ADAPTED STATISTICAL COMPRESSIVE SENSING: LEARNING TO SENSE GAUSSIAN MIXTURE MODELS

Julio M. Duarte-Carvajalino, [1] *Guoshen Yu,* [1] *Lawrence Carin,* [2] *and Guillermo Sapiro*[1]

[1] University of Minnesota and [2] Duke University

ABSTRACT

A framework for learning sensing kernels adapted to signals that follow a Gaussian mixture model (GMM) is introduced in this paper. This follows the paradigm of statistical compressive sensing (SCS), where a statistical model, a GMM in particular, replaces the standard sparsity model of classical compressive sensing (CS), leading to both theoretical and practical improvements. We show that the optimized sensing matrix outperforms random sampling matrices originally exploited both in CS and SCS.

Index Terms— Compressive Sensing, Gaussian Mixture Models, Learning, Structured Sparsity.

1. INTRODUCTION

Compressive sensing (CS) theory states that signals having a concise (sparse) or well approximated (compressible) linear representation on an appropriate sparsifying dictionary, can be recovered with zero (sparse) or minimum (compressible) information loss, from a number of linear projections of dimension considerably lower than the number of samples required by the Shannon–Nyquist Theorem [1].

Besides signal sparsity or compressibility on an appropriated dictionary, CS also requires the sensing matrix (kernel) to be as incoherent as possible with this dictionary. Random matrices, such as Gaussian or ±1 random matrices, are largely incoherent with any fixed sparsifying dictionary with overwhelming probability. On the other hand, it has been shown that deterministic matrices can be more effective than random ones for sensing real signals [2-6], motivating the sensing kernel learning here developed.

It has been noted that off-the-shelf dictionaries are not flexible enough to capture the complexity of natural signals, and learned overcomplete dictionaries are very popular and lead to improved results [7]. Such dictionaries define a large search space [8,9]; the search space on an unstructured overcomplete dictionary with N atoms consists of $\binom{N}{L}$ possible combinations for a signal with sparsity $L \ll N$. This renders the overall representation unstructured and unstable. Structured overcomplete dictionaries have been proposed to reduce the size of the search space and improve the sparse representation of such complex signals [9].

In CS and sparse modeling, the sensed signal is reconstructed via non-linear optimization strategies. A piecewise *linear* inversion model (PLM) based on the maximum a posteriori expectation-maximization method (MAP-EM), for signals following a statistical Gaussian Mixture Model (GMM), was recently introduced [8,10] (see also [11]). The PLM relates to structured sparsity, since each Gaussian defines a PCA, and therefore the GMM can be considered as a dictionary of PCAs. The PLM, and GMMs in general [8,11], have been shown to be very effective and computationally efficient to reconstruct signals that have been degraded by noise, blurring, sub-sampling, or other linear filters, such as CS random matrices. In addition, theoretical analysis in [10,11], which mainly considers random sensing matrices, indicates numerous advantages of such GMM when compared to standard sparsity models.

Motivated by such recent results, as well as the classical studies, popularity, and proven relevance of GMMs, we here develop a framework that simultaneously learns, from data, the GMM and the corresponding sensing kernel adapted to it, replacing the random sensing matrices studied in [8,10,11]. This brings together the benefits of adapted deterministic sensing and those of GMM.

In Section 2 we integrate the PLM [8,10] and the optimization of the sensing matrix [4]. Experimental results are presented in Section 3. We discuss the results, caveats, and future directions in Section 4.

2. SENSING KERNELS FOR GMM/PCA

2.1 General simultaneous dictionary and sensing learning

Let $\mathbf{x} \in \mathbf{R}^n$ be a discrete signal, $\mathbf{D}$ an $n \times N$ overcomplete sparsifying dictionary ($N > n$), $\boldsymbol{\alpha} \in \mathbf{R}^N$ the sparse representation of $\mathbf{x}$ in $\mathbf{D}$, such that $\mathbf{x} = \mathbf{D}\boldsymbol{\alpha}$, and $\boldsymbol{\Phi}$ the $m \times n$ ($m < n$) sensing matrix. We can simultaneously learn, off-line, the dictionary and sensing matrix [4]:

$$(\widehat{\boldsymbol{\Phi}}, \widehat{\mathbf{D}}, \widehat{\boldsymbol{\alpha}}) = \arg\min_{\boldsymbol{\alpha},\mathbf{D},\boldsymbol{\Phi}} (\|\mathbf{y} - \boldsymbol{\Phi}\mathbf{D}\boldsymbol{\alpha}\|_2^2 + \lambda \|\boldsymbol{I} - \boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{\Psi}\|_F^2),$$
$$s.t. \|\mathbf{x} - \mathbf{D}\boldsymbol{\alpha}\|_2^2 \le \epsilon, \|\boldsymbol{\alpha}\|_0 \le L \ll N, \qquad (1)$$

where $\boldsymbol{\Psi} \triangleq \boldsymbol{\Phi}\mathbf{D}$, $\mathbf{y} = \boldsymbol{\Phi}\mathbf{x} + \boldsymbol{w}$ is the compressed signal with additive Gaussian noise $\boldsymbol{w}$, ϵ is a small positive number, λ a weight (trade-off) factor, and $\|\cdot\|_0, \|\cdot\|_2, \|\cdot\|_F$ correspond to the 0 (sparsity), 2, and Frobenius (energy) (pseudo-)norms, respectively. As indicated in [4], the 0-norm can be replaced by the 1-norm.

Notice that in (1), we have imposed the condition that the Gramm matrix ($\mathbf{\Psi}^{\mathrm{T}}\mathbf{\Psi}$) should be as close as possible to the identity. This condition tries to satisfy the incoherence principle in CS (see Section 1). More specifically, this is aimed at satisfying the restricted isometry property (RIP), stating that the power set of columns from $\mathbf{\Psi}$, of cardinality less than L, must be as orthogonal as possible [1]. The RIP ensures low coherence between $\mathbf{\Phi}$ and $\mathbf{D}$, and helps optimization algorithms such as ℓ_1-minimization and (regularized) orthogonal matching pursuit (OMP) to succeed in finding $\mathbf{x}$ from its projection $\mathbf{y}$.

2.2 GMM/PCA as dictionaries

Instead of solving (1) with a basis pursuit algorithm minimizing a Lagrangian penalized by a sparse ℓ_1-norm [12,13], or with a greedy matching pursuit-type algorithms [14] as in [4], we propose here to use the MAP-EM framework introduced in [8,10], which is based on principal component analysis (PCA) and GMMs instead of standard sparsity.

Let us assume that there exist K Gaussian distributions such that the signal of interest $\mathbf{x}$ corresponds to a realization of *exactly* one of these Gaussian distributions (one-block sparsity), i.e.,

$$p(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n|\mathbf{\Sigma}_k|}} exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_k)^{\mathrm{T}}\mathbf{\Sigma}_k^{-1}(\mathbf{x}-\boldsymbol{\mu}_k)\right), \quad (2)$$

where $\boldsymbol{\mu}_k \in \mathbf{R}^n$ is the mean vector and $\mathbf{\Sigma}_k$ the $n\times n$ covariance matrix for the k^{th} Gaussian distribution, $k = 1, \dots, K$, and $|\mathbf{\Sigma}_k|$ represents the determinant of $\mathbf{\Sigma}_k$.

In analogy to structured sparsity, define the dictionary

$$\mathbf{D} = [\mathbf{V}_1 \quad \dots \quad \mathbf{V}_K],\ \mathbf{\Sigma}_k = \mathbf{V}_k\mathbf{\Lambda}_k\mathbf{V}_k^{\mathrm{T}},\ k = 1, \dots, K, \quad (3)$$

based on the PCAs of the covariance matrices. Therefore, $\mathbf{x}$ is represented in this sparsifying dictionary as

$$\mathbf{x} = \mathbf{D}\boldsymbol{\alpha} = [\mathbf{V}_1 \quad \dots \quad \mathbf{V}_K]\left[\mathbf{0} \dots \boldsymbol{\alpha}_k^{\mathrm{T}} \dots \mathbf{0}\right]^{\mathrm{T}}, \quad (4)$$

where $\|\boldsymbol{\alpha}\|_0 = \|\boldsymbol{\alpha}_k\|_0 = L \leq n \ll N$, and L corresponds to the L largest eigenvalues of $\mathbf{\Sigma}_k$, hence, the sparsity condition is met by design. Notice that the dictionary $\mathbf{D}$ has a predefined structure, where each $n\times n$ block corresponds to a PCA basis, which is expected to capture most of the variability of the signals, modeled by the corresponding Gaussian distribution. This structure significantly reduces the search space from $\binom{N}{L}$ on an unstructured overcomplete dictionary to K, simple select the best Gaussian/PCA.

