Sensing Capacity of Sensor Networks: Fundamental Tradeoffs of SNR, Sparsity and Sensing Diversity

Shuchin Aeron Manqi Zhao Venkatesh Saligrama
Department of Electrical and Computer Engineering
Boston University, MA 02215
Email {shuchin,mqzhao,srv}@bu.edu

***Abstract*— A fundamental question in sensor networks is to determine the sensing capacity – the minimum number of sensors necessary to monitor a given region to a desired degree of fidelity based on noisy sensor measurements. In the context of the so called compressed sensing problem sensing capacity provides bounds on the maximum number of signal components that can be identified per sensor under noisy measurements. In this paper we show that sensing capacity is a function of SNR, signal sparsity—the inherent complexity/dimensionality of the underlying signal/information space and its frequency of occurrence and sensing diversity – the number of modalities of operation of each sensor. We derive fundamental tradeoffs between SNR, sparsity, diversity and capacity. We show that the capacity is a monotonic function of SNR and diversity. A surprising result is that as sparsity approaches zero so does the sensing capacity irrespective of diversity. This implies for instance that to reliably monitor a small number of targets in a given region requires disproportionately large number of sensors.**

I. Introduction

Consider a simple sensing architecture as shown in figure 1. The object $\mathbf{X}$ to be estimated/detected belongs to a n dimensional real space. The operator Φ can be thought of as a sensing modality which can be tuned based on the application. The direct observation of the phenomena $\mathbf{X}$ may be corrupted by external factors such as clutter (false events). One wishes to form an estimate $\hat{\mathbf{X}}$ given the corrupted set of observations $\mathbf{Y}$ such that $\frac{1}{n}d(\hat{\mathbf{X}},\mathbf{X}) \leq d_0$ for a given distortion measure. The knowledge of the operator Φ is assumed to be generally available at reconstruction end.

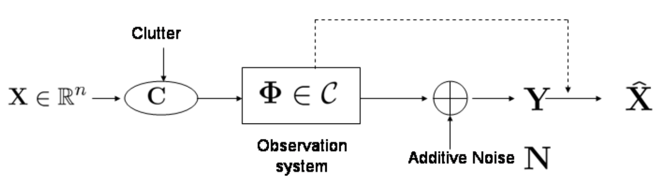

Fig. 1. A sensing architecture with various components illustrated

Generally speaking we are interested in performance characterization in terms of the modality Φ , the nature of the signals $\mathbf{X}$ and the signal to noise ratio. This set up models many statistical inferencing problems of interest, however, in this work we are primarily motivated by sensor network applications. Sensor networks involve judicious utilization of resource constrained communication networks formed out of geographically separated sensors. One could penalize different modalities Φ and formalize resource constraints in sensor networks. Rather than focusing on this aspect one of our primary goals here is to compare different modalities for different types of sensor networking problems such as counting, detection and localization. To concretize what we mean by modalities consider a network of N sensor arrays. Each sensor array has a maximum power P and can operate in different modalities as shown in Figure 2. Examples of modalities could be: (a) *Narrow-Beam High Resolution Modality:* Each sensing array can beamform to a different location and with maximum admissible SNR. Here sensing regions may not overlap. (b) *Wide-Beam Low Resolution:* Each sensing array can have a wider focus but with a corresponding decrease in SNR. Here sensing regions overlap. (c) *Random-Beams:* Each sensing array can beamform to a random set of locations. *Gaussian Modulated Beams:* Each sensing array can take a gaussian average of the beam-formed output for each location. Target localization under the first modality was studied by one of the authors in [8] where each sensor was assumed to have a limited sensing range with non-overlapping sensing regions.

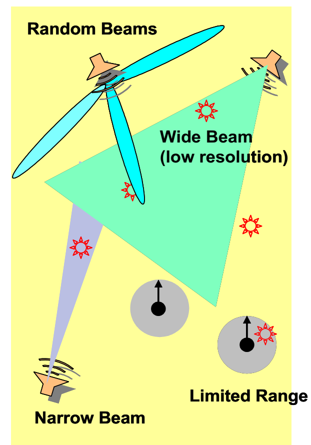

Fig. 2. Sensors with different modalities

It was shown there that when target density is unknown it is impossible to meaningfully control both false alarms and misses.

Another goal of this paper is to develop information-theoretic bounds for the emerging field of compressed sensing. These problems concern continuous natural phenomena. The principle observation here is that most natural phenomena of interest is compressible, i.e. succinctly representable in a natural basis, [13]. The problem of estimating such a phenomena can then be recast in a discrete framework as that of estimating the coefficients in some natural basis in which the signal has unique and sparse representation. Some examples of these applications pertaining to sensor networks are (1) Imaging a scattering medium [12] (2) MIMO radar [15]. (3) Geo-exploration via underground seismics. Compressed sensing deals with computing the minimum number of observations m required for perfect reconstruction of a sparse signal when given level of sparsity. This subject has been a subject of intensive research in the applied mathematics literature, [3], [1], [5], [13], [14], [9] and in signal processing and sensor networks literature [10], [6], [7]. The problem has its roots in the problem of finding sparsest representation in the given basis of signals observed in additive noise and was addressed in [13], [14]. Instead of looking at the entire signal, [3], [1], [5] focused on the *compressed sensing* problem, i.e., reconstruction of sparse signals from a small number of random projections of the signal. They showed that for a k-sparse signal $k+1$ random projections are necessary for exact reconstruction with probability one, in the absence of noise. In [2], [9] these results were extended to the case of noisy random projections.

Much of these developments have been algorithmic, and in a non-bayesian setting motivated by the good performance of L_1 relaxation to a combinatorial problem that involves an L_0 constraint. Our paper deviates from these approaches in many respects. First, we consider a Bayesian setting and adopt an information-theoretic perspective. Therefore the bounds that we derive are independent of any algorithm used for reconstruction. Next we consider both discrete and continuous signal spaces and formalize the problem in terms of *sensing capacity*. The sensing capacity for the discrete alphabets—for instance locations of targets—was introduced in [11]. We extend this definition to the continuous case. Sensing capacity is informally defined as the maximum signal dimension (or complexity)/sensor-measurement that can be recovered to a pre-defined degree of accuracy(distortion). Alternatively, it can be interpreted as the minimal number of sensors required to monitor a given region of interest.

