j,K}\boldsymbol{a}_K + \boldsymbol{n}_j, j \in \mathbb{N}_L \tag{20}$$

Note that the channel model (20) can be equivalently written as

$$\underline{\boldsymbol{b}} = \boldsymbol{H}\underline{\boldsymbol{a}} + \underline{\boldsymbol{n}} \tag{21}$$

where $\underline{\boldsymbol{a}} = [\boldsymbol{a}_1, ..., \boldsymbol{a}_K]^{\intercal}$, $\underline{\boldsymbol{b}} = [\boldsymbol{b}_1, ..., \boldsymbol{b}_L]^{\intercal}$, $\underline{\boldsymbol{n}} = [\boldsymbol{n}_1, ..., \boldsymbol{n}_L]^{\intercal}$, and $\boldsymbol{H} = (\boldsymbol{h}_{j,i}), j \in \mathbb{N}_L, i \in \mathbb{N}_K$. Channel model (21) can be readily recognized as a slow fading Gaussian MIMO channel with channel inputs $\underline{\boldsymbol{a}}$, outputs $\underline{\boldsymbol{b}}$, Gaussian noise $\underline{\boldsymbol{n}} \sim N(\mathbf{0}, \sigma_n^2 I)$ and channel matrix $\boldsymbol{H}$ whose elements are i.i.d. $N(0, \sigma_h^2)$. By matching quantities in (19) and (20) as before, recovering sparse solution with MMV resembles decoding messages passed through a MIMO channel. We consider the MIMO channel (21) as *a MAC with multiple receivers*. This alternative interpretation emphasizes both the competition between users and the advantage of joint decoding using multiple receivers. Similar to single receivers case in (13), we will consider the probability to successfully decode messages from any subset of users assuming receivers cooperate.

To proceed, define $\underline{\boldsymbol{h}}_i \triangleq [\boldsymbol{h}_{1,i}, ..., \boldsymbol{h}_{L,i}]^{\intercal}, i \in \mathbb{N}_K$. For $F \subseteq \mathbb{N}_K$, define $\boldsymbol{H}_F$ as a submatrix consisting of $\boldsymbol{H}$'s columns indexed by set F, i.e. $\boldsymbol{H}_F$ defines a sub MIMO channel that connects a subset of users indexed by F and all receivers. Next, define $\boldsymbol{G}_F = \boldsymbol{H}_F\boldsymbol{H}_F^{\intercal}$, if $|F| \geq L$, or $\boldsymbol{G}_F = \boldsymbol{H}_F^{\intercal}\boldsymbol{H}_F$, if $|F| \leq L$. Similar to (14), by treating (21) as a MAC with multiple receivers, successful transmission requires

$$\sum_{i\in T} R_i \leq 0.5 \log\det\left(I + \mathsf{SNR}\,\boldsymbol{G}_T\right), \,\forall\, T \subseteq \mathbb{N}_K \tag{22}$$

Due to the difficulty in analyzing the exact probability of success associated with capacity region (22), we instead seek an upper-bound and a lower-bound to gain insights. An upper-bound can be obtained by only considering the sum-rate of users, i.e.

$$\begin{aligned} & P\left(\bigcap_{T\subseteq\mathbb{N}_K}\left\{\sum_{i\in T} R_i \leq 0.5\log\det\left(I + \mathsf{SNR}\,\boldsymbol{G}_T\right)\right\}\right) \\ \leq\quad & P\left(\sum_{i\in\mathbb{N}_K} R_i \leq 0.5\log\det\left(I + \mathsf{SNR}\,\boldsymbol{G}_{\mathbb{N}_K}\right)\right) \\ \leq\quad & P\left(0.5r\log\left(1 + \mathsf{SNR}\,\boldsymbol{\lambda}_{\max}\right) \geq K(\log N)/M\right) \end{aligned} \tag{23}$$

where $\boldsymbol{\lambda}_{\max}$ is the largest eigenvalue of $\boldsymbol{G}_{\mathbb{N}_K}$, $r \triangleq \min\{K, L\}$. (23) is obtained by assuming again equal user rate and utilizing eigen-decomposition. By lower-bounding (23) by $(1-\epsilon)$, we reach

$$N \leq (1 + \xi\,\mathsf{SNR})^{\frac{rM}{2K}} \tag{24}$$

where $\int_0^{\xi} p_{\lambda_{\max}}(\lambda)d\lambda = \epsilon$, and $p_{\lambda_{\max}}(\lambda)$ is the probability density function of $\boldsymbol{\lambda}_{\max}$. Note that the dominating effect of L measurement vectors comes through r as a factor in the exponent. It suggests the significant potential increase in the ability of stably recovering sparse solutions from using MMV.

