

BAYESIAN COMPRESSIVE SENSING OF WAVELET COEFFICIENTS USING MULTISCALE LAPLACIAN PRIORS

Esteban Vera¹, Luis Mancera², S. Derin Babacan³, Rafael Molina², and Aggelos K. Katsaggelos³

¹ Dep. of Electrical Engineering
Center for Optics and Photonics
Universidad de Concepción, Chile
estebanvera@udec.cl

² Dep. of Computer Science
and Artificial Intelligence
Universidad de Granada, Spain
mancera@decsai.ugr.es, rms@decsai.ugr.es

³ Dep. of Electrical Engineering
and Computer Science
Northwestern University, Evanston, IL
sdb@northwestern.edu, aggk@eecs.northwestern.edu

ABSTRACT

In this paper, we propose a novel algorithm for image reconstruction from compressive measurements of wavelet coefficients. By incorporating independent Laplace priors on separate wavelet sub-bands, the inhomogeneity of wavelet coefficient distributions and therefore the structural sparsity within images are modeled effectively. We model the problem by adopting a Bayesian formulation, and develop a fast greedy reconstruction algorithm. Experimental results demonstrate that the reconstruction performance of the proposed algorithm is competitive with state-of-the-art methods while outperforming them in terms of running times.

Index Terms— compressive sensing, wavelet transforms, signal reconstruction, bayesian methods.

1. INTRODUCTION

Compressive Sensing (CS) [1, 2] is a new methodology for unifying sampling and compression, allowing sampling at very low rates compared to the Nyquist rate, and requiring that the transmitted signal, or some linear representation of it, is sparse or at least compressible (near-sparse). The CS measurements typically consist of a limited set of random projections, usually less than the length of the input signal, from where the original information should be recovered. Sparsity prior information can be incorporated into the CS reconstruction problem by minimizing the number of non-zero coefficients, or the l_0 -norm, of the estimated (decoded) signal, which results in a highly complex non-convex inverse problem. Hence, as it is commonly done in the literature [3, 4, 5, 6], sparsity can also be imposed by minimizing instead the l_1 -norm, as a convex approximation to the l_0 -norm.

Together with the near-sparsity typically provided by the responses of wavelet coefficients to a wide range of piecewise smooth signals and natural images, we can observe that the

This work has been partially supported by: the Spanish research programme Consolider Ingenio 2010: MIPRCV (CSD2007-00018); the Spanish Ministerio de Educación y Ciencia under contract TIN2007-65533; and the Chilean government programmes Conicyt PBCT and Mecesup FSM601.

distribution of the coefficients is not homogeneous through all scales, and even through different sub-bands at each scale (e.g., [7]). The idea of incorporating *a priori* structural information into the CS decoding process has been recently proposed in the Model-Based CS approach in [8, 9], where a clustered parent-child relationship of the Hidden Markov Tree (HMT) model is used for the wavelet coefficients [10]. Thus, a fixed statistical model for the wavelet tree statistics is obtained after a training stage, leading to an ad-hoc, but efficient, solution. On the other hand, by adopting a similar HMT model, a Bayesian CS approach is developed in [11] which also exploits the parent-child relationships along the wavelet tree, but the inference is performed online by statistical sampling methods, which are computationally expensive.

In this paper, we develop a Bayesian formulation which exploits the multiscale structure of the wavelet coefficients and models each subband separately by utilizing different independent Laplace priors. We propose a reconstruction algorithm which jointly estimates the unknown wavelet coefficients and the required algorithmic parameters. We demonstrate the advantages of separate modeling of wavelet sub-bands with experimental results and show that the proposed method provides state-of-the-art reconstruction performance with much faster convergence rates.

2. PROBLEM FORMULATION

Let \mathbf{w} be an $N \times 1$ signal, Φ an $M \times N$ matrix with $M < N$, and \mathbf{n} a white i.i.d. Gaussian $N \times 1$ noise vector. The CS encoding can be expressed as

$$\mathbf{y} = \Phi \mathbf{w} + \mathbf{n}, \quad (1)$$

where the measurement matrix Φ has to satisfy the Restricted Isometry Property [1] and the signal \mathbf{w} has to be sparse. By considering \mathbf{w} as the wavelet coefficient vector provided by the Wavelet Transform (WT) of a given signal or image \mathbf{x} , and by ignoring the scaling coefficients, without loss of generality we can think of the CS problem as if the measurement matrix is applied directly on the wavelet coefficients instead of the

image. Thus, the classical convex CS decoding problem is to recover \mathbf{w} by minimizing

$$\hat{\mathbf{w}} = \underset{\mathbf{w}}{\operatorname{argmin}} \{ \beta \|\mathbf{y} - \Phi \mathbf{w}\|_2^2 + \lambda \|\mathbf{w}\|_1 \}.$$