Our contribution is two fold. First, we derive fundamental tradeoffs between sensing diversity, sparsity, SNR and sensing capacity. Second, in derivation of these bounds we explore alternate derivations of mutual information that reveal the effect of structure of the sensing matrix. Our results serve as operational guidelines for a network designer to build the most efficient sensing structure or system for particular application(s) of interest.

We reveal a surprising behavior. We show that sensing capacity goes to zero as the sparsity goes to zero independent of SNR and diversity. This means that it requires disproportionately large number of sensors to detect singular events in a large area. We show that small sensing diversity invariably leads to small sensing capacity. In other words, small diversity requires large number of sensors for achieving same accuracy relative to situations with large diversity. The effect of sensing diversity is characterized in a number of different settings.

The paper is organized as follows. In section II we will formalize the problem and give a precise definition of sensing capacity. In section III we will present the main results where we will provide upper and lower bounds to the sensing capacity for full diversity and no correlation. Then in section IV, we will provide results on the effect of the structure of the sensing matrix on sensing capacity. The proofs of our results can be found in BU-CISE technical report, - {http://www.bu.edu/phpbin/cise/search.php}, report number 2007-IR-0010).

II. Problem Set-up

In this section we will make the problem set up precise. We begin by defining signal sparsity and sensing diversity in relation to the set up of figure 1. In this work we will assume that there is no clutter and is modeled as a part of the signal to be estimated.

Definition 2.1: Let k be a positive integer $\leq n$ and $\alpha = \frac{k}{n}$. An n-dimensional signal $\mathbf{X}$ is said to be k-sparse, if each component of X is i.i.d distributed with probability of $(1-\alpha)$ to be zero and the non-zero components are chosen from some distribution $p_X(x)$. We denote α as the *sparsity ratio*.

A. Exploiting Transform Sparsity

Let $\mathbf{Z} \in \mathbb{R}^n$ be the signal of interest. Furthermore assume that we are able to exploit the *transform sparsity* in the recovery of $\mathbf{Z}$, i.e, $\mathbf{Z}$ can be represented as $\mathbf{\Phi X}$, in which $\mathbf{\Phi}$ is some basis and only k components of $\mathbf{X}$ is nonzero. Now the sensing model becomes,

$$\mathbf{Y} = \mathbf{Q\Phi X} + \mathbf{N} = \mathbf{GX} + \mathbf{N} \ (say)$$

B. The choice of sensing matrix

In this work we will restrict our choice to a particular class of random matrices, but we will also talk about non-Gaussian deterministic matrices as well. Specifically, the sensing matrix $\mathbf{Q} \in \mathbb{R}^{m\times n}$ is random matrix with i.i.d Gaussian entries. In addition the matrix $\mathbf{Q}$ is chosen independently of $\mathbf{X}$.

In order that the measurement SNR per sensor remains same for all the choices of sensing diversity factors, we impose that the total power in each row of $\mathbf{Q}$ be equal to P and with respect to the distribution over the non-zero values in $\mathbf{X}$, $\mathbf{E}[X^2] = 1$. Note that when $\mathbf{\Phi}$ is an orthonormal basis, the distribution of $\mathbf{G}$ is same as that of $\mathbf{Q}$. When $\mathbf{\Phi}$ is not orthonormal, in particular normal but not orthogonal, the resulting matrix $\mathbf{G}$ has correlations across the columns, though the rows are independent. In this paper we will mainly focus on the case when there is no correlation in $\mathbf{G}$. We will partially extend the results to correlated $\mathbf{G}$ with correlation induced by $\mathbf{\Phi}$.

Definition 2.2: Let l be a positive integer $\leq n$ and define $\beta = \frac{l}{n}$. For each row of the sensing matrix $\mathbf{G}$, suppose each entry $\mathbf{g}_{ij}$ has probability $1-\beta$ to be zero. We call β as the *sensing diversity* and say that the matrix $\mathbf{G}$ has diversity β.

For sake of brevity, we will denote the sensing matrix by $\mathbf{G}$ in the rest of the paper. In the following, $H_2(\alpha) = \alpha \log_2 \frac{1}{\alpha} + (1-\alpha)\log_2 \frac{1}{1-\alpha}$ is the binary entropy function. All logarithms in this paper are base 2 unless otherwise specified.

C. Sensing capacity and problem formulation

We define the sensing rate as the ratio of signal dimension to the number of sensors, $R = \frac{n}{m}$. We say that a sensing rate R is d_0-achievable if m projections $\{\mathbf{y}_i = \langle \mathbf{g}_i, \mathbf{X}\rangle + \mathbf{n}_i\}_{i=1}^m$ can ensure the reconstruction distortion to within d_0 with probability one. We have the following definition for sensing capacity.

Definition 2.3: The d_0-*distortion sensing capacity* is defined as $C(d_0) = \max R$ over all d_0-achievable R, i.e.,

$$C(d_0) = \limsup \left\{ \frac{n}{m} \mid Prob\{d(\mathbf{X}, \hat{\mathbf{X}}) > d_0\} \to 0, \ \ n, m \in \mathcal{Z}^+ \right\}.$$

Though the above definition includes perfect reconstruction as a special case, i.e. $d_0 = 0$. The problem that is considered in the paper is to find the Sensing Capacity under the linear sensing model proposed. Below we present our main results.

III. Main results

A. Sensing capacity

Theorem 3.1: If $\mathbf{G}$ is a random Gaussian matrix with i.i.d. elements and noise $\mathbf{N}$ being a Gaussian noise $\mathcal{N}(0, I_m)$, then for a sensing diversity of $l = n$, i.e. $\beta = 1$, the d_0-distortion sensing capacity obeys-

$$C \leq \frac{0.5\log(1+\rho\alpha)}{R_X(d_0)}$$

where $R_X(d_0)$ is the corresponding scalar rate distortion function and ρ the signal-to-noise ratio expected power in a row of G over the noise power (each of the non-zero signal components are normalized to unit power).

