AMA Statistical Information based analysis of a Compressive Imaging System

Douglas A. Hope and Sudhakar Prasad
Department of Physics and Astronomy, University of New Mexico
800 Yale Blvd.
Albuquerque, NM 87131 USA
Ph: (808) 573 9540, Fax: (808) 573-9557
dhope@.ifa.hawaii.edu

ABSTRACT

Recent advances in optics and instrumentation have dramatically increased the amount of data, both spatial and spectral, that can be obtained about a target scene. The volume of the acquired data can and, in fact, often does far exceed the amount of intrinsic information present in the scene. In such cases, the large volume of data alone can impede the analysis and extraction of relevant information about the scene. One approach to overcoming this impedance mismatch between the volume of data and intrinsic information in the scene the data are supposed to convey is compressive sensing.

Compressive sensing exploits the fact that most signals of interest, such as image scenes, possess natural correlations in their physical structure. These correlations, which can occur spatially as well as spectrally, can suggest a more natural sparse basis for compressing and representing the scene than standard pixels or voxels. A compressive sensing system attempts to acquire and encode the scene in this sparse basis, while preserving all relevant information in the scene.

One criterion for assessing the content, acquisition, and processing of information in the image scene is Shannon information. This metric describes fundamental limits on encoding and reliably transmitting information about a source, such as an image scene. In this framework, successful encoding of the image requires an optimal choice of a sparse basis, while losses of information during transmission occur due to a finite system response and measurement noise. An information source can be represented by a certain class of image scenes, .e.g. those that have a common morphology. The ability to associate the recorded image with the correct member of the class that produced the image depends on the amount of Shannon information in the acquired data. In this manner, one can analyze the performance of a compressive imaging system for a specific class or ensemble of image scenes.

We present such an information-based analysis of a compressive imaging system based on a new highly efficient and robust method that enables us to evaluate statistical entropies. Our method is based on the notion of density of states (DOS), which plays a major role in statistical mechanics by allowing one to express macroscopic thermal averages in terms of the number of configuration states of a system for a certain energy level. Instead of computing the number of microstates associated with a macroscopic energy of the system, however, we compute here the number of possible configurations (states) in the space of variables characteristic of an image scene and its observations that correspond to a certain probability value. This allows us to compute the statistical entropy of many correlated variables as an essentially one-dimensional (1D) probability integral, as we shall see presently.

We assess the performance of a single pixel compressive sensing (CS) system based on the amount of information encoded and transmitted in parameters that characterize the information in the scene. Specifically, we shall study two applications of the CS approach, namely the problem of faint companion detection and the problem of satellite material disambiguation. Here, we compute the amount of statistical information the single-pixel data convey about the essential parameters of the scene, as a function of the choice of the projective measurement basis and the amount of measurement noise. The noise creates confusion when associating the recorded data with the correct member of the ensemble that produced the image. We show that multiple measurements enable one to mitigate this confusion noise.

1. INTRODUCTION AND MOTIVATION

The ability of an electro-optical (EO) sensing system to measure and extract all relevant information about a target scene is a crucial design requirement. How one quantifies the information content in the scene depends both on the complexity of the scene itself and the task which one wishes to accomplish. One approach to quantifying the ability of an EO system to record relevant information about a target scene is the concept of statistical information as conceived by Shannon [1].

Originally conceived as a tool to investigate the problem of information capacity for a communication channel, statistical information relates the amount of information in the channel output, characterized by random vector (RV), $\underline{Y}$, that is conveyed about channel input $\underline{X}$. Due to noise and other limitations of the channel not all information about $\underline{Y}$ is passed by the channel. To compute this information loss requires one to define statistical entropy for a random variable. For a *K*- dimensional RV $\underline{X} = (x_1, ..., x_K)$ the entropy is defined as,

$$H(x_1, x_2, ..., x_k) = -\sum_{\{x_i\}} p(x_1, x_2, ..., x_k) \log p(x_1, x_2, ..., x_k). \tag{1}$$

