

**STATISTICAL COMPRESSIVE SENSING OF  
GAUSSIAN MIXTURE MODELS**

By

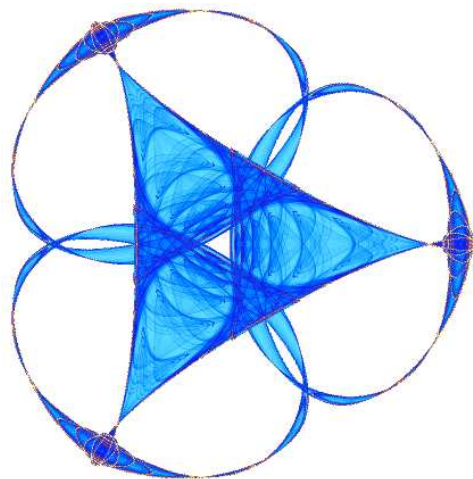
**Guoshen Yu**

and

**Guillermo Sapiro**

**IMA Preprint Series # 2342**

(October 2010)



**INSTITUTE FOR MATHEMATICS AND ITS APPLICATIONS**

UNIVERSITY OF MINNESOTA  
400 Lind Hall  
207 Church Street S.E.  
Minneapolis, Minnesota 55455-0436

Phone: 612-624-6066 Fax: 612-626-7370

URL: <http://www.ima.umn.edu>

# STATISTICAL COMPRESSIVE SENSING OF GAUSSIAN MIXTURE MODELS

Guoshen Yu and Guillermo Sapiro

ECE, University of Minnesota, Minneapolis, MN 55455, U.S.A.

## ABSTRACT

A new framework of compressive sensing (CS), namely *statistical compressive sensing* (SCS), that aims at efficiently sampling a *collection* of signals that follow a statistical distribution and achieving accurate reconstruction *on average*, is introduced. For signals following a Gaussian distribution, with Gaussian or Bernoulli sensing matrices of  $\mathcal{O}(k)$  measurements, considerably smaller than the  $\mathcal{O}(k \log(N/k))$  required by conventional CS, where  $N$  is the signal dimension, and with an optimal decoder implemented with linear filtering, significantly faster than the pursuit decoders applied in conventional CS, the error of SCS is shown tightly upper bounded by a constant times the  $k$ -best term approximation error, with overwhelming probability. The failure probability is also significantly smaller than that of conventional CS. Stronger yet simpler results further show that for *any* sensing matrix, the error of Gaussian SCS is upper bounded by a constant times the  $k$ -best term approximation with probability one, and the bound constant can be efficiently calculated. For signals following Gaussian mixture models, SCS with a piecewise linear decoder is introduced and shown to produce for real images better results than conventional CS based on sparse models.

**Index Terms**— Compressive sensing, Gaussian mixture models

## I. INTRODUCTION

Compressive sensing (CS) attempts to achieve accurate signal reconstruction while sampling signals at a low sampling rate, typically far smaller than the Nyquist/Shannon rate. Let  $\mathbf{x} \in \mathbb{R}^N$  be a signal of interest,  $\Phi \in \mathbb{R}^{M \times N}$  a *non-adaptive* sensing matrix (*encoder*), consisting of  $M \ll N$  measurements,  $\mathbf{y} = \Phi \mathbf{x} \in \mathbb{R}^M$  a measured signal, and  $\Delta$  a *decoder* used to reconstruct  $\mathbf{x}$  from  $\Phi \mathbf{x}$ . CS develops encoder-decoder pairs  $(\Phi, \Delta)$  such that a small reconstruction error  $\mathbf{x} - \Delta(\Phi \mathbf{x})$  can be achieved.

Reconstructing  $\mathbf{x}$  from  $\Phi \mathbf{x}$  is an ill-posed problem whose solution requires some prior information (model) on the signal. Instead of the frequency band-limit signal model assumed in classic Shannon sampling theory, conventional CS adopts a sparse signal model (manifold models have been considered as well [4], [8]), i.e., there exists a dictionary, typically an orthogonal basis  $\Psi \in \mathbb{R}^{N \times N}$ , a linear combination of whose columns generates an accurate approximation of the signal,  $\mathbf{x} \approx \Psi \mathbf{a}$ , with the coefficients  $\mathbf{a}[m]$ ,  $1 \leq m \leq N$ , having their amplitude decaying fast after being sorted. For signals following the sparse model, it has been shown that using some random sensing matrices such as Gaussian and Bernoulli matrices  $\Phi$  with  $M = \mathcal{O}(k \log(N/k))$  measurements, and an  $l_1$  minimization or a greedy matching pursuit decoder  $\Delta$  promoting sparsity, with high probability CS leads to accurate signal reconstruction. The obtained approximation error is tightly upper bounded by a constant times the  $k$ -best term approximation error, the minimum error that one may achieve by keeping the  $k$  largest coefficients in  $\mathbf{a}$  [6], [9], [10].