As shown in [8], for a given *fixed* linear filter $\mathbf{\Phi}$, we can solve (1), in closed form, using a MAP-EM iterative algorithm that alternates between an E-step and an M-step, simultaneously reconstructing the signal and learning the GMM (see [8] for a discussion on the initialization of the MAP-EM). In the E-step, the Gaussian parameters $(\boldsymbol{\mu}_k, \mathbf{\Sigma}_k)$ are assumed known and the MAP estimate of $\boldsymbol{\alpha}$ (for each candidate Gaussian) is found solving

$$\widehat{\boldsymbol{\alpha}}_k = \arg\min_{\boldsymbol{\alpha}_k}\left(\|\mathbf{y} - \mathbf{\Phi}\mathbf{V}_k\boldsymbol{\alpha}_k\|_2^2 + \sigma^2\boldsymbol{\alpha}_k^{\mathrm{T}}\mathbf{\Lambda}_k\boldsymbol{\alpha}_k\right), \quad (5)$$

where the noise $\boldsymbol{w}\sim\mathcal{N}(\mathbf{0},\sigma^2\boldsymbol{I})$ and we have made $\boldsymbol{\mu}_k = \mathbf{0}$ (subtracting the mean from each distribution). Equation (5) can be efficiently solved using the Wiener filter $\mathbf{W}_k$ for each $k = 1, \dots, K$,

$$\widehat{\boldsymbol{\alpha}}_k = \mathbf{W}_k\mathbf{y}_k, \mathbf{W}_k = \left(\mathbf{V}_k^{\mathrm{T}}\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi}\mathbf{V}_k + \sigma^2\mathbf{\Lambda}_k^{-1}\right)^{-1}\mathbf{V}_k^{\mathrm{T}}\mathbf{\Phi}^{\mathrm{T}}. \quad (6)$$

In practice, $\widehat{\boldsymbol{\alpha}}_k$ is estimated solving the equivalent linear system

$$\left(\sigma^{-2}\mathbf{\Lambda}_k\mathbf{V}_k^{\mathrm{T}}\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi}\mathbf{V}_k + \mathbf{I}\right)\widehat{\boldsymbol{\alpha}}_k = \left(\sigma^{-2}\mathbf{\Lambda}_k\mathbf{V}_k^{\mathrm{T}}\mathbf{\Phi}^{\mathrm{T}}\right)\mathbf{y}_k. \quad (7)$$

From the estimated $\widehat{\boldsymbol{\alpha}}_k$, we select the best Gaussian model k (intrinsically incorporating the model/Gaussian complexity [8]), and the corresponding sparse representation of the signal, $\widehat{\boldsymbol{\alpha}}_k$, that minimizes (5), as well as the signal reconstruction.

Once we have chosen the model and sparse representation of the signal (this is done for all the available sensed signals, e.g., all patches in the sensed image), we perform an M-step consisting of re-estimating the Gaussian parameters $(\boldsymbol{\mu}_k, \mathbf{\Sigma}_k)$. This is obtained via the empirical mean and covariance, considering all the S signals that were assigned to the same Gaussian and reconstructed with it during the E-step (see [8] for a discussion on the optimality of this):

$$\widehat{\boldsymbol{\mu}}_k = \frac{1}{S}\sum_{i=1}^{S}\mathbf{x}_i\,,\ \widehat{\mathbf{\Sigma}}_k = \frac{1}{S}\sum_{i=1}^{S}(\mathbf{x}_i - \boldsymbol{\mu}_k)(\mathbf{x}_i - \boldsymbol{\mu}_k)^{\mathrm{T}}. \quad (8)$$

Clearly the GMM-based PLM is computational very efficient, and without the need for learning from large databases, operating just on the sensed image, it has been shown to outperform algorithms based on standard sparse models [8,11].