Proof: See Appendix ■

We point out that the sensing capacity provides asymptotic bounds on the minimum number of observations required to reliably recover the underlying signal/targets to within a distortion d_0, i.e.,

$$m \geq nC(d_0)$$

We illustrate these ideas through some examples. Assume for simplicity that $\mathbf{X} \in \{0,1\}^n$. In the extreme case when the signal $\mathbf{X}$ is 1-sparse, i.e., $k = 1$(or $\alpha = \frac{1}{n} \to 0$), by the above upper bound, we have,

$$C \leq \lim_{\alpha \to 0} \frac{\log(1+\rho\alpha)}{2(H_2(\alpha))} = 0$$

On the other hand, when $k = \frac{n}{2}$(or $\alpha = \frac{1}{2}$), the upper bound of the sensing capacity is given by (note $H_2(\frac{1}{2}) = 1$),

$$C \leq \frac{\log(1+\frac{1}{2}\rho)}{2}$$

Figure 3 shows how the sensing capacity changes with the sparsity ratio α when $\mathcal{X}$ is binary alphabet. Note that the sensing capacity approaches zero for small sparsity. Consequently, *disproportionately large number of sensors are required for monitoring sparse events.* Also note that as $SNR \to \infty$, i.e., when noise approaches zero the capacity approaches infinity. Indeed it is easy to show that a single measurement is sufficient for error-free reconstruction.

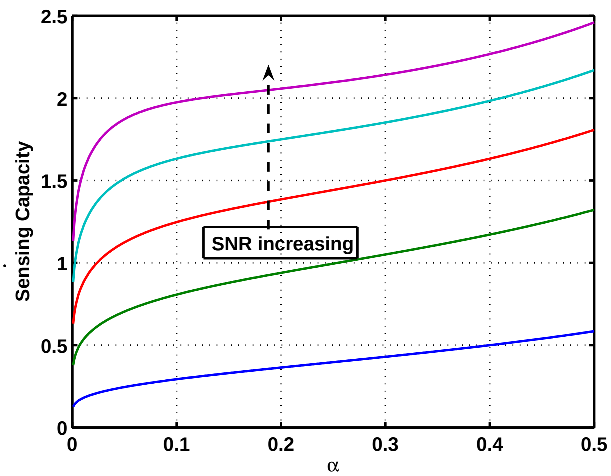

Fig. 3. The plot of Sparsity versus capacity for various SNRs ($\mathcal{X} = \{0,1\}$).

B. Achievable Sensing Capacity : Weak Achievability

In this section we will provide lower bounds to the sensing capacity, for the case $\beta = 1$ and $\mathbf{G}$ being a random Gaussian matrix with i.i.d. elements.

Given N points in $\mathbb{R}^n$, $\{\mathbf{X}_1, .., \mathbf{X}_N\}$. Let $\mathcal{B}_{\Delta, \mathbf{X}_i} = \left\{\mathbf{X} : \frac{1}{n} d(\mathbf{X}_i, \mathbf{X}) \leq \Delta\right\}$ be a distortion ball of average distortion Δ around the point X_i. Given $\epsilon > 0$, let N be the minimal number such that

$$Pr(\bigcup_{i=1}^{N} \mathcal{B}_{\Delta, \mathbf{X}_i}) \geq 1 - \epsilon$$

For sake of brevity denote $\mathcal{B}_{\Delta, \mathbf{X}_i} = \mathcal{B}_i$. Let the underlying vector generating the observation $\mathbf{Y}$ be $\mathbf{X}$. We seek the probability that we decide in favor of $\mathbf{X}'$ which is at an average distance of *at least* 2Δ from $\mathbf{X}$. We have,

$$P_e\left(\mathbf{X} \to \mathbf{X}' | \mathbf{G}, \mathbf{X} \in \mathcal{B}_i, \mathbf{X}' \in \mathcal{B}_j : \frac{1}{n} D(\mathcal{B}_i, \mathcal{B}_j) \geq \Delta\right)$$
$$\leq \max_{\mathbf{X} \in \mathcal{B}_i, \mathbf{X}' \in \mathcal{B}_j} e^{-||\mathbf{G}(\mathbf{X}'-\mathbf{X})||^2 \frac{1}{4N_0}}$$

Since the average set distance $\frac{1}{n} D(.,.)$ is $\geq \Delta$, the worst case probability of error (in set decoding) is determined by the sequences that are at an average distance of Δ. Taking expectation over $\mathbf{G}$ we have for the average pairwise probability of error,

$$P_e\left(\mathbf{X}\to\mathbf{X}'|\mathbf{X}\in\mathcal{B}_i,\mathbf{X}'\in\mathcal{B}_j:\tfrac{1}{n}D(\mathcal{B}_i,\mathcal{B}_j)\geq\Delta\right)$$
$$\leq \mathbf{E_G}exp\left\{-||\mathbf{G}(\mathbf{X}'-\mathbf{X})||^2\tfrac{1}{4N_0}\right\}$$
$$\leq\left(\frac{1}{1+\frac{\Delta P}{2N_0}}\right)^{m/2}$$

Without loss of generality, one can choose $\Delta = c\alpha$ for some $c\leq 2$. From rate distortion theory the minimal number of points such that the covering condition is satisfied with high probability is given by $2^{nR_X(\Delta)}$, where $R_X(\Delta)$ is the rate distortion function. Hence summing over all the distortion balls we get for the average probability of error that,

$$P_e$$
$$\leq \textstyle\sum_{j=1}^N P_e\left(\mathbf{X}\to\mathbf{X}'|\mathbf{X}\in\mathcal{B}_i,\mathbf{X}'\in\mathcal{B}_j:\tfrac{1}{n}D(\mathcal{B}_i,\mathcal{B}_j)\geq c\alpha\right)$$
$$\leq\left(\frac{1}{1+\dfrac{c\alpha}{2}\rho}\right)^{m/2} 2^{nR_{\mathbf{X}}(\Delta)}$$

Thus the achievable sensing capacity is given by

$$C(c\alpha)\geq C_{achieve}=\frac{\frac{1}{2}\log(1+\dfrac{c\alpha}{2}\rho)}{R_X(\Delta)}$$

Also, as $c\alpha\to 0$, $C_{achieve}$ is zero, i.e. zero distortion sensing capacity is zero. Also for any SNR, the ratio of upper and lower bounds satisfies:

$$\frac{C_{upper}}{C_{achiev}}=\frac{\log(1+\alpha\rho)}{\log(1+\frac{c\alpha}{2}\rho)}$$

The gap between upper and lower bounds for the discrete binary alphabet case are shown in figure 4. The achievability is weak in the sense that with the number of projections prescribed by the achievable capacity one can only ensure reconstruction to within distortion 2Δ instead of Δ.