The present paper introduces a novel framework of CS, namely *statistical compressive sensing* (SCS). As opposed to conventional CS that deals with one signal at a time, SCS aims at efficiently sampling a collection of signals and having accurate reconstruction on average. Instead of restricting to sparse models, SCS works with general Bayesian models. Assuming that the signals  $\mathbf{x}$  follow a distribution with probability density function (pdf)  $p(\mathbf{x})$ , SCS designs encoder-decoder pairs  $(\Phi, \Delta)$  so that the average error

$$E_{\mathbf{x}} \|\mathbf{x} - \Delta(\Phi \mathbf{x})\|_X = \int \|\mathbf{x} - \Delta(\Phi \mathbf{x})\|_X p(\mathbf{x}) d\mathbf{x}, \quad (1)$$

where  $\|\cdot\|_X$  is a norm, is small. As an important example, SCS with Gaussian models is here shown to have improved performance (bounds) relative to conventional CS, the signal construction calculated with an optimal decoder  $\Delta$  implemented via a fast linear filtering. Moreover, for

Gaussian mixture models (GMM), SCS with a piecewise linear decoder is introduced and shown to be very effective.

The motivation of SCS with Gaussian models is twofold. First, controlling the average error over a collection of signals is useful in signal acquisition, not only because one is often interested in acquiring multiple signals in real applications, but also because more effective processing of an individual signal, an image or a sound for example, is usually achieved by dividing the signal in local subparts, patches or short-time windows for instance, and a signal can be regarded as a collection of subpart signals [17], [18]. In addition, Gaussian mixture models (GMM), which model signals or subpart signals with a collection of Gaussians, have been shown effective in describing real signals, leading to excellent results in image inverse problems [18] and missing data estimation [13].

After reviewing the optimal decoder for Gaussian signals in Section II, a quick numerical check of the good performance of Gaussian SCS is first given in Section III. Section IV analyzes this performance following a similar mathematical approach as the one adopted in conventional CS performance analysis [9]. This result shows that with the same random matrices as in conventional CS, but with a considerably reduced number  $M = \mathcal{O}(k)$  of measurements, and with the optimal decoder implemented with linear filtering, significantly faster than the decoders applied in conventional CS, the average error of Gaussian SCS is tightly upper bounded by a constant times the  $k$ -best term approximation error with overwhelming probability, the failure probability being orders of magnitude smaller than that of conventional CS. Section V further shows stronger yet simpler results: For *any* sensing matrix, the average error of Gaussian SCS is upper bounded by a constant times the  $k$ -best term approximation with probability one, and the bound constant can be efficiently calculated. Section VI introduces, for SCS with GMM, a piecewise decoding scheme based on an efficient maximum a posteriori expectation-maximization (MAP-EM) algorithm following [18], and shows better results on real images than those obtained with conventional CS.

In the rest of this paper, we will assume without loss of generality Gaussian signal models  $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{S})$ , with mean zero and diagonal covariance matrix  $\mathbf{S} = \text{diag}[\lambda_1, \dots, \lambda_N]$ , with  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N$  the sorted eigenvalues. We can always center the data with respect to the Gaussian distribution and make a basis change with principal component analysis (PCA) [18]. For Gaussian and Bernoulli matrices that are known to be universal, analyzing CS in the canonical basis or the PCA basis is equivalent [2]. The Gaussian distributions are assumed to be full rank, i.e.,  $\lambda_N > 0$ , since a degenerated Gaussian can be regarded as a full-rank Gaussian within a reduced dimension.

## II. OPTIMAL DECODER FOR GAUSSIAN SCS

It is well-known that the optimal decoders for Gaussian signals are calculated with linear filtering:

**Theorem 1.** [12] Let  $\mathbf{x} \in \mathbb{R}^N$  be a random vector with prior pdf  $\mathcal{N}(\mathbf{0}, \mathbf{S})$ , and  $\Phi \in \mathbb{R}^{M \times N}$ ,  $M \leq N$ , be a sensing matrix. From the measured signal  $\mathbf{y} = \Phi \mathbf{x} \in \mathbb{R}^M$ , the optimal decoder  $\Delta$  that minimizes the mean square error (MSE)  $E_{\mathbf{x}}[\|\mathbf{x} - \Delta(\Phi \mathbf{x})\|_2^2] = \min_f E_{\mathbf{x}}[\|\mathbf{x} - f(\Phi \mathbf{x})\|_2^2]$ , as well as the mean absolute error (MAE)  $E_{\mathbf{x}}[\|\mathbf{x} - \Delta(\Phi \mathbf{x})\|_1] = \min_f E_{\mathbf{x}}[\|\mathbf{x} - f(\Phi \mathbf{x})\|_1]$ , where  $f$  is any mapping from  $\mathbb{R}^M \rightarrow \mathbb{R}^N$ , is obtained with a linear MAP estimator,

$$\Delta(\Phi \mathbf{x}) = \arg \max_{\mathbf{x}} p(\mathbf{x}|\mathbf{y}) = \mathbf{S} \Phi^T (\Phi \mathbf{S} \Phi^T)^{-1} (\Phi \mathbf{x}), \quad (2)$$

and the resulting error  $\boldsymbol{\eta} = \mathbf{x} - \Delta(\Phi \mathbf{x})$  is Gaussian with mean zero and with covariance matrix  $\Sigma_{\boldsymbol{\eta}} = E_{\mathbf{x}}[\boldsymbol{\eta} \boldsymbol{\eta}^T] = \mathbf{S} - \mathbf{S} \Phi^T (\Phi \mathbf{S} \Phi^T)^{-1} \Phi \mathbf{S}$ , whose trace yields the MSE of SCS

$$E_{\mathbf{x}}[\|\mathbf{x} - \Delta(\Phi \mathbf{x})\|_2^2] = \text{Tr}(\mathbf{S} - \mathbf{S} \Phi^T (\Phi \mathbf{S} \Phi^T)^{-1} \Phi \mathbf{S}). \quad (3)$$

In contrast to conventional CS, for which the  $l_1$  minimization or greedy matching pursuit decoders, calculated with iterative procedures, have been shown optimal under certain conditions on  $\Phi$  and the signal sparsity [6], [10], Gaussian SCS enjoys the advantage of having an optimal decoder (2) calculated via a *linear* filtering for any  $\Phi$ .