2.3 Putting it all together: Learning to sense the GMM

So far we have assumed a given fixed sensing matrix $\mathbf{\Phi}$ in PLM (see equations (5-7)). We can update, however, the sensing matrix to the GMM, before each E-M step. This is explained next.

We can approximate $\|\boldsymbol{I} - \mathbf{\Psi}^{\mathrm{T}}\mathbf{\Psi}\|_F^2$ on (1) as (the dictionary $\mathbf{D}$ is now composed of the being learned PCAs)

$$\mathbf{\Psi}^{\mathrm{T}}\mathbf{\Psi} = \mathbf{D}^{\mathrm{T}}\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi}\mathbf{D} \approx \mathbf{I} \ \Rightarrow\ \mathbf{D}\mathbf{D}^{\mathrm{T}}\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi}\mathbf{D}\mathbf{D}^{\mathrm{T}} \approx \mathbf{D}\mathbf{D}^{\mathrm{T}}. \quad (9)$$

Let the PCA decomposition of the symmetric matrix $\mathbf{D}\mathbf{D}^{\mathrm{T}}$ be $\mathbf{U}\mathbf{\Delta}\mathbf{U}^{\mathrm{T}}$, we simplify (9) as

$$\mathbf{\Delta}\mathbf{\Gamma}^{\mathrm{T}}\mathbf{\Gamma}\mathbf{\Delta} \approx \mathbf{\Delta}, \quad \mathbf{\Gamma} = \mathbf{\Phi}\mathbf{U}, \quad (10)$$

giving the closed form solution $\mathbf{\Phi} = \left[\mathbf{\Delta}_r^{-1/2}\ \mathbf{0}\right]\mathbf{U}^{\mathrm{T}}$, where, $\mathbf{\Delta}_r$ is the reduced diagonal matrix containing the non-zero eigenvalues of $\mathbf{\Delta}$, ordered in decreasing order. Alternatively, we can update the sensing kernel $\mathbf{\Phi}$ using the iterative algorithm also proposed in [4], which is more stable in the presence of noise. In fact, the iterative algorithm has lower reconstruction error than the optimal CS for structured dictionaries, proposed in [15], based on [4].

This concludes the presentation of the proposed sensing kernel design for GMMs. We now proceed to present numerical examples showing the relevance of this adaptation.

3. EXPERIMENTAL RESULTS

We test on two different data sets. The first corresponds to the MNIST handwritten digits image database,[1] and the second corresponds to natural images taken from the Berkeley segmentation data set.[2] The MNIST digit images were resampled from 28×28 pixels to 8×8 pixels (using subsampling and cubic interpolation). While we present here illustrative results for images, GMMs and the sensing kernel design framework here introduced are applicable to other classes of signals as well.

For the case of the digit images, we use different training ($\mathbf{X}_{tr}$) and testing sets ($\mathbf{X}_{te}$), each one consisting of 2000 images per digit (0-9). We learn the best sensing $\mathbf{\Phi}$ and GMM/PCA dictionary $\mathbf{D}$ from the estimated compressed signals, $\mathbf{Y}_{tr} = \mathbf{\Phi}\mathbf{X}_{tr}$, that are updated as soon the matrix $\mathbf{\Phi}$ is updated. After learning the GMM dictionary $\mathbf{D}$ and sensing kernel $\mathbf{\Phi}$ that minimizes (1), we use this sensing matrix to sense the testing signals $\mathbf{X}_{te}$, from which we obtain $\mathbf{Y}_{te} = \mathbf{\Phi}\mathbf{X}_{te}$. Then, we reconstruct the testing signals $\widehat{\mathbf{X}}_{te}$ from $\mathbf{Y}_{te}$ and the pre-learned $\mathbf{\Phi}$ and $\mathbf{D}$. Since the testing signals are different from the training signals, we can use MAP-EM (PLM) to adapt the dictionary (with a fixed $\mathbf{\Phi}$) to the new data, further improving the quality of the reconstructed signals.

During the training phase, the M-step uses the original signals ($\mathbf{X}_{tr}$) to compute the Gaussian parameters in (8). This is valid in the design phase, we are just learning the sensing and dictionary matrices that are going to be used later on the reconstruction phase, where we only use the sensed signals, $\mathbf{Y}_{te}$, and $\mathbf{\Phi}$ has been fixed (learned).