IV. Effect of Different Sensing Modalities

In this section we will consider different sensing modalities and derive upper bounds for sensing capacity. First consider the case where sensing matrix is identity. This is the case, which arises typically when we have a limited sensing range and the sensing regions of the different sensors does not overlap. First note that from the data processing inequality, $I(X^n,Y^n)\geq I(X^n,GY^n)$ for any $m\times n$ matrix G with $m<n$ and such that each row of G has unit norm. While this implies that our lower bounds for the probability of error (based on generalized Fano's inequality (see appendix))will always be smaller than any other sensing modality G, the sensing capacity is always upper bounded by one. Consequently, we focus on sensing matrices that lead to sensing capacity larger than one.

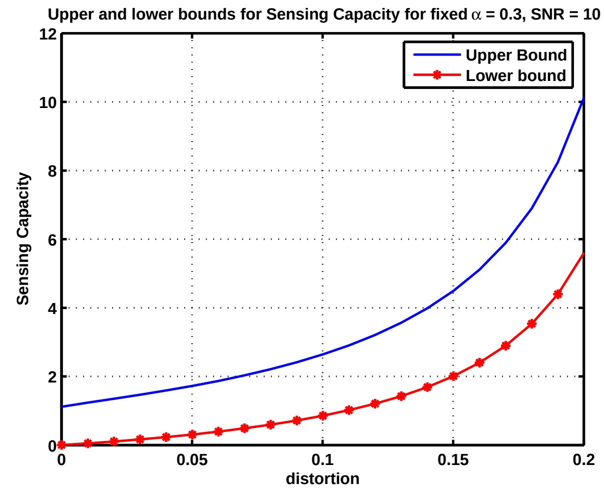

Fig. 4. The gap between Upper and Lower bounds on Sensing Capacity for the binary alphabet case as a function of distortion for a fixed sparsity. The distortion is the average Hamming distortion.

We next consider deterministic matrices G. We denote the rows of G as G_j, $j=1,2,\ldots,m$. Let the cross-correlations of these rows be denoted as:

$$r_k=\frac{G_k^TG_{k+1}}{G_k^TG_k}$$

As before to ensure the SNR, ρ to be fixed we impose $G_k^TG_k=P$ for all k and let $\rho=P/N_0$ where N_0 is the noise power. Then we have the following result:

Proposition 4.1: An upper bound for the sensing capacity for a deterministic sensing matrix $G\in\mathbb{R}^{m\times n}$ is given by:

$$C(d_0)\leq\sum_{k=1}^{m-1}\frac{\log\left(1+\frac{\alpha\rho}{\alpha\rho+1}(1+\alpha\rho(1-r_k))\right)}{R_X(d_0)} \tag{1}$$

The main outline of the proof consists in upperbounding $H(Y^m)$:

$$H(Y^m)\leq H(Y_1)+\sum_{j=1}^{m-1}H(Y_{j+1}\mid Y_j)\leq H(Y_1)+H(Y_{j+1}-\eta_jY_j)$$

where η_jY_j is the best MMSE estimate for Y_{j+1}.

We apply this result to several situations next. First consider a very wide-beam sensor, i.e, $G_j=\frac{1}{\sqrt{n}}[1\,1\,\ldots,1]$. This is a rank deficient situation it is clear that this modality will be unsuccessful in target localization or sparse signal estimation. On the other hand our sensing capacity bound leads to $C(d_0)\leq 1/\log(1/\alpha)$ which is slightly optimistic but does show that capacity is small. Next consider a sensing matrix arising as a consequence of a finite-impulse-response(FIR) filtering operation of the observed data. Suppose the filter length is L with all coefficients equal to one. It follows that the G matrix has a banded structure along the main diagonal. To guarantee coverage we need $m\geq n-L+1$. This implies that L has to be large in order for m to be small. However, this is similar to the previous case the sensing capacity turns out to be smaller than one. This is because the term $(1-r_k)$ in Equation 1 is close to zero because of relatively large overlap between subsequent rows of G. Consequently, we can consider a sub-sampling of the filtered output to reduce the overlap. Sub-sampling at rate L corresponds to no overlap case. In this case $r_k = 0$ and our bounds turn out to be too lose. Indeed, first we require that $m \geq n/L$ to guarantee coverage. This implies that L needs to be large to obtain small m. However, from Fano's inequality it is easy to show that $P_e \not\to 0$ in this case. This is because each observation does not provide sufficient information to localize within the L different locations. While we considered the no-overlap case these conclusions are not substantially for small overlaps (i.e. at sub-sampling rates smaller than L) as well.

To overcome these issues we observe from Equation 1 that a large sensing capacity upperbound requires r_k to be small while maintaining a significant overlap between any two rows. This implies that the cross-correlation between any two subsequent rows have to be negligible. By improving these bounds one can argue that indeed, the cross-correlations between any two rows must be negligible. This suggests that the components of the G matrix should be chosen randomly to obtain the best upperbounds. Consequently, we consider the effect of diversity on the sensing capacity. We will show that low diversity leads to a low sensing capacity. We will finally consider the effect of correlation on the achievable sensing capacity. We show that high level of correlation across the columns reduces achievable sensing capacity. To this end we have the following lemma.

