

SPARSE REPRESENTATION OF MEDICAL IMAGES VIA COMPRESSED SENSING USING GAUSSIAN SCALE MIXTURES

George Tzagkarakis and Panagiotis Tsakalides

Department of Computer Science, University of Crete & Institute of Computer Science - FORTH
e-mail: gtzag@csd.uoc.gr, tsakalid@ics.forth.gr

ABSTRACT

The increased high-resolution capabilities of modern medical image acquisition systems raise the crucial tasks of effectively storing and interacting with large databases of such data. The ease of image storage and query would be unfeasible without compression, which represents high-resolution images with a relatively small set of significant transform coefficients. Due to the specific content of medical images, compression often results in highly sparse representations in appropriate orthonormal bases. The inherent property of compressed sensing (CS) working simultaneously as a sensing and compression protocol using a small subset of random projection coefficients, enables a potentially significant reduction in storage requirements. In this paper, we introduce a Bayesian CS approach for obtaining highly sparse representations of medical images based on a set of noisy CS measurements, where the prior belief that the vector of transform coefficients should be sparse is exploited by modeling its probability distribution by means of a Gaussian Scale Mixture. The experimental results show that the proposed approach maintains the reconstruction performance of other state-of-the-art CS methods while achieving significantly sparser representations of medical images with distinct content.

Index Terms— Bayesian compressed imaging, Gaussian scale mixture, medical imaging, sparse Bayesian learning, sparse representation.

1. INTRODUCTION

The design of modern high-resolution imaging devices in medical applications have increased the amount of image data at an explosive rate. The storage and interaction with large databases of medical image data necessitates the development of efficient compression techniques and standards [1, 2]. However, even higher compression rates could potentially suffice to carry out a specific task, such as image classification and retrieval, where a high-quality reconstruction of the still images is not necessary.

Several studies [3] have shown that appropriate transforms (e.g., wavelets and sinusoids) of many natural signals often reveal certain structures allowing for compact and sparse representations. This also holds for many medical images, since they consist primarily of edges on a relatively homogeneous background. For instance, the 2-D Discrete Wavelet Transform (DWT) of such images results in a large number of coefficients with negligible amplitude and a small number of large-amplitude coefficients concentrated about the

edges. The common approach to compressing such a sparse image is to compute its transform coefficients and then store only the most “significant” ones. However, this is an inherently wasteful process (in terms of both sampling rate and computational complexity), since one gathers and processes the entire image even though an exact representation is not required explicitly.

Compressed sensing (CS) is a recently introduced framework for simultaneous sensing and compression [4, 5] enabling a potentially significant reduction in the sampling and computation costs. In particular, a signal having a sparse representation in a transform basis can be reconstructed from a small set of projections onto a second, measurement basis that is incoherent with the first one. The majority of previous studies about the sparse representation and reconstruction of a signal in an over-complete dictionary using CS, solve constrained-based optimization problems.

Several recent papers exploit the sparsity of images using CS to increase the compression rates [6, 7]. In addition, the CS framework has been already applied in the field of Magnetic Resonance Image (MRI) reconstruction with very promising results [8, 9]. Recently, a Bayesian CS (BCS) framework was introduced [10] resulting in certain improvements when compared with norm-based CS methods. In particular, the prior belief that the vector of transform coefficients should be sparse was expressed by employing a hierarchical model as a sparsity-enforcing prior distribution on the sparse coefficients vector. In the present work, we model directly the coefficients vector using a Gaussian Scale Mixture (GSM). The experimental results reveal that this approach yields a significantly sparser representation of several medical images with distinct content, while also maintaining a high reconstruction performance.

The paper is organized as follows: in Section 2, we briefly review the main concepts of BCS and introduce the GSM-based BCS method. In Section 3, we compare the performance of the proposed approach with recent state-of-the-art CS methods in terms of the degree of sparsity and the reconstruction quality, while we conclude in Section 4.

