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 $O(k)$ measurements, considerably smaller than the $O(k\log(N/k))$ 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 to be 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 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.

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 $x \in \mathbb{R}^N$ be a signal of interest, $\Phi \in \mathbb{R}^{M \times N}$ a non-adaptive sensing matrix (encoder), consisting of $M \ll N$ measurements, $y = \Phi x \in \mathbb{R}^M$ a measured signal, and $\Delta$ a decoder used to reconstruct $x$ from $\Phi y$. CS develops encoder-decoder pairs $(\Phi, \Delta)$ such that a small reconstruction error $x - \Delta(\Phi y)$ can be achieved.

Reconstructing $x$ from $\Phi y$ 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) where $N$ 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. 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 and missing data estimation.

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. This result shows that with the same random matrices as in conventional CS, but with a considerably reduced number $M = O(k)$ 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 VII 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 approximation with probability one, and the bound constant can be efficiently calculated. Section VII introduces, for CS 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 $x \sim \mathcal{N}(0, \Sigma)$, with mean zero and diagonal covariance matrix $\Sigma = \text{diag}[\lambda_1, \ldots, \lambda_N]$, with $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_N$ 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). For Gaussian and Bernoulli matrices that are known to be universal, analyzing CS in the canonical basis or the PCA basis is equivalent.

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

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

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

$$
E_x[\|x - \Delta(\Phi y)\|^2] = Tr(S - S \Phi (\Phi S \Phi^T)^{-1} \Phi S).
$$
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 \([5, 10]\), 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 \( x \), with all the other conditions as in Theorem 1 the MSE of SCS is

\[
E_{x,\Phi}[\|x - \Delta(\Phi x)\|^2_F^2] = E_\Phi[T_\Phi R(\Phi)S_\Phi^{-1}F_\Phi]
\]

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 \( x \sim \mathcal{N}(0, \mathbf{S}) \) with sorted eigenvalues \( \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_N \), it is well-known that the best \( k \)-term approximation error,

\[
\sigma_k(\{x\})_x = E_{x,\Phi}[\|x - x_k\|_2], \quad \sigma_k(\{x\})_x^2 = E_{x,\Phi}[\|x - x_k\|_2^2],
\]

is obtained with a linear projection to the first \( k \) entries, i.e., \( x_k[n] = x[n] \), \( \forall n \in K = \{1, \ldots, k\} \), and \( x_k[n] = 0 \) otherwise \([13]\). Note that \( \sigma_k(\{x\})_x^2 = \sum_{m=k+1}^N \lambda_m \). 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 \([13]\).

\[
\lambda_m = m^{-\alpha}, \quad 1 \leq m \leq N,
\]

\( N = 64 \), is assumed in the Monte Carlo simulation. An independent random Gaussian matrix realization \( \Phi \) is applied to sense each signal \( x \) \([9]\), 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 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 \( \Phi x \), i.e., \( \Phi(\Delta(\Phi x)) = \Phi x \), satisfied by the optimal \( \Delta \) in \([2]\) for Gaussian signals \( x \), \( \forall \Phi \).

**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 \( \|\cdot\|_2 \), with a constant \( C_0 \), if

\[
E_{x,\Phi}[\|x - \Delta(\Phi x)\|_2^2] \leq C_0 \sigma_k(\{x\})_x^2.
\]

The next section evaluates these values through Monte Carlo simulations, preceding the theoretical bounds later developed.

**Corollary 2.** For Gaussian signals \( x \sim \mathcal{N}(0, \mathbf{S}) \), if an \( M \times N \) sensing matrix \( \Phi \) satisfies the null space property in expectation \([9]\) holds with constant \( C_0 \), is 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 \( k \)-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.** Given an \( M \times N \) matrix \( \Phi \), a norm \( \|\cdot\|_2 \), 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 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.

**Corollary 2.** For Gaussian signals \( x \sim \mathcal{N}(0, \mathbf{S}) \), 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 \( \Delta = \Phi^T(\Phi \Phi^T)^{-1} \) 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 \([2, 10]\). The new \( \infty \) 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 \( K_3 \) defined this way may have support with less than \( k \) indices. An \( M \times N \) matrix \( \Phi \) is said to have \( \infty \) RIP of order \( k \) with constant \( \delta \) if

\[
(1 - \delta)\|x\|_2 \leq \|\Phi x\|_\infty \leq (1 + \delta)\|x\|_2, \quad \forall x \in \bigcup_{j=1}^k K_j.
\]

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 of a matrix $\Phi$ to its null space property in expectation [2].