Lemma 4.1: Suppose the sensing matrix has diversity $\beta = \frac{l}{n}$ and $\mathbf{X} \in \{0,1\}^n$, then we have,

$$I(\mathbf{X};\mathbf{Y}|\mathbf{G}) \leq \frac{m}{2}\mathbf{E}_j\left[\log\left(\frac{\rho}{l}j+1\right)\right], \quad (2)$$

where the expectation is evaluated over the distribution

$$\Pr(j) = \frac{\binom{k}{j}\binom{n-k}{l-j}}{\binom{n}{l}}$$

We provide the main idea behind the proof here and the details can be found in BU-CISE technical report, {http://www.bu.edu/phpbin/cise/search.php}, report number 2007-IR-0010). The main idea is that since the matrix $\mathbf{G}$ is chosen independently of $\mathbf{X}$ we can expand the mutual information between $\mathbf{X}$ and $\mathbf{Y},\mathbf{G}$ in two different ways as follows:

$$\begin{aligned} I(\mathbf{X};\mathbf{Y},\mathbf{G}) &= \underbrace{I(\mathbf{X};\mathbf{G})}_{=0} + I(\mathbf{X};\mathbf{Y}|\mathbf{G}) \\ &= I(\mathbf{X};\mathbf{Y}) + I(\mathbf{X};\mathbf{G}|\mathbf{Y}) \end{aligned}$$

This way of expanding gives us a handle on evaluating the mutual information with respect to the structure of the resulting sensing matrix $\mathbf{G}$. From above we get that,

$$\begin{aligned} I(\mathbf{X};\mathbf{Y}|\mathbf{G}) &= I(\mathbf{X};\mathbf{Y}) + I(\mathbf{X};\mathbf{G}|\mathbf{Y}) \\ &= h(\mathbf{Y}) - h(\mathbf{Y}|\mathbf{X}) + h(\mathbf{G}|\mathbf{Y}) - h(\mathbf{G}|\mathbf{X},\mathbf{Y}) \end{aligned}$$

The proof essentially requires evaluation of the last term in the above expression. The term $I(\mathbf{X};\mathbf{Y})$ turns out to be negligible since the sensing matrix is unknown.

A. Effect of sensing diversity

The following lemma follows from lemma 4.1 and lemma 6.1.

Lemma 4.2: If $\mathbf{G}$ is random Gaussian matrix with i.i.d. elements and the diversity factor is $\beta = \frac{l}{n}$, then the d_0-distortion Sensing Capacity for diversity of l obeys-

$$C \leq \frac{0.5\mathbf{E}_j \log(1+\frac{j\rho}{l})}{R_X(d_0)}$$

where the expectation is evaluated over the distribution

$$\Pr(j) = \frac{\binom{k}{j}\binom{n-k}{l-j}}{\binom{n}{l}}$$

To illustrate the effect of low sensing diversity on sensing capacity let $\beta = \frac{1}{n}$.From the upper bound for the mutual information in the partial diversity case given by lemma 4.1, we have

$$\begin{aligned} &I(\mathbf{X};\mathbf{Y}|\mathbf{G}) \\ &\leq \tfrac{m}{2}\mathbf{E}_i\left[\log\left(\tfrac{\rho}{l}i+1\right)\right] \\ &= \tfrac{m}{2}\mathbf{E}_i\left[\log\left(\rho i+1\right)\right] \\ &= \tfrac{m}{2}[(1-\alpha)\log(\rho\cdot 0+1)+\alpha\log(\rho+1)] \\ &= \tfrac{m\alpha}{2}\log(1+\rho) \end{aligned}$$

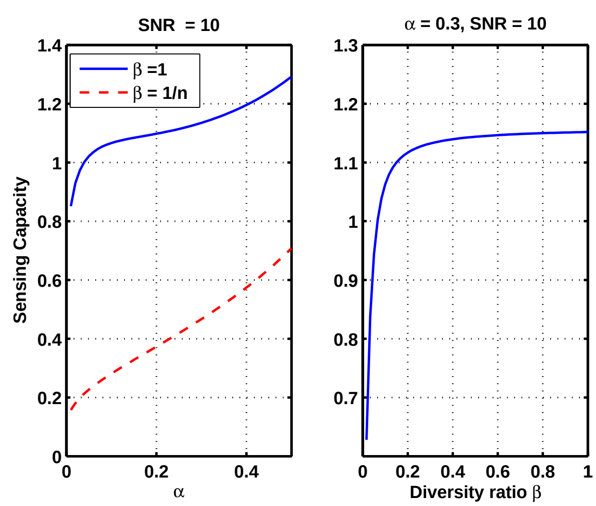

Fig. 5. The gap between sensing capacities in low diversity and full diversity for the binary alphabet case. Shown also is the Sensing Capacity as a function of diversity for fixed sparsity. Note the saturation effect of diversity.

The sensing capacity is ,

$$C \leq \frac{\frac{\alpha}{2}\log(1+\rho)}{H_2(\alpha)}$$

Compared with the sensing capacity inequality in theorem 3.1, we can see that low diversity will reduce the sensing capacity. The gap between these two cases (i.e., $l = 1$ v.s. $l = n$) are shown in figure 5. An interesting point to notice is that one can achieve almost all the sensing capacity with small (but sufficient) enough diversity. The dependence of Sensing Capacity on both sparsity and diversity is shown in figure 6.

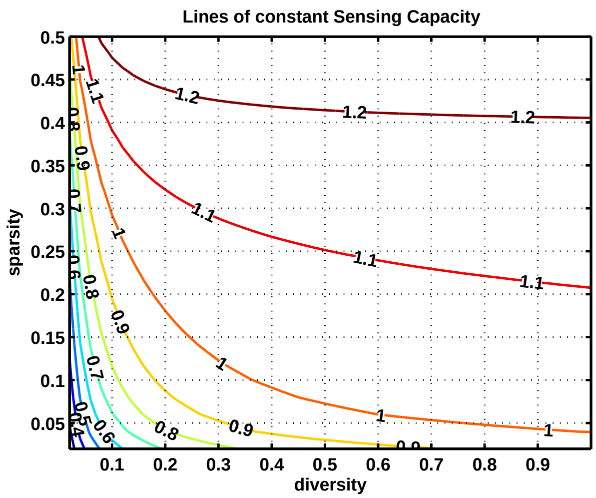

Fig. 6. Lines of constant sensing capacity as a function of sparsity and diversity for the binary alphabet case.