2. BAYESIAN CS RECONSTRUCTION

Let Ψ be a $N \times N$ matrix, whose columns correspond to the transform basis functions. Then, a given image $\vec{f} \in \mathbb{R}^N$ (reshaped as a column vector) can be represented as $\vec{f} = \Psi \vec{w}$, where $\vec{w} \in \mathbb{R}^N$ is the coefficient vector. Obviously, \vec{f} and \vec{w} are equivalent representations of the image, with \vec{f} being in the space domain and \vec{w} in the (transform) Ψ domain. As mentioned above, for natural images with specific content, such as edges and lines in the case of many medical images, the majority of the components of \vec{w} have negligible amplitude. In particular, \vec{f} is L -sparse in basis Ψ if \vec{w} has L non-zero components ($L \ll N$). In a real-world scenario \vec{f} is not strictly L -sparse, but it is said to be *compressible* when the re-ordered com-

This work was funded by the Greek General Secretariat for Research and Technology under Program ΠΕΝΕΔ-Code 03ΕΔ69 and by the Marie Curie TOK-DEV “ASPIRE” grant (MTKD-CT-2005-029791) within the 6th European Community Framework Program.

ponents of \vec{w} decay at a power-law.

Consider also an $M \times N$ ($M < N$) measurement matrix Φ (the over-complete dictionary), where the rows of Φ are incoherent with the columns of Ψ . For instance, let Φ be a matrix with independent and identically distributed (i.i.d.) Gaussian entries. Such matrices are incoherent with any fixed transform matrix Ψ with high probability (universality property) [5].

If \vec{f} is compressible in Ψ , then it is possible to perform directly a compressed set of measurements \vec{g} , resulting in a simplified image acquisition system. The relation between the original image \vec{f} and the CS measurements \vec{g} is obtained through random projections, $\vec{g} = \Phi \Psi^T \vec{f} = \Phi \vec{w}$, where $\Phi = [\phi_1, \dots, \phi_M]^T$ and $\phi_m \in \mathbb{R}^N$ is a random vector with i.i.d. components. Thus, the reconstruction of \vec{f} from \vec{g} reduces to estimating the sparse weight vector \vec{w} .

Most of the recent literature on CS [11, 12] has concentrated on solving constrained-based optimization problems for sparse signal representation. For instance, in the case of CS measurements corrupted by additive noise $\vec{\eta}$ with unknown variance σ_η^2 , $\vec{g} = \Phi \vec{w} + \vec{\eta}$, the ℓ_1 -norm minimization approach seeks a sparse vector \vec{w} by solving the following optimization problem,

$$\tilde{\vec{w}} = \arg \min_{\vec{w}} \|\vec{w}\|_1, \text{ s.t. } \|\vec{g} - \Phi \vec{w}\|_\infty \leq \epsilon, \quad (1)$$

where ϵ is the noise level ($\|\vec{\eta}\|_2 \leq \epsilon$). The main approaches for the solution of such optimization problems include linear programming [13] and greedy algorithms [14], resulting in a *point estimate* of the weight vector \vec{w} .

On the other hand, when working in a probabilistic framework, then given the prior belief that \vec{w} is sparse in basis Ψ and the set of CS measurements \vec{g} , the objective is to formulate a *posterior probability distribution* for \vec{w} . This improves the accuracy over a point estimate and provides confidence intervals (error bars) in the approximation of \vec{f} , which can be used to guide the optimal design of additional CS measurements with the goal of reducing the uncertainty in reconstructing \vec{f} .

Under the common assumption of a zero-mean Gaussian noise, we obtain the following Gaussian likelihood model,

$$p(\vec{g}|\vec{w}, \sigma_\eta^2) = (2\pi\sigma_\eta^2)^{-M/2} \cdot \exp\left(-\frac{1}{2\sigma_\eta^2} \|\vec{g} - \Phi \vec{w}\|^2\right). \quad (2)$$

Assuming that Φ is known, the quantities to be estimated, given the CS measurements \vec{g} , are the sparse weight vector \vec{w} and the noise variance σ_η^2 . This is equivalent to seeking a full posterior density function for \vec{w} and σ_η^2 .