Here, the summation is over all possible realizations of the random vector. Likewise, one computes the entropy of the channel output $\underline{Y}$, the conditional entropies, $H(\underline{X} \mid \underline{Y})$ and $H(\underline{Y} \mid \underline{X})$, and the joint entropy $H(\underline{X}, \underline{Y})$ for the conditional and joint probability distributions respectively. The amount of information successfully transmitted by the channel is quantified as the mutual information (MI),

$$\begin{aligned} I(\underline{X};\underline{Y}) &= H(\underline{X}) - H(\underline{X} \mid \underline{Y}) \\ &= H(\underline{Y}) - H(\underline{Y} \mid \underline{X}) \\ &= H(\underline{X}) + H(\underline{Y}) - H(\underline{X}, \underline{Y}). \end{aligned} \tag{2}$$

2. COMPUTING STATISTICAL ENTROPY

Although, statistical information is the basis for modern communication theory, its use in the field of imaging has been somewhat limited due to difficulties in computing the statistical entropies, particularly for large dimensions. One typically evaluates this sum by evaluating the probability value $p(x_1, x_2, ..., x_k)$ at B^K configuration space bins, where *B* defines the number of such bins along each dimension. In all except cases where *K* is small, this evaluation becomes computationally intractable (e.g. when (K=5 and *B*=100) the entropy computation entails 10^{10} terms).

We propose a method for overcoming this limitation. Our approach reformulates the expression for entropy given by Eq. (1) from a B^K fold sum into a 1D sum problem, thus yielding a more efficient and robust computation of the needed statistical entropies especially for large values of *K*. Our formulation is motivated by statistical mechanics, a field of physics that has been highly successful in deriving certain thermodynamic quantities and relations of interest for physical systems that are comprised of a large number of degrees of freedom (DOF).

We make essential use of the notion of density of states (DOS). In statistical mechanics, the DOS of a physical system is defined as the number of microstates, or quantum states, per unit energy interval. For our problem, let us define $\Omega(p)$ as the total number of *K*-dimensional configuration space bins for which the probability value is p or larger. The density of states (DOS) is the number of states (NOS) per unit probability interval,

$$\omega_h(p) = -\frac{d\Omega_h(p)}{dp}. \tag{5}$$

Such a formulation allows us to recast Eq. (1.1) in terms of the NOS,

$$H(x_1, x_2, ..., x_k) = -\sum_{i=1}^{N_p} \Delta\Omega_i \, p_i \log p_i. \tag{6}$$

This transforms the K-dimensional summation problem into a 1-dimensional sum over the N_p probability values where $\Delta\Omega_i$ denotes the number of configuration space bins (or possible states of the system) corresponding to the i[th] probability interval $(p_i, p_i + dp_i)$. We then implement a Monte-Carlo method for computing $\Delta\Omega_i$ [2]. Specifically we define a domain of volume *V* in the configuration space and throw *N* random darts uniformly over this volume. At each dart location we compute the probability value and count the number of darts falling in each of the probability bins, denoted by N_i, for the i[th] bin.

The estimated volume of configuration space corresponding to the i^{th} probability bin is $\widehat{V}_i = \dfrac{N_i}{N}$ V and the estimated NOS is,

$$\Delta\Omega_i = \frac{\widehat{V}_i}{\Delta^K} \tag{7}$$

where Δ^K denotes the volume of each configuration space bin. As a demonstration of our DOS approach to computing statistical entropies we show in Fig. 1 the numerically computed value for $\Delta\Omega_i \, p_i \log p_i$ plotted against $\log_{10}(p_i)$ for the Gaussian PDF with dimensions *K=1, 2, 4, 8,* and *16*.