B. Effect of Correlation on achievable sensing capacity

Now we will show that correlation in sensing matrix $\mathbf{G}$ reduces achievable capacity. Consider the numerator of the exponent in the upper bound to the probability of pair-wise error in max-likelihood decoding -

$$\sum_{i=1}^{m} |\sum_{j=1}^{n} g_{ij}\Delta_j|^2$$

The terms $\sum_{j=1}^{n} g_{ij}\Delta_j$ for each i are independent Gaussian random variables with zero mean and variance given by- $\Delta^T \Sigma_{g_i} \Delta$ where Δ is the vector consisting of $\Delta_j, j = 1, .., n$ and Σ_{g_i} is the covariance matrix (symmetric and positive semi-definite) of the i-th row of $\mathbf{G}$. Let $\frac{1}{n}\Delta^T\Delta = c\alpha$. Then we can normalize the variance of each of $g_{ij} \sim \mathcal{N}(0, \dfrac{P}{n})$. Then, we can get an upper bound to the probability of error via

$$\min\ \Delta^T \tilde{\Sigma}_{g_i} \Delta = \lambda_{\min} \Delta^T \Delta$$

where $\lambda_{\min}$ is the minimum eigenvalue of the normalized covariance matrix $\tilde{\Sigma}_{g_i}$. Thus a worst case bound on the variance (which is related to the expected distance between $\mathbf{GX}_1$ and $\mathbf{GX}_2$) is $c\alpha P \lambda_{min}$. Now taking the expectation we have the worst case upper bound to the pairwise probability of error as,

$$P_e(pair) \leq \left(\frac{1}{1 + \frac{c\alpha P \lambda_{\min}}{2N_0}} \right)^{m/2}$$

For the discrete alphabet case summing over all the k sparse sequences we get for the average probability of error that

$$P_e \leq 2^{-\frac{m}{2}\log\left(1+\frac{c\alpha\rho\lambda_{min}}{2}\right)} 2^{nR_X(d_0)}$$

Thus the achievable capacity scales as

$$C_s \geq \frac{\frac{1}{2}\log(1 + \frac{c\alpha\lambda_{min}}{2})}{R_X(d_0)}$$

which in general is smaller than the case when the columns of $\mathbf{G}$ are uncorrelated, i.e. $\lambda_{min} \leq 1$. In the case when all the elements of $\mathbf{G}$ are say $\frac{1}{\sqrt{n}}$ (in the full diversity case), then $\lambda_{\min} = 0$. In this case the sensing capacity is zero.

C. Effect of Sensing diversity on Achievable Sensing Capacity

In this section we will provide achievable bounds to the sensing capacity when the diversity is β – fraction of non-zero elements in each row of G. Let $l = \beta n$ be the number of non-zero elements in each row of the matrix $\mathbf{G}$. Now consider the numerator of the exponent in the upper bound to the pair-wise probability of error -

$$\sum_{i=1}^{m} |\sum_{j=1}^{n} g_{ij}\Delta_j|^2$$

Note for each i the random variables $z = |\sum_{j=1}^{n} g_{ij}\Delta_j|^2$ are independent. We have for the average pairwise probability of error that

$$P_e \quad \leq \prod_{i=1}^{m} \mathbf{E}(e^{-|\sum_{j=1}^{n} g_{ij}\Delta_j|^2/4N_0})$$

Now $\Delta_j = \mathbf{X}_{1j} - \mathbf{X}_{2j}$. Without loss of generality let the positions in which X_1 and X_2 differ be equal to $nc\alpha$. Note that $c \leq 2$ and $\alpha \leq \frac{1}{2}$. Also g_{ij} are non-zero in l positions. Let r denote the number of overlaps of the vector $\mathbf{g}_i$ and the vector $\mathbf{X}_1 - \mathbf{X}_2$. Conditioned on r , $\sum_{j=1}^{n} g_{ij}\Delta_j$ is a Gaussian random variable with zero mean and variance $\frac{P}{l}r$. Taking the expectation with respect to this random variable we have for the upper bound on the probability of error that,

$$\begin{aligned} P_e \quad &\leq \prod_{i=1}^{m} \mathbf{E}_r \left(\frac{1}{1+\frac{Pr}{2lN_0}} \right)^{1/2} \\ &\leq \prod_{i=1}^{m} e^{-\frac{1}{2}\mathbf{E}_r \log(1+\frac{Pr}{2lN_0})} \end{aligned}$$

where the last inequality follows from Jensen's inequality. Note that $Pr(r) = \dfrac{\binom{nc\alpha}{r}\binom{n-nc\alpha}{l-r}}{\binom{n}{l}}$. Thus the achievable bound on the capacity decreases.

V. Comparison with Existing Compressed Sensing Bounds

The bounds obtained in this paper can be related to those obtained in the context of compressed sensing problem, [9], [6], [5]. There the bounds on the number of projections required for reconstructing a k-sparse signal was addressed in the framework of complexity regularized estimation of the sparse vector. The regularization term is the L_0 norm on the reconstructed signal with a weight factor of $\frac{k \log n}{m}$. They have the following result: The expected mean squared error in reconstruction is upper bounded by,

$$\mathbf{E}\left[\frac{1}{n}||\mathbf{X}-\hat{\mathbf{X}}||^2\right] \leq C_1 C_2 \frac{k\log n}{m} \quad (3)$$

where $C_1 \sim 1$ and $C_2 \sim 50(P+\sigma)2\left\{(1+p)log2+4\right\}$, under normalization of the signal and the noise power and p is the number of quantization levels, [9].

We argue that the above bound requires a normalization with respect to sparsity rate, $\alpha = k/n$ to be meaningful. This is because for an extremely sparse situation, for instance, k=const, the average distortion metric in Equation 3 implies that an all zero estimate can ensure negligible error. Specifically, (1) for $\mathbf{X}$ is extremely sparse, i.e. $\alpha << 1$ and the sparsity rate approaches zero the average distortion over the number of non-zero elements is

$$\mathbf{E}\left[\frac{1}{\alpha n}||\mathbf{X}-\hat{\mathbf{X}}||^2\right]$$

where $\alpha = \frac{k}{n}$, is the sparsity ratio. Using this as the performance metric we have from equation 3,

$$\mathbf{E}\left[\frac{1}{\alpha n}||\mathbf{X}-\hat{\mathbf{X}}||^2\right] \leq C_1 C_2 \frac{n\log n}{m} \quad (4)$$

In this case the the average number of projections required such that the per non-zero element distortion is bounded by a constant, scales as $\mathcal{O}(n\log n)$. This is indeed consistent with our sensing capacity results as well. The sensing capacity goes to zero as $\frac{1}{\log n}$. (2)$\mathbf{X}$ is sparse, i.e. $\alpha < 1$ but not very small the above normalization does not produce any significant effect. In this case since the number of non-zero elements is a fixed fraction of n, an average distortion criteria is a reasonable distortion measure. Applying Markov inequality we have that

$$Pr\left(\frac{1}{n}||\mathbf{X}-\hat{\mathbf{X}}||^2 \geq d_0\right) \leq \frac{\mathbf{E}\frac{1}{n}||\mathbf{X}-\hat{\mathbf{X}}||^2}{d_0}$$