On the other hand, to obtain a lower-bound of probability of success, we proceed as follows,

$$\begin{aligned} & P\left(\bigcap_{T\subseteq\mathbb{N}_K}\left\{\sum_{i\in T} R_i \leq 0.5\log\det\left(I + \mathsf{SNR}\,\boldsymbol{G}_T\right)\right\}\right) \\ \geq\, & P\left(\bigcap_{T\subseteq\mathbb{N}_K}\left\{\sum_{i\in T} R_i \leq 0.5\log\left(1 + \mathsf{SNR}\sum_{j\in T}||\underline{\boldsymbol{h}}_j||^2\right)\right\}\right) \quad (25) \\ \geq\, & P\left(||\underline{\boldsymbol{h}}_1||^2 \geq (N^{2K/M} - 1)\mathsf{SNR}^{-1}K^{-1}\right)^K \quad (26) \end{aligned}$$

where (25) follows from the fact that the region is strictly contained in the original capacity region, (26) follows the same vein of (15) and (16). Lower-bounding (26) by $(1-\epsilon)$ yields an explicit tradeoff between model parameters

$$\gamma\left(\frac{L}{2}, \frac{N^{2K/M} - 1}{2K\sigma_h^2\mathsf{SNR}}\right) \leq \Gamma(L/2)\left(1 - \sqrt[K]{1-\epsilon}\right) \tag{27}$$

where $\gamma(\cdot,\cdot)$ and $\Gamma(\cdot)$ are two types of Gamma functions. For $L = 1$, (27) reduces to (17). For $L = 2$, (27) has closed-form as follows,

$$N \leq \left(1 - 2\log\left(1-\epsilon\right)\sigma_h^2\mathsf{SNR}\right)^{\frac{M}{2K}} \tag{28}$$

Comparisons between (17), (24) and (28) explain the improved performance by using MMV. The MIMO channel provides multiple data paths between sources and receivers, and greatly improves the system capacity. Especially, adding one measurement vector may greatly boost the recoverability when $L < K$.

5. SUMMARY

We established the connection between inverse problems with sparsity requirement and channels of wireless communication. Channel capacity along with outage analysis is utilized to uncover fundamental limits to the stable recovery problem in overcomplete representations and to understand the tradeoffs between the various parameters involved. The analysis also explains the gain in performance resulting from using MMV.

6. REFERENCES

[1] S. F. Cotter, B. D. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Trans. Sig. Proc.*, 2005.

[2] S. Sarvotham, D. Baron, and R. G. Baraniuk, "Measurements vs. bits: Compressed sensing meets information theory," *Proceedings of 44th Allerton Conf. Comm., Ctrl., Computing*, 2006.

[3] D. L. Donoho, "For most large underdetermined systems of linear equations the minimal l^1-norm solution is also the sparsest solution," *Preprint*, 2004.

[4] E. J. Candes, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Comm. Pure Appl. Math*, 2006.

[5] J. A. Tropp, "Greedy is good: Algorithmic results for sparse approximation," *IEEE Trans. Info. Theo.*, 2004.

[6] D. Wipf and B. D. Rao, "An empirical bayesian strategy for solving the simultaneous sparse approximation problem," *IEEE Trans. Sig. Proc.*, vol. 55, no. 7, pp. 3704–3716, 2007.

[7] M. Wainwright, "Information-theoretic bounds on sparsity recovery in the high-dimensional and noisy setting," *Proc. Int'l. Symp. Info. Theo.*, June 2007.

[8] J.A. Tropp, "Just relax: Convex programming methods for identifying sparse signal in noise," *IEEE Trans. Info. Theo.*, 2006.

[9] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*, Cambridge University Press, 2005.