In our approach we divide \mathbf{w} into its B sub-bands, $\mathbf{w} = (\mathbf{w}_1^t, \dots, \mathbf{w}_B^t)^t$, and we enforce sparsity independently on each sub-band by formulating the problem

$$\hat{\mathbf{w}} = \underset{\mathbf{w}}{\operatorname{argmin}} \{ \beta \|\mathbf{y} - \Phi \mathbf{w}\|_2^2 + \sum_{b=1}^B \lambda_b \|\mathbf{w}_b\|_1 \}. \quad (2)$$

Note that no inter-scale constraints are imposed between the sub-bands, but different regularizers are associated with each sub-band. We will examine next these concepts from the Bayesian point of view, and also provide estimates for all the needed parameters.

3. BAYESIAN MODELING

We formulate the problem in Eq. (2) following the Bayesian approach in [6, 12]. The observation model is given by

$$p(\mathbf{y}|\mathbf{w}, \beta) = \mathcal{N}(\mathbf{y}|\Phi \mathbf{w}, \beta^{-1}),$$

and the l_1 regularization term can be reformulated by assuming independent Laplacian priors for each sub-band \mathbf{w}_b

$$\begin{aligned} p(\mathbf{w}|\lambda) &= \prod_{b=1}^B p(\mathbf{w}_b|\lambda_b) = \prod_{b=1}^B \prod_{i=1}^{N_b} \lambda_b \exp(-\lambda_b |w_{bi}|) \\ &= \prod_{b=1}^B \lambda_b^{N_b} \exp(-\lambda_b \|\mathbf{w}_b\|_1), \end{aligned} \quad (3)$$

where B is the number of sub-bands and N_b the number of elements in each sub-band \mathbf{w}_b .

To perform inference in a tractable way using the Laplace prior, we utilize the hierarchical form by first defining

$$p(\mathbf{w}_b|\gamma_b) = \prod_{i=1}^{N_b} \mathcal{N}(w_{bi}|0, \gamma_{bi}),$$

where $\gamma_b = (\gamma_{b1}, \dots, \gamma_{bN_b})$, and then using

$$\begin{aligned} p(\gamma_{bi}|\lambda_b) &= \Gamma(\gamma_{bi}|1, \lambda_b/2) = \frac{\lambda_b}{2} \exp\left(-\frac{\lambda_b \gamma_{bi}}{2}\right), \\ &i = 1, \dots, N_b. \end{aligned} \quad (4)$$

Notice that by integrating out the auxiliary parameters γ_{bi} , we finally obtain the desired Laplacian distribution prior on every sub-band b , which is governed by the parameter λ_b :

$$\begin{aligned} p(\mathbf{w}_b|\lambda_b) &= \int p(\mathbf{w}_b|\gamma_b) p(\gamma_b|\lambda_b) d\gamma_b \\ &= \prod_{i=1}^{N_b} \int p(w_{bi}|\gamma_{bi}) p(\gamma_{bi}|\lambda_b) d\gamma_{bi} \\ &= \frac{\lambda_b^{N_b/2}}{2^{N_b}} \exp\left(-\sqrt{\lambda_b} \sum_{i=1}^{N_b} |w_{bi}|\right). \end{aligned}$$

In addition, a Jeffrey's prior is assumed for each one of the parameters λ_b , that is $p(\lambda_b) \propto 1/\lambda_b$.

The joint distribution of our hierarchical Bayesian model is defined by $p(\mathbf{w}, \gamma, \lambda, \beta, \mathbf{y}) = p(\mathbf{y}|\mathbf{w}, \beta) p(\mathbf{w}|\gamma) p(\gamma|\lambda) p(\lambda)$.