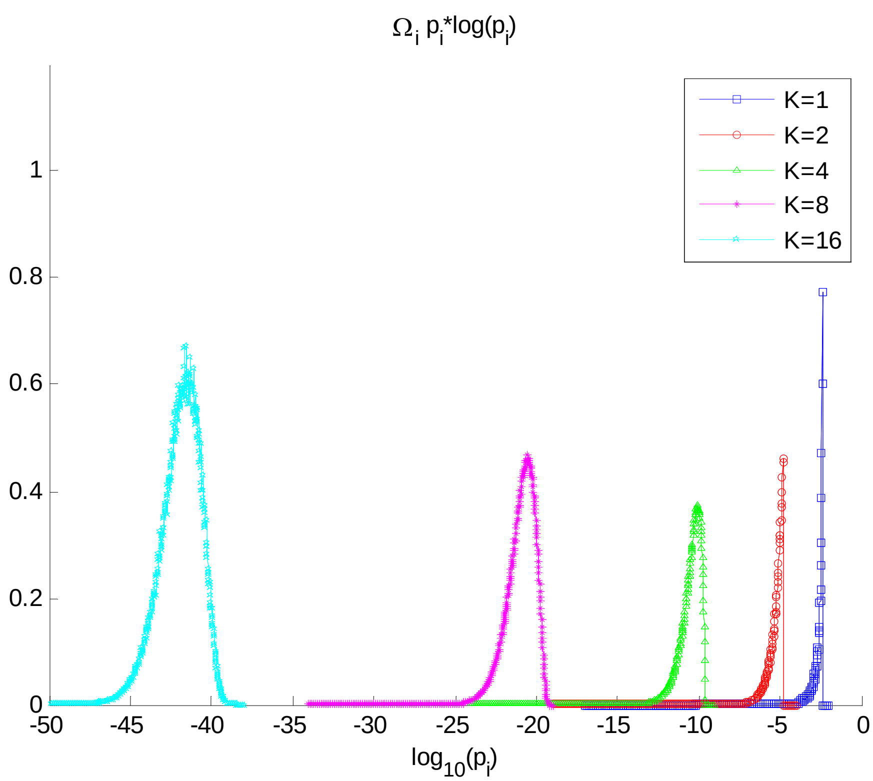

Figure 1 Shown are the product of the (NOS) term $\Delta\Omega_i$ and the $p_i \log p_i$ vs the dimensionality of the configuration space.

The location of the peak of this product shows the probability bins with the greatest contribution to the entropy computation. For the Gaussian case it is possible to derive an analytic expression [3] for the scaling of the peak location p^* and its width δp in the limit of large *K*,

$$p^* = p_{\max} \exp(-K/2) \text{ and } \delta p = D^{1/2} p^* \quad (8)$$

where,

$$p_{\max} = \frac{\Delta^K}{(2\pi)^{K/2} (\det \mathbf{C})^{1/2}} \quad (9)$$

and C is the covariance matrix for the Gaussian RV. As a check we plot the numerically computed ratio $p^* / p_{\max}$ against the analytic result for the Gaussian PDF in Fig 2.
Even for dimensions as small as K=2, our numerical results agree with analytic predictions of the scaling of p*and give confidence in our DOS approach for computing entropies for large numbers of random variables. The results shown here were computed on a desktop computing environment. Further numerical computations using 100 nodes on a supercomputer enabled us to compute the entropy for up to K=24 variables. With our efficient approach to compute statistical entropies we turn our attention to computing mutual information quantities for image data.

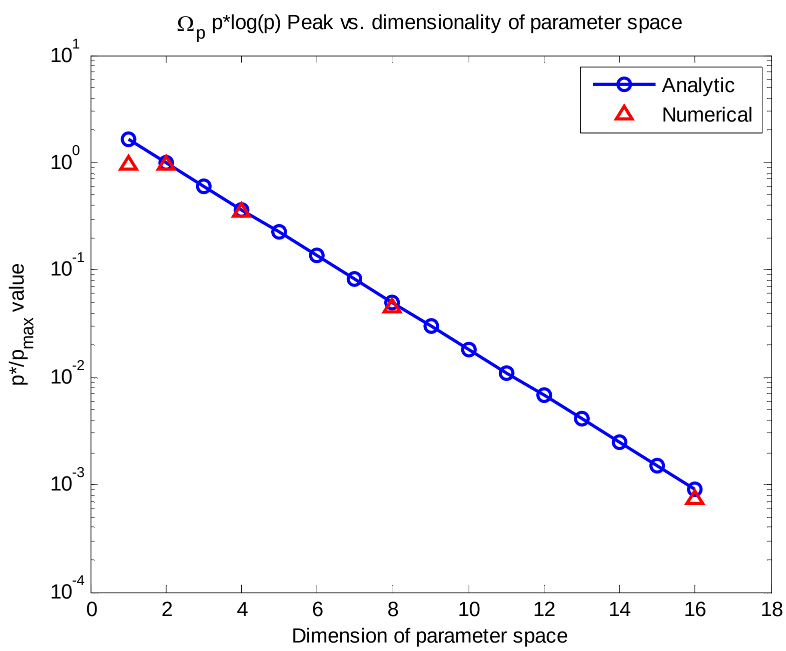

Figure 2 Shown on this log plot is the exact value of p* (the location of the maximum in probability space of terms in the sum (6)) for the Gaussian PDF (blue circles) verses the computed p* using our DOS approach