**Corollary 1.** *If a random matrix  $\Phi \in \mathbb{R}^{M \times N}$  is drawn independently to sense each  $\mathbf{x}$ , with all the other conditions as in Theorem 1, the MSE of SCS is*

$$E_{\mathbf{x}, \Phi} \|\mathbf{x} - \Delta(\Phi \mathbf{x})\|_2^2 = E_{\Phi} [\text{Tr}(\mathbf{S} - \mathbf{S}\Phi^T(\Phi\mathbf{S}\Phi^T)^{-1}\Phi\mathbf{S})]. \quad (4)$$

The MSE of Gaussian SCS has closed-forms (3), (4), however, a mathematical analysis of (3) and (4) seems uneasy due to the complexity of the involved matrix expression. The next section evaluates these values through Monte Carlo simulations, preceding the theoretical bounds later developed.

### III. PERFORMANCE OF GAUSSIAN SCS – A NUMERICAL ANALYSIS AT FIRST

This section numerically evaluates the MSE of Gaussian SCS, and compares it with the minimal MSE generated by the best  $k$ -term approximation. Under the Gaussian signal model  $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{S})$  with sorted eigenvalues  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N$ , it is well-known that the best  $k$ -term approximation error,

$$\sigma_k(\{\mathbf{x}\})_X = E_{\mathbf{x}} \|\mathbf{x} - \mathbf{x}_K\|_X \quad \text{and} \quad \sigma_k(\{\mathbf{x}\})_2^2 = E_{\mathbf{x}} \|\mathbf{x} - \mathbf{x}_K\|_2^2, \quad (5)$$

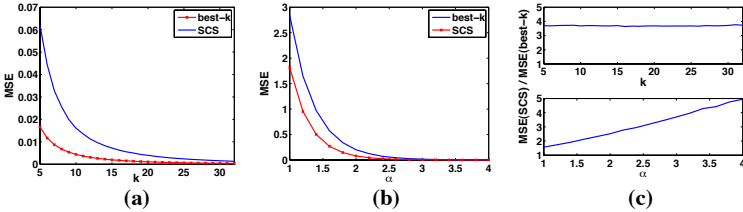
is obtained with a *linear* projection to the first  $k$  entries, i.e.,  $\mathbf{x}_K[n] = \mathbf{x}[n]$ ,  $\forall n \in K = \{1, \dots, k\}$ , and  $\mathbf{x}_K[n] = 0$  otherwise [14]. Note that  $\sigma_k(\{\mathbf{x}\})_2^2 = \sum_{m=k+1}^N \lambda_m$ . This best  $k$ -term approximation sensing is impractical with GMM, since the Gaussian assignment of a signal is unknown at the acquisition moment, GMM describing real data much better than a single Gaussian model [18].

A power decay of the eigenvalues [14],

$$\lambda_m = m^{-\alpha}, \quad 1 \leq m \leq N, \quad (6)$$

$N = 64$ , is assumed in the Monte Carlo simulation. An independent random Gaussian matrix realization  $\Phi$  is applied to sense each signal  $\mathbf{x}$  [9], and (4) is used to calculate the MSE of SCS.

Figures 1 (a) and (c)-top plot the MSE of SCS and that of the best  $k$ -term approximation, as well as their ratio as a function of  $k$  for Gaussian signals with a typical eigenvalue decay  $\alpha = -3$ . With  $k$  increasing, both MSEs decrease, their ratio being almost constant at about 3.7. The same is plotted in figures 1 (b) and (c)-bottom, with  $k$  fixed at a typical value of 10, and  $\alpha$  varying from 1 to 4. When  $\alpha$  increases, the eigenvalues decay faster, and the MSEs for both methods decrease. The ratio increases almost linearly with  $\alpha$ .



**Fig. 1.** Comparison of the MSE of SCS and that of the best  $k$ -term approximation for Gaussian signals. See text for details.

These results indicate a good performance of Gaussian SCS, its MSE is only a small number of times larger than that of the best  $k$ -term approximation. The next sections provide mathematical analysis of this performance.

### IV. PERFORMANCE BOUNDS OF GAUSSIAN SCS

Following [9], this section shows that with Gaussian and Bernoulli random matrices of  $\mathcal{O}(k)$  measurements, considerably smaller than the  $\mathcal{O}(k \log(N/k))$  required by conventional CS, the average error of Gaussian SCS is tightly upper bounded by a constant times the  $k$ -best term approximation with overwhelming probability, the failure probability being orders of magnitude smaller than that of conventional CS. The proofs of the theorems will not be given due to the space limitations. We consider only the encoder-decoder pairs  $(\Phi, \Delta)$  that preserve  $\Phi \mathbf{x}$ , i.e.,  $\Phi(\Delta(\Phi \mathbf{x})) = \Phi \mathbf{x}$ , satisfied by the optimal  $\Delta$  in (2) for Gaussian signals  $\mathbf{x}$ ,  $\forall \Phi$ .