In this probabilistic framework, the assumption that \vec{w} is sparse is formalized by modeling its distribution using a sparsity-enforcing prior. A common choice of this prior is the Laplace density [15]. However, the use of a Laplace prior density raised the problem that the Bayesian inference may not be performed in closed form, since the Laplace prior is not conjugate¹ to the Gaussian likelihood model. The treatment of the CS measurements \vec{g} from a Bayesian viewpoint, while overcoming the problem of conjugateness, was introduced in [10] by replacing the Laplace prior of \vec{w} with a hierarchical model, which had similar properties as the Laplace but allowed convenient conjugate-exponential analysis [16].

2.1. BCS sparse representation using GSM priors

In the present work, the sparse representation of \vec{w} is also performed in a Bayesian framework. However, in contrast to previous meth-

¹In probability theory, a family of prior probability distributions $p(s)$ is said to be conjugate to a family of likelihood functions $p(x|s)$ if the resulting posterior distribution $p(s|x)$ is in the same family as $p(s)$.

ods, the proposed method consists in modeling directly the prior of \vec{w} with a heavy-tailed distribution, which promotes its sparsity. For this purpose, we approximate the prior distribution of \vec{w} by means of a Gaussian Scale Mixture (GSM). This means that \vec{w} can be written in the form $\vec{w} = \sqrt{A} \vec{G}$, where A is a positive random variable and $\vec{G} = (G_1, G_2, \dots, G_N)$ is a zero-mean Gaussian random vector, independent of A , with covariance matrix Σ . The additional assumption that the components of \vec{G} are independent yields a diagonal covariance matrix $\Sigma = \text{diag}(\sigma_1^2, \dots, \sigma_N^2)$.

From the above, the density of \vec{w} conditioned on the variable A is a zero-mean multivariate Gaussian given by,

$$p(\vec{w}|A) = \frac{\exp(-\frac{1}{2} \vec{w}^T (A \Sigma)^{-1} \vec{w})}{(2\pi)^{N/2} |A \Sigma|^{1/2}}, \quad (3)$$

where $|\cdot|$ denotes the determinant of a matrix. From (3), we obtain the following simple expression for the maximum likelihood (ML) estimate of the variable A ,

$$\hat{A}(\vec{w}) = (\vec{w}^T \Sigma^{-1} \vec{w}) / N. \quad (4)$$

Assuming that the noise variance σ_η^2 , the value of A and the covariance matrix Σ have been estimated, given the CS measurements \vec{g} and the matrix Φ , the posterior of \vec{w} is given by the Bayes' rule,

$$p(\vec{w}|\vec{g}, A, \Sigma, \sigma_\eta^2) = \frac{p(\vec{g}|\vec{w}, \sigma_\eta^2) p(\vec{w}|A, \Sigma)}{p(\vec{g}|A, \Sigma, \sigma_\eta^2)}, \quad (5)$$

which is a multivariate Gaussian distribution whose mean $\vec{\mu}$ and covariance \mathbf{P} are given by,

$$\vec{\mu} = \sigma_\eta^{-2} \mathbf{P} \Phi^T \vec{g}, \quad (6)$$

$$\mathbf{P} = (\sigma_\eta^{-2} \Phi^T \Phi + \mathbf{M})^{-1}, \quad (7)$$

where $\mathbf{M} = \text{diag}((A\sigma_1^2)^{-1}, \dots, (A\sigma_N^2)^{-1})$. The estimated vector \vec{w} is equal to the most probable value of the above multivariate Gaussian model, that is, $\vec{w} \equiv \vec{\mu}$.

The critical advantage offered by a Bayesian CS method, when compared with the constrained-based optimization approaches in the processing of medical images is that it fits better the true heavy-tailed statistics of the sparse vector \vec{w} . The use of the GSM model could enhance the (sparse) representation performance, since it provides an additional degree of freedom through the scale parameter A , and thus it results in a more accurate modeling of the true sparsity of the original image in the (wavelet) transform domain.