3. EXTRACTION OF STATISTICAL INFORMATION FROM IMAGE DATA

Due to the many advances in optics and instrumentation, current state of the art imaging systems can acquire vast amounts of spatial and spectral data about a target scene. In fact, in some cases the sheer volume data can impede the analysis and extraction of key parameters of interest about the target scene. One approach to obtaining only the intrinsic information in the target scene is the concept of compressive sensing. In this imaging strategy one attempts to more or less directly extract the intrinsic parameters of interest. How well such a system accomplishes this task can be analyzed in terms of statistical information, which we now turn our attention to.

We consider two relevant problems in space surveillance. The detection and identification of faint targets near a brighter well known target and the detection of a material boundary on a target scene. In the former problem the parameters of interest are the total flux, location, width and its shape, while in the latter they are the location of the material boundary and the disparity between the brightnesses of pixels belonging to different materials. Accurate estimation of these parameters in the presence of noise is critical to successfully accomplishing the detection and identification tasks. Shannon's statistical information theory can provide insight on how much information about these parameters can be obtained by a CS system.

4. PROBLEM OF FAINT COMPANION DETECTION

For the problem of faint companion detection the task of interest is the extraction of the intensity of the relatively faint companion of a relatively bright primary star from CS data that are K single-pixel projective measurements using a set of K image-plane masks.

We describe each of the two sources comprising the target scene spatially by the functions, $w^0(\mathbf{x})$ and $w^1(\mathbf{x})$, with the superscripts 0 and 1 referring to the brighter component and the fainter companion respectively. We take these functions to be normalized such that their integrals over the target scene are unity,

$$\int w^{(i)}(\mathbf{x})d\mathbf{x} = 1, \quad i = 0,1. \tag{10}$$

If the integrated source fluxes are I_0 and I_1, then the spatial brightness distribution for the target scene may be expressed as,

$$f(\mathbf{x}) = I_0 w^{(0)}(\mathbf{x}) + I_1 w^{(1)}(\mathbf{x}), \tag{11}$$

where $\mathbf{x}$ denotes the position vector. In our model for computing mutual information we shall take our source, image, and mask profiles to be pixelated, taking discrete values, $x_n = n\Delta x, n = 1,\ldots,N$ over an N-pixel image field, Δx being the pixel size. In this sense, the prior continuous integrals are to be understood as discrete sums, namely $\sum_n w_n^{(i)} = 1$ with the pixel size Δx absorbed inside the discrete variables $w_n^{(i)}$.

The basis of CS is the choice of an optimal measurement basis in which one acquires all relevant information about the target scene [3,4,5]. This information is encoded by K projective measurements of the scene $f(\mathbf{x})$ on to the set of K basis vectors $\{r_k(\mathbf{x})\}$,

$$\begin{aligned} m_k &= \sum_{\mathbf{x}} f(\mathbf{x}) r_k(\mathbf{x}) + n_k \\ &= I_0 p_k^{(0)} + I_1 p_k^{(1)} + n_k \end{aligned} \tag{12}$$

where the quantities $p_k^{(0)}$ and $p_k^{(1)}$ are the overlap integrals between the k-th mask function and the primary and companion source profiles,

$$p_k^{(0)} = \sum_{\mathbf{x}} w^{(0)}(\mathbf{x}) r_k(\mathbf{x}); \quad p_k^{(1)} = \sum_{\mathbf{x}} w^{(1)}(\mathbf{x}) r_k(\mathbf{x}). \tag{13}$$

The quantity n_k represents the detection noise associated with the measurement, which we take to be independently identically distributed (IID) according to a zero-mean Gaussian PDF with variance equal to σ_N^2 . With this assumption we implicitly assume that the images are reasonable faint overall such that the additive noise of detection dominates the shot noise of photon counting. Furthermore, we assume that the imaging system is perfect with an infinitely sharp PSF, so that the image function prior to detection is an exact reproduction of the target intensity distribution $f(\mathbf{x})$. The set of functions $\{r_k(\mathbf{x})\}$ denote a set of basis vectors onto which the target scene is projected. We chose these vectors based on a central result of CS, that the optimal choice of basis vectors, when little is known a priori about the target scene, is a set of vectors where each component is a realization of a uniform random number drawn repeatedly in an independent, identically distributed (IID) manner [4]. Alternatively, one can define a random mask in the Fourier domain

$$r_k(\mathbf{x}) = \sum_{\mathbf{u}} R_k(\mathbf{u}) \exp\left(\frac{2\pi i}{N}\mathbf{x}\cdot\mathbf{u}\right). \tag{14}$$