For the case of natural images, we use overlapping patches from each image to learn the GMM dictionary and adapted sensing kernel that minimizes (1). In this case the initialization provided by K=18 directional PCAs following [8] leads to good enough reconstructions. Hence, the original signals are not used in the M-step during training.

The reconstruction error is measured in terms of the peak signal to noise ratio (PSNR), given by

$$\text{PSNR} = 10\log_{10}\left({I_{max}^2}/{MSE}\right), \tag{12}$$

where I_{max} corresponds to the maximum image intensity.

Given that different random sensing matrices can produce slightly different reconstructions, we report, when comparing with the adapted sensing case, the average reconstruction error obtained using 20 different realizations of random Gaussian sensing matrices. For the case of overlapping patches extracted from natural images, we report the average reconstruction error obtained using 20 different images, each one using a different random matrix.

Since we are averaging the reconstruction error from a large amount of signals (2000 handwritten digit images per digit, and $\sim 10^5$ overlapping image patches), and for 20 different random realizations, by the Central Limit Theorem, we can assume that these errors follow a normal distribution, and hence, we can test the statistical significance of these average differences using a paired t-test.

Table 1 compares the mean PSNR of the reconstructed (testing) digits using random versus optimized sensing matrices, at three different levels of compression n/m. This table also indicates the significance level (p-value) of the mean differences found using a paired t-test. This p-value indicates that the differences found are very statistically significant. A per-digit comparison is presented in Figure 1, improvements of up to 3.35 dbs. are observed.

Table 1: Mean PSNR digit reconstruction

PSNR (dbs.)				
Compression	Random	Optimized	Difference (dbs.)	p-value
8.0	21.05	21.74	0.69	0.0001
5.3	23.10	23.85	0.75	0.0001
3.2	26.69	27.77	1.08	0.0001

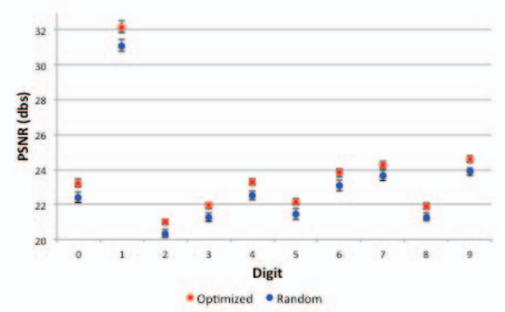

Figure 1: Per-digit comparison between the learned and random sensing kernels at n/m = 5.3 (bar errors correspond to three standard deviations of the estimated distribution of the mean).

Table 2 compares the mean PSNR of the reconstructed natural image patches using random and optimized sensing matrices, again at three compression levels. The paired t-test also indicates here that these differences are very significant statistically (improvements of up to 1.25 dbs. are observed).

Table 2: Mean PSNR natural image patches reconstruction

PSNR (dbs.)				
Compression	Random	Optimized	Difference (dbs.)	p-value
8.0	28.04	28.74	0.70	0.0001
5.3	29.86	30.42	0.56	0.0001
3.2	32.90	33.47	0.57	0.0001

[1] http://yann.lecun.com/exdb/mnist/

[2] http://www.eecs.berkeley.edu/Research/Projects/CS/vision/bsds/

4. DISCUSSION AND CONCLUSIONS

The reported results clearly indicate that the simple GMM-adapted sensing matrices achieve higher accuracy than random sensing ones. We plan to study in the future a structured sensing matrix that satisfies physical sensor constrains and that better exploits the underlying dictionary structure and GMM paradigm. The model can be further improved with a larger number of Gaussians (only K=18, following [8], have been used for the experiments here reported), and considering two or more mixed Gaussian models to represent the signals (going from one-block sparsity to more general structured sparsity models). As discussed in [8], the initialization of the PLM is critical, and this has been designed for natural images. For digits, as well as other signals, such initialization has to be re-designed. This might explain in part the observation that in Table 1 the PSNR difference between random and optimized sensing matrices increases as the compression rate decreases, while in Table 2, the highest PSNR difference corresponds to the maximum compression. Such difference also depends on the level of sparsity in the signal's class.