The sparse representation of the wavelet coefficient vector \vec{w} reduces to estimating the model parameters A, Σ, σ_η^2 . The unknown parameters $\sigma_\eta^2, \{\sigma_i^2\}_{i=1}^N$ can be estimated iteratively by maximizing the following marginal log-likelihood function with respect to them:

$$\begin{aligned} \mathcal{L}(\sigma_\eta^2, \{\sigma_i^{-2}\}_{i=1}^N) &= \log[p(\vec{g}|A, \sigma_\eta^2, \{\sigma_i^{-2}\}_{i=1}^N)] \\ &= -\frac{1}{2} [M \log(2\pi) + \log(|\mathbf{C}|) + \vec{g}^T \mathbf{C}^{-1} \vec{g}], \end{aligned} \quad (8)$$

where $\mathbf{C} = \frac{\sigma_\eta^2}{A} \mathbf{I} + \Phi \Sigma^{-1} \Phi^T$. As it can be seen, the proposed model is a scaled version of the previous hierarchical model by a factor of $1/A$. This factor is important, since it controls the heavy-tailed behavior of the diagonal elements of \mathbf{M} and consequently of the covariance matrix \mathbf{P} , and thus the sparsity of the estimated vector $\vec{w} \equiv \vec{\mu}$. A fast incremental algorithm is used for the addition and deletion of candidate basis functions (columns of Φ) to monotonically increase the marginal likelihood (8), by noting that the marginal log-likelihood can be decomposed in two terms,

$$\mathcal{L}(\sigma_\eta^2, \{\sigma_i^{-2}\}_{i=1}^N) = \mathcal{L}(\sigma_\eta^2, \{\sigma_i^{-2}\}_{i=1, i \neq i'}) + l(\sigma_{i'}^{-2}), \quad (9)$$

Algorithm 1 Estimation of a sparse vector \vec{w} via BCS-GSM

Input: $\Phi, \vec{g}, c \sim 10^{-3}$ **Output:** $\hat{\vec{w}} \equiv \vec{\mu}, \mathbf{P}, \sigma_n^2, \mathcal{B}$ {the set of significant basis functions}**Initialize:** $\sigma_n^2 = c \cdot \text{Var}(\vec{g})$, select basis vector $\vec{\phi}_{\cdot, i_1}$ (i_1 -th column of Φ) s.t.

$$i_1 = \arg \max_{i=1, \dots, N} \frac{\|\vec{\phi}_{\cdot, i}\|^2}{(\|\vec{\phi}_{\cdot, i}^T \vec{g}\|^2 / \|\vec{\phi}_{\cdot, i}\|^2) - \sigma_n^2},$$

set $\sigma_{i_1}^{-2} = \frac{\|\vec{\phi}_{\cdot, i_1}\|^2}{(\|\vec{\phi}_{\cdot, i_1}^T \vec{g}\|^2 / \|\vec{\phi}_{\cdot, i_1}\|^2) - \sigma_n^2}$ (all other $\{\sigma_i^{-2}\}_{i \neq i_1}$ are set to infinity), $\mathcal{B} = \{i_1\}$ 1: Compute \mathbf{P} (Eq. (7)), $\vec{\mu}$ (Eq. (6)) (initially scalars) and estimate A from Eq.(4)2: **repeat**3: **for** $i = 1, \dots, N$ **do**4: Compute $\xi_i = q_i^2 - s_i$ 5: **if** $\xi_i > 0$ and $\sigma_i^{-2} < \infty$ **then**6: re-estimate σ_i^{-2} 7: **else if** $\xi_i > 0$ and $\sigma_i^{-2} = \infty$ **then**8: add i -th basis in the model ($\mathcal{B} \leftarrow \mathcal{B} \cup \{i\}$) and update σ_i^{-2} 9: **else if** $\xi_i \leq 0$ and $\sigma_i^{-2} < \infty$ **then**10: delete i -th basis from the model ($\mathcal{B} \leftarrow \mathcal{B} \setminus \{i\}$) and set $\sigma_i^{-2} = \infty$ 11: **end if**12: Update \mathbf{P} , $\vec{\mu}$ and A (in this order)13: Update $\sigma_n^2 = \frac{\|\vec{g} - \Phi \vec{\mu}\|^2}{N - \text{card}(\mathcal{B}) + \sum_{n \in \mathcal{B}} A^{-1} \sigma_n^{-2} \mathbf{P}_{nn}}$ {*card* denotes the cardinality of a set}14: Update \mathbf{D} by performing the scaling $A \sigma_i^2$ 15: **end for**16: **until** convergence