This implies that

$$Pr\left(\frac{1}{n}||\mathbf{X}-\hat{\mathbf{X}}||^2 \geq d_0\right) \leq C_1 C_2 \frac{k\log n}{d_0 m}$$

On the other hand from the results on achievable sensing capacity we have that

$$\begin{aligned} &Pr\left(\tfrac{1}{n}||\mathbf{X}-\hat{\mathbf{X}}||^2 \geq d_0\right) \\ &\leq -\tfrac{m}{2}\log(1+d_0 SNR/2) + nR_X(d_0) \end{aligned}$$

In order to compare the results we fix a performance guarantee of ϵ, i.e., $Pr(d(\mathbf{X},\hat{\mathbf{X}}) \geq d_0) \leq \epsilon$ for comparison purposes. For a given $\epsilon > 0$, we have for the minimal number of projections required that,

$$m \geq \frac{2\left(\log(1/\epsilon) + nR_X(d_0)\right)}{\log(1+d_0 SNR/2)}$$

from our results. From results in [9] it follows that,

$$m \geq C_1 C_2 \frac{\alpha n \log n}{d_0 \epsilon}$$

For the special case of binary alphabet the number of projections based on the achievable sensing capacity is $m_1 \geq \mathcal{O}(nH_2(\alpha))$ and from results in [9] we have $m_2 \geq \mathcal{O}(\alpha n\log n)$. A plot of these orders as a function of α for a fixed n is shown in figure, 7.

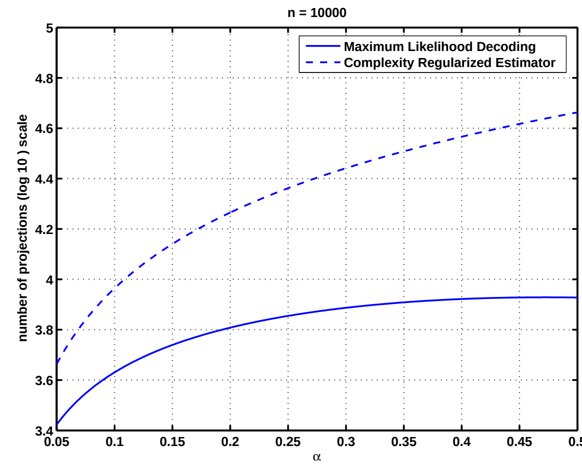

Fig. 7. The figure compares the performance in terms of scaling of the number of projections as a function of the sparsity rate for max-likelihood procedure and complexity regularized estimators of [9].

VI. Appendix

We have the following lemma for discrete alphabet case. The lemma can be extended along similar lines for continuous alphabet case by using Asymptotic Equipartition Property over the i.i.d. realization of $\mathbf{X}$. The proof then closely follows the proof for the discrete case below, where we start with a distortion error event based on $\frac{1}{n}d(\hat{\mathbf{X}}(\mathbf{Y}),\mathbf{X}) \geq d_0$ and then evaluate conditional entropy of a *rate-distortion mapping* conditioned on the error event and the observation $\mathbf{Y}$. The use of AEP allows us to separate various conditional entropies in terms of the cardinalities of the rate distortion mapping as minimal cover over the AEP set.

Lemma 6.1: For hamming distortion measure $d_H(.,.)$ and for distortion levels, $d_0 \leq (|\mathcal{X}|-1)\min_{X\in\mathcal{X}} P_X$,

$$Pr\left(\tfrac{1}{n}d_H(X^n,\hat{X}^n(\mathbf{Y}) \geq d_0)\right) \geq \frac{nR_X(d_0) - I(X^n;\mathbf{Y}) - 1}{n\log(|\mathcal{X}|) - n\left(h(d_0) + d_0\log(|\mathcal{X}|-1)\right)}$$

Proof: Given an observation $\mathbf{Y}$ about the event $\mathbf{X} \doteq X^n$. Define an error event,

$$E = \begin{cases} 1 \text{ if } \frac{1}{n}d_H(X^n,\hat{X}^n(\mathbf{Y})) \geq d_0 \\ 0 \text{ otherwise} \end{cases}$$

Expanding $H(X^n, E|\mathbf{Y})$ in two different ways we get that,

$$H(X^n|\mathbf{Y}) \leq 1 + nP_e\log(|\mathcal{X}|) + (1-P_e)H(X^n|E=0,\mathbf{Y})$$

Now the term

$$\begin{aligned}&(1-P_e)H(X^n|E=0,\mathbf{Y})\\&\leq(1-P_e)\binom{n}{d_0 n}(|\mathcal{X}|-1)^{nd_0}\\&\leq n(1-P_e)\left(h(d_0)+d_0\log(|\mathcal{X}|-1)\right)\end{aligned}$$

where $h(d_0)\doteq d_0\log_2\frac{1}{d_0}+(1-d_0)\log_2\frac{1}{1-d_0}$. Then we have for the lower bound on the probability of error that,

$$P_e\geq\frac{H(X^n|\mathbf{Y})-n\left(h(d_0)+d_0\log(|\mathcal{X}|-1)\right))-1}{n\log(|\mathcal{X}|)-n\left(h(d_0)+d_0\log(|\mathcal{X}|-1)\right)}$$

Since $H(X^n|\mathbf{Y})=H(X^n)-I(X^n;\mathbf{Y})$ we have

$$P_e\geq\frac{n\left(H(X)-h(d_0)-d_0\log(|\mathcal{X}|-1)\right)-I(X^n;\mathbf{Y})-1}{n\log(|\mathcal{X}|)-n\left(h(d_0)+d_0\log(|\mathcal{X}|-1)\right)}$$