### IV-A. From Null Space Property to Instance Optimality

The *instance optimality in expectation* bounds the average error of SCS with a constant times that of the best  $k$ -term approximation (5), defining the desired SCS performance:

**Definition 1.** *We say that  $(\Phi, \Delta)$  is instance optimal in expectation of order  $k$  in  $\|\cdot\|_X$ , with a constant  $C_0$ , if*

$$E_{\mathbf{x}, (\Phi)} \|\mathbf{x} - \Delta(\Phi \mathbf{x})\|_X \leq C_0 \sigma_k(\{\mathbf{x}\})_X, \quad (7)$$

the expectation with respect to  $\mathbf{x}$ , and to  $\Phi$  if one random  $\Phi$  is drawn independently for each  $\mathbf{x}$ . Similarly, the MSE instance optimality of order  $k$  is defined as

$$E_{\mathbf{x}, (\Phi)} \|\mathbf{x} - \Delta(\Phi \mathbf{x})\|_2^2 \leq C_0 \sigma_k(\{\mathbf{x}\})_2^2. \quad (8)$$

The *null space property in expectation* defined next will be shown equivalent to the instance optimality in expectation.

**Definition 2.** *We say that  $\Phi$  in  $(\Phi, \Delta)$  has the null space property in expectation of order  $k$  in  $\|\cdot\|_X$ , with constant  $C$ , if*

$$E_{\mathbf{x}, (\Phi)} \|\eta\|_X \leq C \sigma_k(\{\eta\})_X, \quad \forall \eta = \mathbf{x} - \Delta(\Phi \mathbf{x}), \quad (9)$$

the expectation with respect to  $\mathbf{x}$ , and to  $\Phi$  if one random  $\Phi$  is drawn independently for each  $\mathbf{x}$ . Note that  $\eta \in \text{Null}(\Phi)$ . Similarly, the MSE null space property of order  $k$  is defined as

$$E_{\mathbf{x}, (\Phi)} \|\eta\|_2^2 \leq C \sigma_k(\{\eta\})_2^2, \quad \forall \eta = \mathbf{x} - \Delta(\Phi \mathbf{x}). \quad (10)$$

**Theorem 2.** *Given an  $M \times N$  matrix  $\Phi$ , a norm  $\|\cdot\|_X$ , and a positive integer  $k \leq N$ , a sufficient condition that there exists a decoder  $\Delta$  such that the instance optimality in expectation (7) holds with constant  $C_0$ , is that the null space property in expectation (9) holds with  $C = C_0/2$  for this  $(\Phi, \Delta)$ . A necessary condition is the null space property in expectation (9) with  $C = C_0$ . Similar results hold between MSE instance optimality (8) and null space property (10), with the constant  $C = C_0/4$  in the sufficient condition.*

Comparing to conventional CS that requires the null space property to hold with the best  $2k$ -term approximation error [9], the requirement for Gaussian SCS is relaxed to  $k$ , thanks to the linearity of the best  $k$ -term approximation for Gaussian signals.

Theorem 2 proves the existence of the decoder  $\Delta$  for which the instance optimality in expectation holds for  $(\Phi, \Delta)$ , given the null space property in expectation. However, it does not explain how such decoder is implemented. The following Corollary, a direct consequence of theorems 1 and 2, shows that for Gaussian signals the optimal decoder (2) leads to the instance optimality in expectation.

**Corollary 2.** *For Gaussian signals  $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{S})$ , if an  $M \times N$  sensing matrix  $\Phi$  satisfies the null space property in expectation (9) of order  $k$  in  $\|\cdot\|_1$ , with constant  $C_0/2$ , or the MSE null space property (10) of order  $k$  with constant  $C_0/4$ , then the optimal and linear decoder  $\Delta = \mathbf{S}\Phi^T(\Phi\mathbf{S}\Phi^T)^{-1}$  satisfies the instance optimality in expectation (7) in  $\|\cdot\|_1$ , or the MSE instance optimality (8).*

### IV-B. From RIP to Null Space Property

The Restricted Isometry Property (RIP) of a matrix measures its ability to preserve distances, and is related to the null space property in conventional CS [7], [10]. The new *linear* RIP of order  $k$  restricts the requirement of conventional RIP of order  $k$ , to a union of  $k$ -dimensional linear subspaces with consecutive supports:

**Definition 3.** *Let  $k \leq N$  be a positive integer. Let  $K_1$  define a linear subspace of functions with support in the first  $k$  indices in  $[1, N]$ ,  $K_2$  a linear subspace of functions with support in the next  $k$  indices, and so on. The functions in the last linear subspace  $K_J$  defined this way may have support with less than  $k$  indices. An  $M \times N$  matrix  $\Phi$  is said to have linear RIP of order  $k$  with constant  $\delta$  if*

$$(1 - \delta) \|\mathbf{x}\|_2 \leq \|\Phi \mathbf{x}\|_2 \leq (1 + \delta) \|\mathbf{x}\|_2, \quad \forall \mathbf{x} \in \cup_{j=1}^J K_j. \quad (11)$$

The linear RIP is a special case of the block RIP [11], with block sparsity one and blocks having consecutive support of the same size.

The following theorem relates the linear RIP (11) of a matrix  $\Phi$  to its null space property in expectation (9).