with the first term depending on all $\vec{\phi}$ except for the i' -th variance, while the second term depends only on the i' -th variance. The iterative scheme for the estimation of the weight vector \vec{w} proceeds as shown by Algorithm 1, where the following notation is used: $s_i = \vec{\phi}_{\cdot, i}^T \mathbf{C}_{-i}^{-1} \vec{\phi}_{\cdot, i}$ and $q_i = \vec{\phi}_{\cdot, i}^T \mathbf{C}_{-i}^{-1} \vec{g}$, where \mathbf{C}_{-i} is \mathbf{C} with the contribution of the i -th basis vector ignored.

Several convergence criteria can be employed to terminate the execution of the algorithm, such as when the number of iterations exceeds a predefined maximum or when the relative decrease of the marginal log-likelihood function from one iteration to the next one falls below a small positive threshold. In our implementation we adopt the second approach, since it results in an increased reconstruction performance, while the first one could be used to reduce the computational cost.

3. EXPERIMENTAL RESULTS

In this section, we evaluate the performance of BCS-GSM by applying it on a set of six medical images of size 128×128 , which are shown in Figure 1. Each image is sparsified in the 2-D DWT domain by decomposing them in 5 scales using the Daubechies' "db 4" wavelet. The detail wavelet coefficients represent the high-frequency content of a given image and they are characterized by a highly sparse behavior, whereas the approximation coefficients correspond to a coarse representation of it. Thus, the CS algorithms are applied on the detail coefficients only and the reconstruction of the original image is performed by adding the approximation coefficients to the reconstructed image obtained from the detail coefficients. Except for the original (noiseless) images we generate two

noisy versions of them by adding zero-mean Gaussian noise resulting in SNR = 7.5, 15 dB.

In the subsequent experiments we apply several CS algorithms using a portion of the detail coefficients. In particular, if N_{detail} is the number of the detail coefficients we evaluate the performance using a subset of size $c \cdot N_{detail}$ with $c \in \{0.3, 0.4, 0.5, 0.65\}$ (or equivalently by employing 30%, 40%, 50% and 65% of the detail coefficients). The proposed BCS-GSM method is compared with the following CS techniques: 1) standard BCS, 2) BP, 3) StOMP (combined with a CFAR thresholding scheme), 4) ℓ_1 -norm minimization using the primal-dual interior point method (LIEQ-PD) and 5) the linear reconstruction, which is simply the inverse 2-D DWT and gives the optimal reconstruction².

The CS measurements \vec{g} are acquired by applying measurement matrices Φ with their columns being drawn randomly from the unit sphere on the wavelet coefficients vector \vec{w} . The quality of the reconstructed image (of size $P \times Q$) is measured via the Peak Signal-to-Noise Ratio (PSNR), which is defined as follows (in dB):

$$\text{PSNR} = 20 \log_{10} \left(\frac{\max\{I\}}{\sqrt{\frac{1}{PQ} \sum_{p=1}^P \sum_{q=1}^Q |I(p, q) - \hat{I}(p, q)|^2}} \right), \quad (10)$$

where I and \hat{I} denote the original and reconstructed image, respectively, $\max\{I\}$ is the maximum pixel value of image I and $I(p, q)$ is the pixel value at the position (p, q) . Due to space limitations, we plot the results for the images of the top row only. However, similar performance is achieved for the other three images.

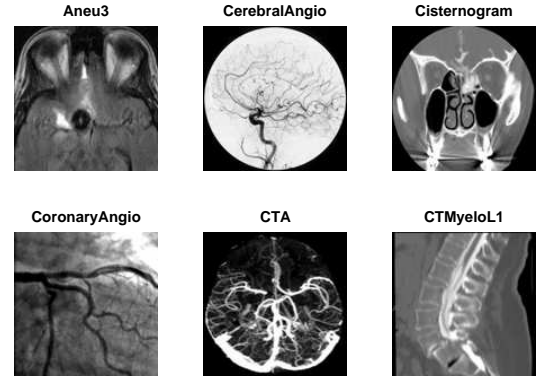


Fig. 1. Medical images (128×128) used for evaluation of the performance of BCS-GSM.