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

For Gaussian signals $x \in \mathcal{N}(0, S)$, with $\Phi$ Gaussian or Bernoulli matrices, one realization drawn independently for each $x$, and with $\Delta$ the optimal decoder [2], the decay of $E_x(\Phi x(\eta^T x))$ 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

$$Pr(\|\Phi x\|_2^2 - \|x\|_2^2) \leq \epsilon \|x\|_2^2, \quad \forall x \in \mathbb{R}^N,$$

(12)

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

$$\|\Phi K\|_2 \leq \|x\|_2 \leq (1 + \delta)\|x\|_2, \quad \forall x \in \mathcal{N}(K),$$

(13)

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

The linear RIP of order $k$ requires that $\|x\|_2$ holds with probability greater than or equal to $1 - 2N e^{-c_0M}$. Let $\Phi$ be a 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 $k$ 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 $k$ subspaces [2], [7], [10], the number of subspaces in the linear RIP (11) is sharply reduced to $N/k$ for Gaussian CS. 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 shown 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], $\Phi$ satisfies the instance optimality in expectation of order $k$ in $\| \cdot \|_1$ 1, with constant $C_0 = 2(1 + k^{1/2} 1/2^{1/2})$. 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 [3] 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 $x \in \mathbb{R}^N$ 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 $\eta = x - \Delta(\Phi x)$. $\Phi$ in $(\Phi, \Delta)$ is said to have RIP in expectation in $K$ with constant $c_K$ if

$$E_\eta(\Phi x(\eta^T x)) = c_K E_\eta(\Phi x(\eta^T x)), \quad \forall \eta = x - \Delta(\Phi x),$$

(14)

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

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 $x \sim \mathcal{N}(0, S)$, $\Phi$ an $M \times N$ sensing matrix and $\Delta$ the optimal decoder. Then $\Phi$ in $(\Phi, \Delta)$ satisfies the RIP in expectation in $K$. Let

$$(\Phi) \left[ \operatorname{Tr} \left( \Phi R_k S_k^T \Phi^T - \Phi R_k S_k^T \Phi^T \Phi S_k^T \Phi^T \right) \right] = c_K (\Phi) \left[ \operatorname{Tr} \left( R_k S_k^T R_k S_k^T \Phi S_k^T \Phi^T \right) \right],$$

(15)

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

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

**Theorem 6.** Let $x \in \mathbb{R}^N$ be a random vector that follows a certain distribution, $\Phi$ an $M \times N$ sensing matrix, and $\Delta$ a decoder. Assume $E_\eta(\Phi x(\eta^T x)) = 0$ and $E_\eta(\Phi x(\eta)^T x) = 0$, for some $K \subset \{1, \ldots, N\}$. Assume that $\Phi$ in $(\Phi, \Delta)$ has the RIP in expectation in $K$ with constant $c_K > 0$, and in $K^c = \{1, \ldots, N\} \setminus K$ with constant $b_K > 0$: $E_\eta(\Phi x(\eta^T x)) = c_K$, $E_\eta(\Phi x(\eta)^T x) = b_K$, $\forall \eta = x - \Delta(\Phi x)$.

Then $\Phi$ satisfies

$$E_\eta(\Phi x(\eta)^T x) = c_K E_\eta(\Phi x(\eta)^T x),$$

(17)

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

$$E_\eta(\Phi x(\eta)^T x) = C_0 \sigma_2(\{1\})^2.$$  

(18)

Let us check the MSE null space property constant $C_0$ of different sensing matrices in SCS for Gaussian signals $x \in \mathbb{R}^N \sim \mathcal{N}(0, S)$, assuming the eigenvalues of $S$ follow a power decay [9] 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 $x$. Note that since the linear 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 $x$, and is replaced by $\Phi^T \Psi$ 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 of $M$ 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 $C_0$ values, varying typically with little $k$. 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 3 we obtain the next concluding result:

**Theorem 7.** Assume $x \sim \mathcal{N}(0, S)$. 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 $b_K$
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 [13], an image is decomposed into $\sqrt{N} \times \sqrt{N} = 8 \times 8$ patches $\{x_i\}_{1 \leq i \leq I}$ assumed to follow a GMM: There exist $K$ Gaussian distributions $\{N(\mu_k, \Sigma_k)\}_{1 \leq k \leq K}$, and each image patch $x_i$ is independently drawn from one of these Gaussians with an unknown index $k$. SCS samples each patch $y_i = \Phi_k x_i$, with a possibly different $\Phi$ for each $x_i$. The decoding scheme (i) estimates $\{N(\mu_k, \Sigma_k)\}_{1 \leq k \leq K}$, and (ii) identifies the Gaussian distribution $k_i$ that generates the $i$-th patch, and reconstructs $x_i$ with the optimal and linear decoder 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 (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. 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].

SCS using Gaussian sensing matrices and non-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. Wang, 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 mixture models. IEEE Trans. on Info. Theo., 52(8):3588–3606, 2006.
[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.