**Theorem 3.** *Let  $\mathbf{x} \in \mathbb{R}^N$  be a random vector that follows a certain distribution. Let  $\Phi$  be an  $M \times N$  matrix that satisfies the linear RIP of order  $2k$  with  $\delta < 1$ , and let  $\Delta$  be a decoder. Let  $\eta = \mathbf{x} - \Delta(\Phi\mathbf{x})$ . Assume further that  $E_{\mathbf{x},(\Phi)}|\eta[n]|$  decays in  $n$ :  $E_{\mathbf{x},(\Phi)}|\eta[n+1]| < E_{\mathbf{x},(\Phi)}|\eta[n]|$ ,  $\forall n < N$ . Then  $\Phi$  satisfies the null property in expectation of order  $k$  in  $\|\cdot\|_1$  (9), with constant  $C_0 = 1 + k^{1/2} \frac{1+\delta}{1-\delta}$ .*

For Gaussian signals  $\mathbf{x} \in \mathcal{N}(\mathbf{0}, \mathbf{S})$ , with  $\Phi$  Gaussian or Bernoulli matrices, one realization drawn independently for each  $\mathbf{x}$ , and with  $\Delta$  the optimal decoder (2), the decay of  $E_{\mathbf{x},\Phi}|\eta[n]|$  assumed in Theorem 3 is verified through Monte Carlo simulations.

#### IV-C. From Random Matrices to RIP

The next Theorem shows that Gaussian and Bernoulli matrices satisfy the RIP for one subspace with overwhelming probability.

**Theorem 4.** [2] *Let  $\Phi$  be a random matrix of size  $M \times N$  drawn according to any distribution that satisfies the concentration inequality*

$$\Pr(\|\Phi\mathbf{x}\|_2 - \|\mathbf{x}\|_2 \geq \epsilon \|\mathbf{x}\|_2) \leq 2e^{-Mc_0(\delta/2)}, \quad \forall \mathbf{x} \in \mathbb{R}^N, \quad (12)$$

where  $0 < \delta < 1$ , and  $c_0(\delta/2) > 0$  is a constant depending only on  $\delta/2$ . Then for any set  $K \subset \{1, \dots, N\}$  with  $|K| = k < M$ , we have

$$(1 - \delta)\|\mathbf{x}\|_2 \leq \|\Phi\mathbf{x}\|_2 \leq (1 + \delta)\|\mathbf{x}\|_2, \quad \forall \mathbf{x} \in \mathcal{X}_K, \quad (13)$$

where  $\mathcal{X}_K$  is the set of all vectors in  $\mathbb{R}^N$  that are zero outside of  $K$ , with probability greater than or equal to  $1 - 2(12/\delta)^k e^{-c_0(\delta/2)M}$ . Gaussian and Bernoulli matrices satisfy the concentration inequality (12).

The linear RIP of order  $k$  (11) requires that (13) holds for  $N/k \leq N$  subspaces. The next Theorem follows from Theorem 4 by simply multiplying by  $N$  the probability that the RIP fails to hold for one subspace.

**Theorem 5.** *Suppose that  $M$ ,  $N$  and  $0 < \delta < 1$  are given. Let  $\Phi$  be a random matrix of size  $M \times N$  drawn according to any distribution that satisfies the concentration inequality (12). Then there exist constants  $c_1, c_2 > 0$  depending only on  $\delta$  such that the linear RIP of order  $k$  (11) holds with probability greater than or equal to  $1 - 2Ne^{-c_2M}$  for  $\Phi$  with the prescribed  $\delta$  and  $k \leq c_1M$ .*

Comparing with conventional CS, where the null space property requires that the RIP (13) holds for  $\binom{N}{k}$  subspaces [2], [7], [10], the number of subspaces in the linear RIP (11) is sharply reduced to  $N/k$  for Gaussian SCS. In consequence, with the same number of measurements  $M$ , the probability that a Gaussian or Bernoulli matrix  $\Phi$  satisfies the linear RIP is substantially higher than that for the conventional RIP. Equivalently, given the same probability that  $\Phi$  satisfies the linear RIP or the conventional RIP of order  $k$ , the required number of measurements for the linear RIP is  $M \sim \mathcal{O}(k)$ , substantially smaller than the  $M \sim \mathcal{O}(k \log(N/k))$  required for the conventional RIP. Similar improvements have been obtained with model-based CS that assumes structured sparsity on the signals [3].

With the results above, we have showed that for Gaussian signals, with sensing matrices satisfying the linear RIP (11) of order  $2k$ , for example Gaussian or Bernoulli matrices with  $\mathcal{O}(k)$  rows, with overwhelming probability, and with the optimal decoder (2), SCS leads to the instance optimality in expectation of order  $k$  in  $\|\cdot\|_1$  (7), with constant  $C_0 = 2(1 + k^{1/2} \frac{1+\delta}{1-\delta})$ . By the definition of CS,  $k^{1/2}$  is typically small.

#### V. PERFORMANCE BOUNDS OF GAUSSIAN SCS WITH RIP IN EXPECTATION

This section shows that with an RIP in expectation, a matrix isometry property more adapted to SCS, the Gaussian SCS MSE instance optimality (8) of order  $k$  and constant  $C_0$ , holds in the  $l_2$  norm with probability one for any matrix.  $C_0$  has a closed-form and can be easily computed numerically.