Fig. 2 shows the PSNRs between the reconstructed (noiseless and noisy) images and the corresponding original (noiseless) image, for the BCS-GSM, as well as for the other five reconstruction approaches, as a function of the number of measurements for the two SNR values. First, we observe that for the selected images the reconstruction performance of all methods decreases as the SNR decreases, something that we expected. However, it is clear that the proposed BCS-GSM method achieves practically the same PSNR with the selected CS methods and the optimal linear reconstruction. In particular, the difference in PSNR with the linear reconstruction is less than 1 dB in the noiseless case, while it is negligible in the two noisy cases. Besides, the increased number of measurements

²For the implementation of the other CS methods we used the code included in the SparseLab package that is available online at <http://sparselab.stanford.edu/>.

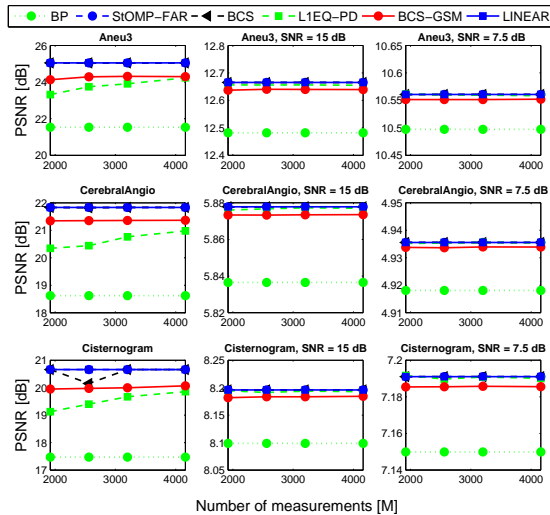


Fig. 2. PSNRs for “Aneu3”, “CerebralAngio” and “Cisternogram” as a function of M and for SNR = 7.5, 15 dB.

in the selected range $M \in [1900, 4200]$ does not affect the reconstruction PSNR as much as one would expect. A justification for this behavior is that all of the selected images consist of lines and edges spread over a relatively homogeneous background resulting in highly sparse coefficient vector. Thus, a small number of measurements is adequate in capturing the sparsity, while the addition of more measurements above a threshold improves the reconstruction quality only slightly.

The increased ability of BCS-GSM to provide a highly sparse representation in the case of medical images is highlighted in Fig. 3 depicting the corresponding CS ratio values, which we define as the ratio of the number of measurements M over the number of non-zero components of \vec{w} (sparsity) returned by the algorithm. The larger the CS ratio value, the higher the sparsity for a fixed value of M . Obviously, BCS-GSM outperforms all the other CS methods, increasing the sparsity of the representation by as much as 15 times as the number of measurements increases. In addition, this significantly improved performance is robust even in the low-SNR regime.

4. CONCLUSIONS AND FUTURE WORK

In this work, we described a probabilistic method for CS sparse representation of medical images using a GSM, which models directly the sparse coefficient vector with a heavy-tailed distribution that enforces its sparsity. The experimental results revealed a critical property of the proposed BCS-GSM approach when compared with other CS reconstruction methods. In particular, we showed that the BCS-GSM implementation maintains comparable reconstruction performance, while using much fewer basis functions and thus, resulting in an increased sparsity. The subject of our ongoing research is to apply the increased sparsity for classification and retrieval purposes reducing the storage requirements and the computational cost.

5. REFERENCES

- [1] D. Taubman and M. Marcellin, “JPEG 2000: Image Compression Fundamentals, Standards and Practice”, (Int. Series in Eng. and Comp. Sci.), Norwell, MA: Kluwer, 2002.
- [2] S. Grgic, K. Kers and M. Grgic, “Image compression using wavelets”, *Proc. of IEEE Int. Symp. on Industr. Elec.*, Vol. 1, pp.99–104, 1999.
- [3] S. Mallat, “A Wavelet Tour of Signal Processing”, 2nd ed., New York: Academic Press, 1998.