Here, the function $R_k(\mathbf{u})$ is the Fourier amplitude which for each value of k, is equal to 1 at a single spatial frequency chosen between DC and the highest spatial frequency and zero at the remaining frequencies.
We consider both types of random masks, namely random values for pixels chosen in the image domain and randomly chosen sinusoids in the Fourier domain.

5. PROBLEM OF MATERIAL BOUNDARY DETECTION

We next consider the problem of extracting information from CS measurements of a two material component satellite. We model each component of the satellite using the profiles, $w^0(\mathbf{x})$ and $w^1(\mathbf{x})$, which are of lengths α and L-α where L is the length of the satellite which is constrained. The parameters of interest in this problem are the location of the boundary, α, and the flux difference of the two components $\Delta i = i_1 - i_2$. Total flux of the satellite F_0 is constrained such that,

$$i_1\alpha + i_2(L-\alpha) = F_0. \tag{15}$$

The projective measurements recorded by the CS system are,

$$m_k = i_1\sum_{j=1}^{\alpha} r_j^{(k)} + i_2\sum_{j=\alpha+1}^{L} r_j^{(k)} + n^{(k)} \tag{16}$$

Where the measurements are now expressed in terms of the overlap integrals, which are functions of the nonlinear parameter α.

6. RESULTS: FAINT COMPANION DETECTION

In our analysis of the information content of projective measurements for the faint companion problem we assume the flux I_0 of the brighter primary source is well known and that also the profile of the source, $w_0(\mathbf{x})$, is known. We assume the companion profile $w_1(\mathbf{x})$ is known (a single pixel) but its separation distance of the companion from the primary source α is modeled as a uniform random variable. The companion flux I_1 is taken to be distributed statistically according to the negative exponential PDF, with mean $\langle I_1 \rangle$. We define the signal-to-noise ratio (SNR) of the measured data as $\frac{\langle I_1 \rangle}{\sigma_N}$.

In our analysis of the problem of faint companion detection we consider two types of target morphologies, one where the companion has a profile width of 1 pixel and where the profile is extended over 8 pixels. We compute using a forward CS model the amount of information in the parameters, I_1 and α, for different numbers of measurements using random masks defined in both the image domain and in the Fourier domain.

We plot the computed mutual information in the measurements in terms of fractional information recoverable from the measurements, namely

$$f_{I_1} = \frac{I(\mathbf{m} \mid I_1)}{H(I_1)} \text{ and } f_\alpha = \frac{I(\mathbf{m} \mid \alpha)}{H(\alpha)} \quad . \tag{17}$$

The choice of an optimal measurement basis depends on the sparse nature of the target scene. Scenes that are sparse in their native pixel representation, such as our faint companion problem, should be best sampled using a random pixel mask defined in the image domain. An example of this result is shown in Fig. 3 where the companion, with a width of 1 pixel, has a sparse representation in the pixel basis. The fractional MI recovered for the companion flux is comparable; while for the position parameter α, the MI recovered using the pixel mask is slightly higher than the case where the image scene is sampled on random sinusoids.