**Definition 4.** *Let  $\mathbf{x} \in \mathbb{R}^N$  be a random vector that follows a certain distribution. Let  $\Phi$  be an  $M \times N$  sensing matrix and let  $\Delta$  be a decoder.*

Let  $\eta = \mathbf{x} - \Delta(\Phi\mathbf{x})$ .  $\Phi$  in  $(\Phi, \Delta)$  is said to have RIP in expectation in  $K$  with constant  $c_K$  if

$$E_{\mathbf{x},(\Phi)}\|\Phi\eta_K\|_2^2 = c_K E_{\mathbf{x},(\Phi)}\|\eta_K\|_2^2, \quad \forall \eta = \mathbf{x} - \Delta(\Phi\mathbf{x}), \quad (14)$$

where  $K \subset \{1, \dots, N\}$ ,  $\eta_K \in \mathbb{R}^N$  is the signal  $\eta$  restricted to  $K$  ( $\eta_K[n] = \eta[n]$ ,  $\forall n \in K$ , and 0 otherwise), and the expectation is with respect to  $\mathbf{x}$ , and to  $\Phi$  if one random  $\Phi$  is drawn independently for each  $\mathbf{x}$ .

The conventional RIP is known to be satisfied only by some random matrices, Gaussian and Bernoulli matrices for example, with high probability. For a given matrix, checking the RIP property is however NP-hard [2]. By contrast, the constant of the RIP in expectation (14) can be measured for any matrix via a fast simulation, the quick convergence guaranteed by the concentration of measure [16]. The next proposition, directly following from (3) and (4), further shows that for Gaussian signals, the RIP in expectation has its constant in a closed form.

**Proposition 1.** *Assume  $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{S})$ ,  $\Phi$  is an  $M \times N$  sensing matrix and  $\Delta$  is the optimal decoder (2). Then  $\Phi$  in  $(\Phi, \Delta)$  satisfies the RIP in expectation in  $K$ ,*

$$(E_\Phi) [Tr(\Phi\mathbf{R}_K\mathbf{S}\mathbf{R}_K^T\Phi^T - \Phi\mathbf{R}_K\mathbf{S}\Phi^T(\Phi\mathbf{S}\Phi^T)^{-1}\Phi\mathbf{S}\mathbf{R}_K^T\Phi^T)] = c_K(E_\Phi) [Tr(\mathbf{R}_K\mathbf{S}\mathbf{R}_K^T - \mathbf{R}_K\mathbf{S}\Phi^T(\Phi\mathbf{S}\Phi^T)^{-1}\Phi\mathbf{S}\mathbf{R}_K^T)], \quad (15)$$

where  $\mathbf{R}_K$  is an  $N \times N$  extraction matrix giving  $\eta_K = \mathbf{R}_K\eta$ , i.e.,  $\mathbf{R}_K(i, i) = 1$ ,  $\forall i \in K$ , all the other entries being zero. The expectation with respect to  $\Phi$  is calculated if one random  $\Phi$  is drawn independently for each  $\mathbf{x}$ .

The next Theorem shows that the RIP in expectation leads to the MSE null space property holding in equality.

**Theorem 6.** *Let  $\mathbf{x} \in \mathbb{R}^N$  be a random vector that follows a certain distribution,  $\Phi$  an  $M \times N$  sensing matrix, and  $\Delta$  a decoder. Assume  $E_{\mathbf{x},(\Phi)}\|\eta_K\|_2^2 \neq 0$  and  $E_{\mathbf{x},(\Phi)}\|\eta_{K^C}\|_2^2 \neq 0$ , for some  $K \subset \{1, \dots, N\}$ . Assume that  $\Phi$  in  $(\Phi, \Delta)$  has the RIP in expectation in  $K$  with constant  $a_K > 0$ , and in  $K^C = \{1, \dots, N\} \setminus K$  with constant  $b_K > 0$ :*

$$\frac{E_{\mathbf{x},(\Phi)}\|\Phi\eta_K\|_2^2}{E_{\mathbf{x},(\Phi)}\|\eta_K\|_2^2} = a_K, \quad \frac{E_{\mathbf{x},(\Phi)}\|\Phi\eta_{K^C}\|_2^2}{E_{\mathbf{x},(\Phi)}\|\eta_{K^C}\|_2^2} = b_K, \quad \forall \eta = \mathbf{x} - \Delta\Phi\mathbf{x}. \quad (16)$$

Then  $\Phi$  satisfies

$$E_{\mathbf{x},(\Phi)}\|\eta\|_2^2 = C_0 E_{\mathbf{x},(\Phi)}\|\eta_K\|_2^2, \quad (17)$$

where  $C_0 = 1 + b_K/a_K$ . In particular, if  $K = \{1, \dots, k\}$ , then  $\Phi$  satisfies the MSE null space property of order  $k$ , which holds with equality,

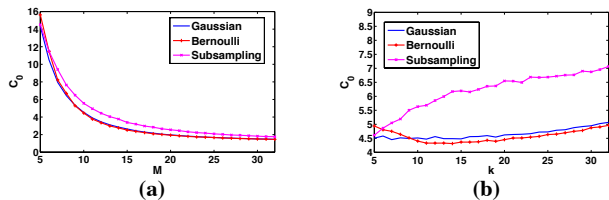
$$E_{\mathbf{x},(\Phi)}\|\eta\|_2^2 = C_0 \sigma_k(\{\eta\})_2^2. \quad (18)$$