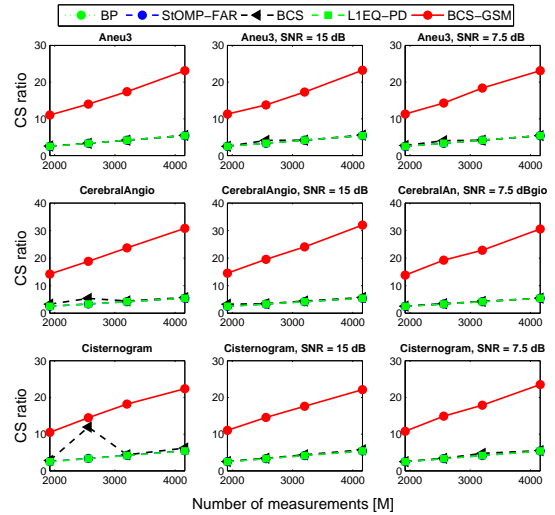


Fig. 3. CS ratios for “Aneu3”, “CerebralAngio” and “Cisternogram” as a function of M and for SNR = 7.5, 15 dB.

- [4] E. Candès, J. Romberg and T. Tao, “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information”, *IEEE Trans. Inform. Th.*, Vol. 52, No. 2, pp. 489–509, Feb. 2006.
- [5] D. Donoho, “Compressed Sensing”, *IEEE Trans. Inform. Th.*, Vol. 52, No. 4, pp. 1289–1306, Apr. 2006.
- [6] J. Haupt and R. Nowak, “Compressive sampling vs. conventional imaging”, *Int. Conf. on Image Proc. (ICIP)*, Atlanta, Oct. 2006
- [7] M. Duarte *et al.*, “Single-pixel imaging via compressive sampling”, *IEEE Signal Proc. Mag.*, Vol. 25, No. 2, pp. 83–91, Mar. 2008.
- [8] M. Lustig, D. Donoho and J. Pauly, “Sparse MRI: The application of compressed sensing for rapid MR imaging,” *Magn. Res. in Medicine*, Vol. 58, No. 6, pp. 1182–1195, Dec. 2007.
- [9] J. Trzasko, A. Manduca and E. Borisch, “Highly undersampled magnetic resonance image reconstruction via homotopic ell-0-minimization,” *IEEE Trans. Med. Imaging*, Vol. 28, No. 1, pp. 106–121, 2009.
- [10] S. Ji, Y. Xue and L. Carin, “Bayesian Compressive Sensing”, *IEEE Trans. on Signal Proc.*, Vol. 56, No. 6, pp. 2346–2356, June 2008.
- [11] Y. Tsaig, and D. L. Donoho, “Extensions of compressed sensing”, *Signal Proc.*, Vol. 86, No. 3, pp. 549–571, Mar. 2006.
- [12] J. Haupt, and R. Nowak, “Signal reconstruction from noisy random projections”, *IEEE Trans. Inform. Theory*, Vol. 52, No. 9, pp. 4036–4048, Sept. 2006.
- [13] S. Chen, D. Donoho and M. Saunders, “Atomic decomposition by Basis Pursuit”, *SIAM J. on Sci. Comp.*, Vol. 20, No. 1, pp. 33–61, 1999.
- [14] J. Tropp and A. Gilbert, “Signal recovery from partial information via Orthogonal Matching Pursuit”, Apr. 2005 [Online]. Available: <http://www.personal.umich.edu/~jtropp/papers/TG06-Signal-Recovery.pdf>
- [15] M. Figueiredo, “Adaptive sparseness using Jeffreys prior”, in *Advances in Neural Inf. Proc. Systems (NIPS 14)*, 2002.
- [16] M. Tipping, “Sparse Bayesian Learning and the Relevance Vector Machine”, *J. Mach. L. Res.*, Vol. 1, pp. 211–244, 2001.
- [17] M. Tipping and A. Faul, “Fast marginal likelihood maximisation for sparse Bayesian models”, in *Proc. 9th Int. Work. on Artif. Intell. and Stat.*, C. Bishop and B. Frey Eds., 2003.
- [18] D. Donoho *et al.*, “Sparse solution of underdetermined linear equations by Stagewise Orthogonal Matching Pursuit”, Tech. Report 06-02, Dept. of Statistics, Stanford Univ., 2006.