While here we considered mixtures of Gaussians, the framework can be extended to other mixtures as well following the recent results in [16]. In addition, the off-line kernel learning here reported can be augmented by an on-line paradigm [16,17]. The combination of the results here reported with the techniques in [16,17] is the subject of current efforts. Additional details such as computational complexity and extensions can be found in [18].

To conclude, we have shown that an optimized sensing matrix can outperform random sensing matrices within the statistical compressive sensing framework, considering GMMs in particular. The framework analyzed here, using small digit images and overlapping patches, can be used as well for other classes of signals. The PLM and underlying GMM has been shown to be very powerful also for matrix completion [19] and audio (see [11] for other applications and datasets), further supporting the need to optimize kernel design for this type of statistical signal models.

Acknowledgments. The authors would like to thank David Brady, Robert Calderbank, Wotao Yin, and Stan Osher for feedback on the work here reported. This work is partially supported by NSF, ONR, NGA, ARO, DARPA, and NSSEFF.

REFERENCES

[1] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Proc. Mag.*, vol. 25, no. 2, pp. 21–30, 2008.

[2] M. Elad, "Optimized projections for compressed sensing," *IEEE Trans. Signal Proc.*, vol. 55, no. 12, pp. 5695–5702, 2007.

[3] R. A. DeVore, "Deterministic constructions of compressed sensing matrices," *J. Complexity*, vol. 23, nos. 4-6, pp. 918–925, Aug. 2007.

[4] J. M. Duarte-Carvajalino and G. Sapiro, "Learning to sense sparse signals: Simultaneous sensing matrix and sparsifying dictionary optimization," *IEEE Trans. Image Proc.*, vol. 18, no. 7, pp. 1395 – 1408, 2009.

[5] G. Peyré, "Best basis compressed sensing," *IEEE Trans. Image Proc.*, vol. 58, no. 5, 2010.

[6] S. D. Howard, R. Calderbank, and S. J. Searle, "A fast reconstruction algorithm for deterministic compressive sensing using second order Redd-Muller codes," *CISS*, pp. 11-15, 2008.

[7] J. Mairal, M. Elad, and G. Sapiro, "Sparse representation for color image restoration," *IEEE Trans. Image Proc.*, vol. 17, no. 1, pp. 53–69, 2008.

[8] G. Yu, G. Sapiro, and S. Mallat, "Solving inverse problems with piecewise linear estimators: From Gaussian mixture models to structured sparsity," arXiv:1006.3056, June 2010.

[9] M. F. Duarte and Y. C. Eldar, "Structured compressed sensing: from theory to applications," arXiv:1106.6224v2, June 2011.

[10] G. Yu and G. Sapiro, "Statistical compressed sensing of Gaussian mixture models," arXiv:1101.5785v1, Jan. 2011, *IEEE Trans. Signal Proc.,* to appear.

[11] 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. Signal Proc.*, December 2010.

[12] S. S. Chen, D.L. Donoho, and M.A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comp.*, vol. 20, pp. 33-61, 1999.

[13] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. of the Royal Stat. Society*, vol. 58, no. 1, pp. 267–288, 1996.

[14] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Proc.*, vol. 41, no. 12, pp. 3397–3415, 1993.

[15] L. Zelnik-Manor, K. Rosenblum, and Y. C. Eldar, "Sensing matrix optimization for block-sparse decoding," *IEEE Trans. Signal Proc.*, vol. 59, no. 9, pp. 4300–4312, 2011.

[16] W. R. Carson, M. R. D. Rodrigues, M. Chen, L. Carin, and R. Calderbank, "How to focus the discriminative power of a dictionary," *submitted*, September 2011.

[17] J. M. Duarte-Carvajalino, G. Yu, L. Carin, and G. Sapiro, "Online adaptive statistical compressed sensing of Gaussian mixture models,'' *arXiv*:1112.5895v1, Dec. 2011.

[18] J. M. Duarte-Carvajalino, G. Yu, L. Carin, and G. Sapiro, "Task-driven adaptive statistical compressive sensing of Gaussian mixture models," available in arxiv.org.

[19] F. Leger, G. Yu, and G. Sapiro, "Efficient matrix completion with Gaussian models," Proc. *IEEE ICASSP 2011*, Prague, May 2011.