Let us check the MSE null space property constant  $C_0$  of different sensing matrices in SCS for Gaussian signals  $\mathbf{x} \in \mathbb{R}^N \sim \mathcal{N}(\mathbf{0}, \mathbf{S})$ , assuming the eigenvalues of  $\mathbf{S}$  follow a power decay (6) with  $\alpha = -3$ , and  $N = 64$ . Gaussian, Bernoulli and random subsampling matrices  $\Phi$  of size  $M \times N$  are considered, and the optimal and linear decoder  $\Delta$  (2) is applied to reconstruct the signals. For each matrix distribution, a different random matrix realization  $\Phi$  is applied to sense each signal  $\mathbf{x}$ . Note that since the random subsampling matrix  $\Phi$ , each row containing one entry with value 1 at a random position and 0 otherwise, has the maximal coherence with the canonical basis, this matrix is not suitable for directly sensing  $\mathbf{x}$  [5], and is replaced by  $\Phi\Psi$  in the simulation, with  $\Psi$  a DCT basis having low coherence with  $\Phi$ .

Figure 2 (a) plots  $C_0 = 1 + b_K/a_K$ , obtained by simulating (16), with  $k = 10$ , for different values of  $M$ . When the number  $M$  of SCS measurements increases, the reconstruction error of SCS decreases, resulting in a smaller ratio over the best- $k$  term approximation error with a fixed  $k$ . Gaussian and Bernoulli matrices lead to similar  $C_0$  values, slightly smaller than that of random subsampling matrices. Figure 2 (b) plots  $C_0$ , as a function of  $k$ , with  $M = k$ . Gaussian and Bernoulli matrices lead to similar  $C_0 \approx 4.5$  that varies little with  $k$ . For random subsampling matrices  $C_0$  slowly increases, almost linearly, and is equal to 5.5 for a typical value  $k = 10$ , about 20% larger than that of Gaussian and Bernoulli matrices.

From Corollary 2 and Theorem 6, we obtain the next concluding result:

**Theorem 7.** *Assume  $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{S})$ . Let  $\Phi$  be an  $M \times N$  sensing matrix and  $\Delta$  the optimal and linear decoder (2). Then  $\Phi$  satisfies the MSE instance optimality of order  $k$  (8) with constant  $C_0 = 4(1 + b_K/a_K)$ ,  $a_K$  and  $b_K$*



**Fig. 2.** The MSE null space property constant  $C_0$  (18) of Gaussian, Bernoulli, and random subsampling matrices, as a function of  $M$  (left), and of  $k$  with  $M = k$  (right).

given in (16).

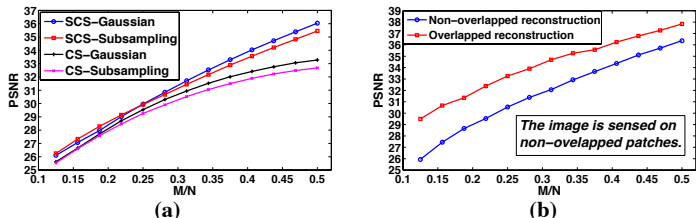
## VI. SCS WITH GMM

### – EXPERIMENTS ON REAL IMAGE DATA

This section applies SCS with GMM in real image sensing and compares it with conventional CS based on sparse models.

Following [18], an image is decomposed into  $\sqrt{N} \times \sqrt{N} = 8 \times 8$  patches  $\{\mathbf{x}_i\}_{1 \leq i \leq I}$  assumed to follow a GMM: There exist  $K$  Gaussian distributions  $\{\mathcal{N}(\mu_k, \Sigma_k)\}_{1 \leq k \leq K}$ , and each image patch  $\mathbf{x}_i$  is independently drawn from one of these Gaussians with an unknown index  $k$ . SCS samples each patch  $\mathbf{y}_i = \Phi_i \mathbf{x}_i$ , with a possibly different  $\Phi_i$  for each  $\mathbf{x}_i$ . The decoding scheme (i) estimates  $\{(\mu_k, \Sigma_k)\}_{1 \leq k \leq K}$ , and (ii) identifies the Gaussian distribution  $k_i$  that generates the  $i$ -th patch, and reconstructs  $\mathbf{x}_i$  with the optimal and linear decoder (2) associated to the appropriate Gaussian,  $\forall i$ . The resulting piecewise linear decoder is implemented with a computationally efficient MAP-EM algorithm alternating between (i) and (ii) [18]. The algorithm has fast convergence and the MAP probability of the measured data increases as the algorithm iterates [18] (refer to [18] Sec. 2 for more details).

The dictionary for conventional CS is learned with K-SVD [1] from 720,000 image patches, extracted from the entire standard Berkeley segmentation database containing 300 natural images [15]. The decoder is calculated with an  $l_1$  minimization.

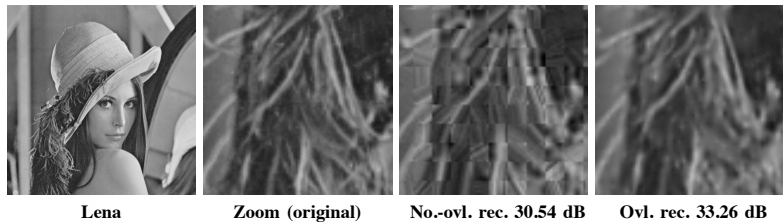


**Fig. 3.** (a) Patch PSNR (dB) vs sampling rate for SCS and CS using Gaussian and random subsampling sensing matrices. (b) Image PSNR (dB) vs sampling rate, for SCS using Gaussian sensing matrices with non-overlapping reconstruction, and subsampling random matrices with overlapping reconstruction.