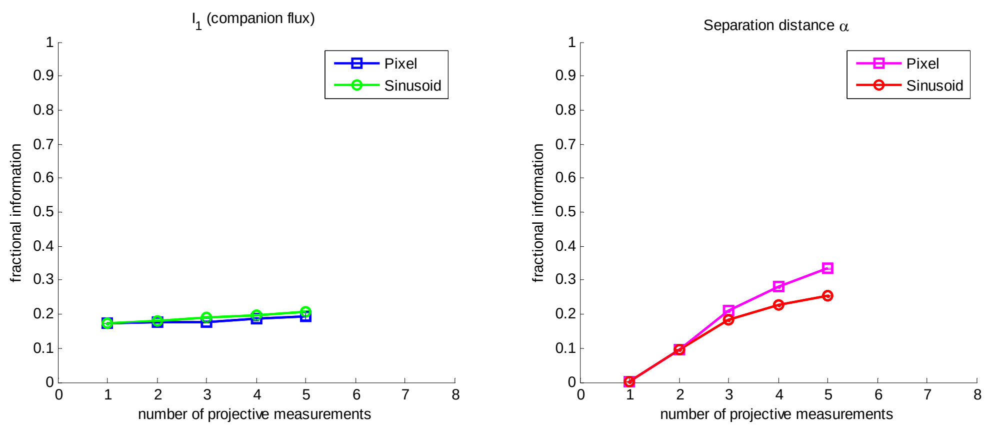

Figure 3 Results for faint companion detection where profile width of companion is 1 pixel. Left is the fractional information about I_1 in the projective measurements. Right: The fractional information about the position parameter α of the companion in the projective measurements. (SNR=5)

As the width of the faint companion becomes larger, i.e., its representation in the pixel basis becomes less sparse, we expect the random Fourier-based masks to outperform the random pixel masks defined in the image domain. In Fig. 4 we show results for a broader companion width. Here, the fractional information recovered in the parameter I_1 for both masks is approximately the same. However, for the nonlinear companion position parameter ∝ the Fourier-based mask outperforms the random image pixel mask. The random-pixel mask produces simply too much scatter in the measurement vector in such cases to yield comparable information about the position variable.

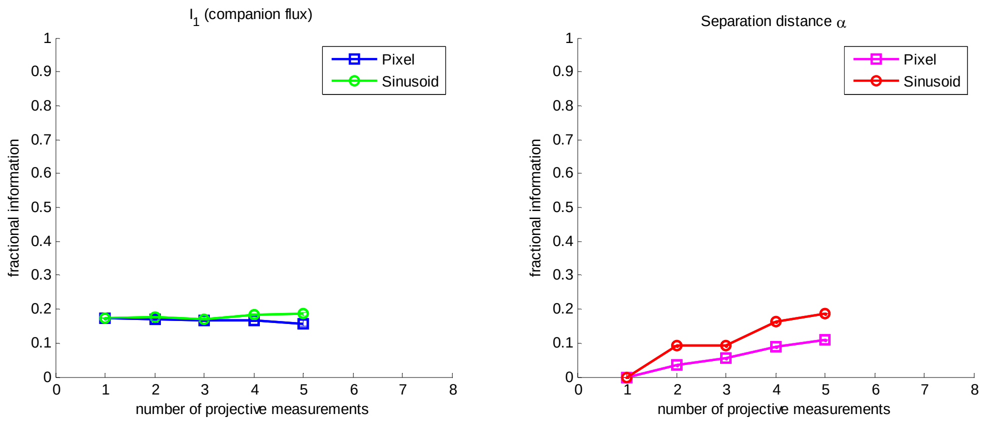

Figure 4 Results for faint companion detection where profile width of companion is 8 pixels. Left is the fractional information about I_1 in the projective measurements. Right: The fractional information in the projective measurements about the position parameter α of the companion. SNR=5

7. RESULTS: MATERIAL BOUNDARY DETECTION

Another problem we consider is the problem of detecting a material boundary on a 1-d satellite model, that consists of two different material components. The location of the boundary is parameterized by α which is drawn from a uniform distribution. The flux difference $\Delta i = i_1 - i_2$ of the two components is assumed to have a Gaussian distribution with zero mean and $\sigma_{\Delta i} = 1$. The SNR of the measured data depends on both the length *L* of the satellite and the number *K* of measurements,

$$SNR = \frac{\sigma_{\Delta i}}{\sigma_N}\sqrt{\frac{LK}{12}}\,. \tag{18}$$

As before we define the fractional information as $f_{\Delta i} = \frac{I(\mathbf{m} \mid \Delta i)}{H(\Delta i)}$. Based on our analysis of the binary problem, the material boundary problem does not have a sparse representation in the natural pixel basis. For such morphology the Fourier-based random mask is expected to be a better choice of basis for recording the measurements since it tends to average systematically, rather than in randomly noisy fashion, over the uniformly bright pixels while still being sensitive to the location of the boundary. The SNR is approximately 3 for this problem.