It is known that $R_X(d_0)\geq H(X)-h(d_0)-d_0\log(|\mathcal{X}|-1)$, with equality iff

$$d_0\leq(|\mathcal{X}|-1)\min_{X\in\mathcal{X}}P_X$$

see e.g., [4]. Thus for those values of distortion we have for all n,

$$P_e\geq\frac{nR_X(d_0)-I(X^n;\mathbf{Y})-1}{n\log(|\mathcal{X}|)-n\left(h(d_0)+d_0\log(|\mathcal{X}|-1)\right)}$$

■

A. Proof of theorem 3.1

Proof: From modified Fano's inequality (see lemma 6.1 it follows that

$$P_e\geq\frac{R_{\mathbf{X}}(d_0)-I(\mathbf{X};\mathbf{Y}|\mathbf{G})-1}{n\log(|\mathcal{X}|)-n\left(h(d_0)+d_0\log(|\mathcal{X}|-1)\right)}$$

Let the average rate distortion function for $\mathbf{X}$ be $R_X(d_0)$, i.e., $R_{\mathbf{X}}(d_0)=nR_X(d_0)$. For the worst case lower bound on P_e we maximize the mutual information $I(\mathbf{X};\mathbf{Y}|\mathbf{G})$. Now note that

$$\max_{\Sigma_x:trace\Sigma_x\leq kP}I(\mathbf{X};\mathbf{Y}|\mathbf{G})=\frac{1}{2}\log det\left(I_m+\frac{k}{n}\frac{\mathbf{G}I_n\mathbf{G}^*}{N_0}\right)$$

i.e., the maximization is achieved when the input covariance matrix is an identity matrix. The normalization factor of $\alpha=\frac{k}{n}$ appears since the total power is kP. In the above expression I_n denotes $n\times n$ identity matrix. Taking expectation with respect to random realizations of $\mathbf{G}$ we have for the expected probability of error,

$$P_e\geq\frac{R_{\mathbf{X}}(d_0)-\mathbf{E}_{\mathbf{G}}\frac{1}{2}\log det\left(I_m+\frac{k}{n}\frac{\mathbf{G}I_n\mathbf{G}^*}{N_0}\right)-1}{n\log(|\mathcal{X}|)-n\left(h(d_0)+d_0\log(|\mathcal{X}|-1)\right)}$$

Since $\log det(.)$ is a concave function by Jensen's inequality we can further lower bound the probability of error by

$$P_e\geq\frac{R_{\mathbf{X}}(d_0)-\frac{1}{2}\log det\left(I_m+\frac{k}{n}\mathbf{E}_{\mathbf{G}}\frac{\mathbf{G}I_n\mathbf{G}^*}{N_0}\right)-1}{n\log(|\mathcal{X}|)-n\left(h(d_0)+d_0\log(|\mathcal{X}|-1)\right)}$$

Note that for the lower bound to be zero the numerator in the above expression should be zero. This implies

$$R_{\mathbf{X}}(d_0)-\frac{1}{2}\log det\left(I_m+\frac{k}{n}\mathbf{E}_{\mathbf{G}}\frac{\mathbf{G}I_n\mathbf{G}^*}{N_0}\right)-1\leq 0$$

Let $R_{\mathbf{X}}(d_0)=nR_X(d_0)$ where $R_X(d_0)$ is the scalar rate distortion function. Also note that $\mathbf{E}\mathbf{G}\mathbf{G}^*=I_m$. Then we get that,

$$nR_X(d_0)\leq\frac{m}{2}\log\left(1+\alpha\rho\right)+1$$

It implies,

$$\frac{n}{m}\leq\frac{0.5\log(1+\alpha\rho)+\frac{1}{n}}{R_X(d_0)}$$

Taking the limits $n,k\to\infty:k=\alpha n$ and we get the upper bound to the d_0 distortion sensing capacity.

■

References

[1] E. Candes, J. Romberg, and T. Tao, *Robust uncertainity principles: Exact signal reconstruction from highly incomplete frequency information*, IEEE Transactions on Inforamtion Theory **52** (2006), no. 2, 489–509.

[2] ———, *Stable signal recovery from incomplete and inaccuarte measurements*, Communications on Pure and Applied Mathematics **59** (2006), no. 8, 1207–1223.

[3] E. Candes and T. Tao, *Near optimal signal recovery from random projections: Universal encoding strategies?*, preprint (2004).

[4] I. Csiszár and J. Jänos Korner, *Information theory: Coding theorems for discrete memoryless systems*, Academic Press, New York, 1981.

[5] D. Donoho, *Compressed sensing*, IEEE Transactions on Information Theory **52** (2006), no. 4, 1289–1306.

[6] M.F. Duarte, M.A. Davenport, and R.G. Baraniuk, *Sparse signal detection from incoherent projections*, Intl. Conf. on Acoustics Speech and Signal Processing, May 2006.

[7] M.F. Duarte, M.B. Walkin, D. Baron, and R.G. Baraniuk, *Universal distributed sensing via random projections*, International conference on Information Processing in Sensor Networks, April 2006.

[8] E. B. Ermis and V. Saligrama, *Distributed detection of localized phenomena with limited range sensors*, submitted to IEEE Transactions on Information Theory, 2006.

[9] J. Haupt and R. Nowak, *Signal reconstruction from noisy random projections*, IEEE Transactions on Inforamtion Theory **52** (2006), no. 9, 4036–4068.

[10] M. Rabbat, J. Haupt, A. Singh, and R. Nowak, *Decentralized compression and predistribution via randomized gossiping*, International Conference on Information Processing in Sensor Networks, April 2006.

[11] Y. Rachlin, R. Negi, and P. Khosla, *Sensing capacity for discrete sensor network applications*, Proceedings of the Fourth International Symposium on Information Processing in Sensor Networks, 2005.

[12] R. Rangarajan, R. Raich, and A.O. Hero, *Sequential design of experiments for a rayleigh inverse scattering problem*, IEEE Workshop on Statistical Signal Processing (SSP), July 2005.

[13] J. A. Tropp, *Recovery of short linear combinations via l_1 minimization*, IEEE Transactions on Inforamtion Theory **51** (2005), no. 4, 1568–1570.

[14] ———, *Just relax: Convex programming methods for identifying sparse signals*, IEEE Transactions on Inforamtion Theory **51** (2006), no. 3, 1030–1051.

[15] Y. Yang and R.S. Blum, *Waveform design for mimo radar based on mutual information and minimum mean-square error estimation*, Conference on Information System and Sceinces.