Figure 3 (a) shows the sensing performance on about 260,000 (sliding) patches, regarded as signals, extracted from the standard image Lena, as shown in Figure 4. The PSNRs generated by SCS and CS using Gaussian and random subsampling sensing matrices, one independent realization for each patch, are plotted as a function of the sampling rate  $M/N$ . At the same sampling rate, SCS outperforms SC. The gain increases from about 0.5 dB at very low sampling rates ( $M/N \approx 0.1$ ), learning GMM from the poor-quality measured data being more challenging, to more than 3.5 dB at high sampling rates ( $M/N \approx 0.5$ ). (SC using an “oracle” dictionary learned from the full Lena itself, undoable in practice, improves its performance from 0.2 dB at low sampling rates to 1.3 dB at high sample rates, still lower than SCS.) For both SCS and CS, Gaussian and random subsampling matrices lead to similar PSNRs at low sampling rates ( $M/N < 0.25$ ), and at higher sampling rates the former gains about 0.5 dB. Recall that SCS is not just more accurate and significantly faster, but also only uses the compressed image, while CS uses a pre-learned dictionary from a large database.

Figure 3 (b) shows the sensing performance on the whole Lena image. The sensing is performed on *non-overlapped* patches. Random subsampling matrices, which are diagonal operators (one non-zero entry per row), have the advantage of being able to lead to reconstruction on overlapped patches, averaging the overlapped reconstructed patches further improving the image estimation. The PSNRs generated by SCS using random subsampling matrices and overlapped reconstruction are plotted, in comparison with those obtained using Gaussian sensing matrices and non-overlapped

reconstruction. The former improves from about 3.5 to 1.5 dB, at a cost of  $N = 64$  times computation. As illustrated in Figure 4, the overlapped reconstruction removes the block artifacts and considerably improves the reconstructed image.



**Fig. 4.** From left to right. Lena, a zoomed crop, reconstructed image by SCS using Gaussian sensing matrices and non-overlapping reconstruction, and by SCS using subsampling random matrices and overlapping reconstruction. The image is sensed on *non-overlapped* patches at a sampling rate of  $M/N = 0.25$ .

**Acknowledgements:** Work partially supported by NSF, ONR, NGA, ARO, and NSSEFF. The authors thank very much Stéphane Mallat for co-developing the GMM framework for solving inverse problems.

## VII. REFERENCES

- [1] M. Aharon, M. Elad, and A. Bruckstein. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans. on Signal Proc.*, 54(11):4311, 2006.
- [2] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin. A simple proof of the restricted isometry property for random matrices. *Constructive Approximation*, 28(3):253–263, 2008.
- [3] R.G. Baraniuk, V. Cevher, M.F. Duarte, and C. Hegde. Model-based compressive sensing. *IEEE Trans. on Info. Theo.*, 56(4):1982–2001, 2010.
- [4] R.G. Baraniuk and M.B. Wakin. Random projections of smooth manifolds. *Foundations of Comp. Math.*, 9(1):51–77, 2009.
- [5] E. Candès and J. Romberg. Sparsity and incoherence in compressive sampling. *Inverse Problems*, 23:969, 2007.
- [6] E.J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. on Info. Theo.*, 52(2):489–509, 2006.
- [7] E.J. Candès and T. Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Trans. on Info. Theo.*, 52(12):5406–5425, 2006.
- [8] M. Chen, J. Silva, J. Paisley, C. Wainwright, D. Dunson, and L. Carin. Compressive sensing on manifolds using a nonparametric mixture of factor analyzers: Algorithm and performance bounds. *IEEE Trans. on Signal Proc.*, (99):1, 2010.
- [9] A. Cohen, W. Dahmen, and R. DeVore. Compressed sensing and best  $k$ -term approximation. *J. of the Am. Math. Soc.*, 22(1):211–231, 2009.
- [10] D.L. Donoho. Compressed sensing. *IEEE Trans. on Info. Theo.*, 52, 2006.
- [11] Y.C. Eldar and M. Mishali. Robust recovery of signals from a structured union of subspaces. *IEEE Trans. on Info. Theo.*, 55(11):5302–5316, 2009.
- [12] S.M. Kay. *Fundamentals of Statistical Signal Processing, Volume 1: Estimation Theory*. Prentice Hall, 1998.
- [13] F. Léger, G. Yu, and G Sapiro. Efficient matrix completion with Gaussian models. *Submitted*, Oct., 2010.
- [14] S. Mallat. *A Wavelet Tour of Signal Processing: The Sparse Way, 3rd Edition*. Academic Press, 2008.
- [15] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proc. ICCV*, 2001.
- [16] M. Talagrand. A new look at independence. *Ann. Prob.*, 24:1, 1996.
- [17] G. Yu, S. Mallat, and E. Bacry. Audio denoising by time-frequency block thresholding. *IEEE Trans. on Signal Proc.*, 56(5):1830, 2008.
- [18] G. Yu, G. Sapiro, and S. Mallat. Solving inverse problems with piecewise linear estimators: from Gaussian mixture models to structured sparsity. *Submitted, arxiv.org/abs/1006.3056*, June, 2010.