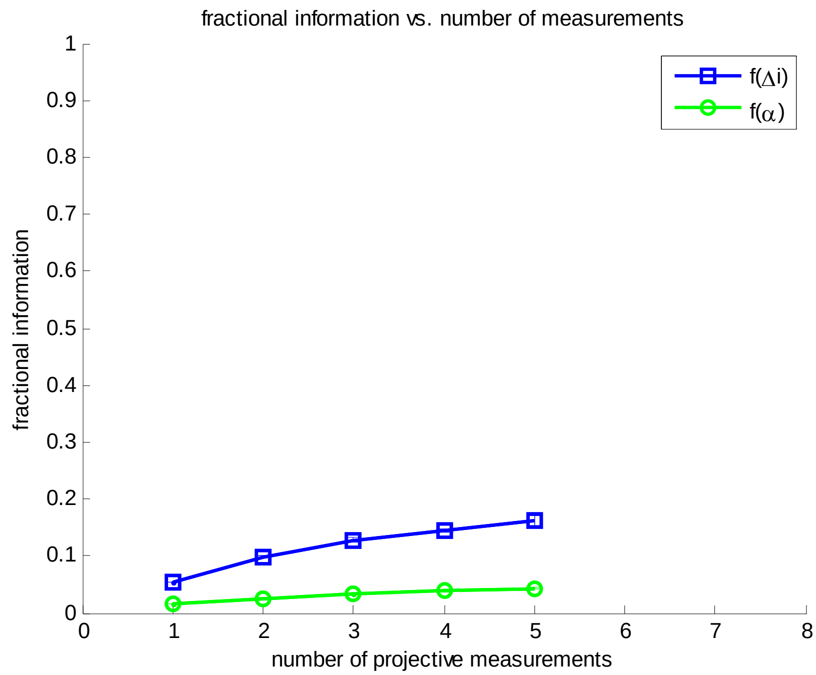

Figure 3 Plotted is the fractional information recovered for parameters Δi and α for the material boundary problem using Fourier masks.

8. CONCLUSIONS

We have presented results from a new numerical computational technique for computing statistical entropies for random variables of high dimension. This capability has enabled us to study the performance of a compressive sensing system from an information-based viewpoint. We have shown how the choice of an optimal measurement basis depends on the morphology of the target scene. In scenes which are highly sparse in the natural pixel basis, except for the limiting case of an unresolved (one-pixel) object, the random masks defined in the image plane tend to yield poorer results than the random Fourier-based masks. In contrast, for scenes that have very low sparsity in the pixel basis, the random-pixel masks may yield better results.

A heuristic argument for this result is that for scenes with a highly sparse representation in the pixel basis, the random pixel masks are best able to capture information about these different sparse components of the scene. For scenes with highly extended features over many pixels, the projection of these features onto the random masks can greatly reduce the value of the overlap integrals in Eq. 13 but the scatter of these values about the expected value of 0 may still provide a CS signal vector that potentially enables recovery of scene information. By using random sinusoids in such cases, where in the image domain pixel values of the random mask vary more smoothly, the overlap integrals tend to vanish in a systematic fashion. However in the intermediate-sparsity regime, such as the faint companion problem where the component width was 8 pixels, the Fourier-based random masks performs better than the pixel-based mask. For the satellite material-boundary problem with uniform intensity over extended domains interrupted only by boundary steps, Fourier masks of any finite frequency tend to average out the contribution of the uniform-intensity pixels while preserving information about these sharp intensity steps. For this reason, we expect such masks to outperform other masks, including random-pixel masks, for this problem.

Acknowledgement: Research supported by AFOSR award FA9550-09-1-0495

9. BIBLIOGRAPHY

1. *C. Shannon, The Bell System Technical Journal,Vol. 27, pp. 379–423, 623–656, July, October, 1948.*
2. *Metropolis, N.; Ulam, S. (1949). "The Monte Carlo Method". Journal of the American Statistical Association* ***44*** *(247): 335–341.*
3. *S. Prasad, D. Hope, "A Dimensionally Robust Numerical Approach for Computing Statistical Entropies" (in preparation)*
4. *E. Candes, K. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," IEEE Trans. Inform. Theory, vol. 52, no. 2, pp. 489-509 (2006)*
5. *E. Candes, K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," Comm. Pure Appl. Math, vol. 59, no. 8, pp. 1207-1223 (2006)*
6. *E. Candes and T. Tao, "Near-optimal signal recovery from random projections and universal encoding strategies," IEEE Trans. Inform. Theory, vol. 52, no. 2, pp. 489-509 (2006)*