3.1. Bayesian Inference

Following the work in [6, 12, 5, 13], the inference procedure is based on the following decomposition:

$$p(\mathbf{w}, \gamma, \lambda, \beta|\mathbf{y}) = p(\mathbf{w}|\gamma, \lambda, \beta, \mathbf{y}) p(\gamma, \lambda, \beta|\mathbf{y}).$$

The term $p(\mathbf{w}|\gamma, \lambda, \beta, \mathbf{y})$ is found to be proportional to the joint distribution $p(\mathbf{w}, \gamma, \lambda, \beta, \mathbf{y})$, and is a multivariate Gaussian distribution with parameters

$$\begin{aligned} \Sigma &= (\beta \Phi^T \Phi + \Lambda)^{-1}, \\ \mu &= \Sigma \beta \Phi^T \mathbf{y}, \end{aligned}$$

with Λ a block diagonal matrix where each diagonal block has the form $\Lambda_b = \operatorname{diag}(1/\gamma_{bi})$. Since $p(\gamma, \lambda, \beta|\mathbf{y})$ is proportional to the joint distribution $p(\mathbf{y}, \gamma, \beta, \lambda)$, then the hyperparameters can be estimated by maximizing

$$\begin{aligned} \mathcal{L} &= \ln \int p(\mathbf{y}|\mathbf{w}, \beta) p(\mathbf{w}|\gamma) p(\gamma|\lambda) p(\lambda) d\mathbf{w} \\ &= -\frac{1}{2} \ln |\mathbf{C}| - \frac{1}{2} \mathbf{y}^T \mathbf{C}^{-1} \mathbf{y} + \sum_{b=1}^{N_b} (N_b - 1) \ln \lambda_b \\ &\quad - \sum_{b=1}^{N_b} \frac{\lambda_b}{2} \sum_i \gamma_{bi}, \end{aligned} \quad (5)$$

where $\mathbf{C} = (\beta^{-1} \mathbf{I} + \Phi \Lambda^{-1} \Phi^T)$. By maximizing \mathcal{L} for each λ_b , we obtain the following update equation:

$$\lambda_b = \frac{N_b - 1}{\sum_{i=1}^{N_b} \gamma_{bi}/2}. \quad (6)$$

For convenience, and to match the iterative update estimation of the parameters γ_{bi} , we rewrite matrix \mathbf{C} as:

$$\mathbf{C} = \beta^{-1} \mathbf{I} + \sum_{l_j \neq b_i} \gamma_{lj} \phi_{lj} \phi_{lj}^T + \gamma_{bi} \phi_{bi} \phi_{bi}^T = \mathbf{C}_{-bi} + \gamma_{bi} \phi_{bi} \phi_{bi}^T.$$

Thus, for each iteration on bi , given band $b \in \{1, \dots, B\}$ and coefficient $i \in \{1, \dots, N_b\}$, we can define

$$s_{bi} = \phi_{bi}^t \mathbf{C}_{-bi}^{-1} \phi_{bi} \quad q_{bi} = \phi_{bi}^t \mathbf{C}_{-bi}^{-1} \mathbf{y}.$$

Using these quantities as in [5, 6], if $q_{bi}^2 - s_{bi} < \lambda_b$, then γ_{bi} is set equal to zero and the corresponding basis vector ϕ_{bi} is pruned from the model. Otherwise, if $q_{bi}^2 - s_{bi} \geq \lambda_b$, then γ_{bi} is updated using

$$\gamma_{bi} = \frac{-s_{bi}(s_{bi} + 2\lambda_b) + s_{bi} \sqrt{(s_{bi} + 2\lambda_b)^2 - 4\lambda_b(s_{bi} - q_{bi}^2 + \lambda_b)}}{2\lambda_b s_{bi}^2}. \quad (7)$$

Note that this procedure provides a systemic way of adding relevant bases to the model, as well as pruning irrelevant ones. At each iteration only one basis is updated, which is chosen as the one that results in the maximum increase of the likelihood in Eq. 5. Then, the corresponding parameters γ_{bi} and λ_b are calculated using Eq. 7 and 6, respectively. Finally, the noise parameter β is approximated using the observation only.

4. EXPERIMENTAL RESULTS

For both computational and storage reasons, we restricted the experiments to image patches of size 64×64 pixels, thus using up to 4096 wavelet coefficients. Four standard test images were used, namely: *Barbara*, *Boat*, *Cameraman* and *House*, whose respective crops begin at pixels: (27,253), (361,181), (95,36) and (11,91). To assure a near sparse behavior on the wavelet sub-bands for such small image patches, we only performed a two scale decomposition using the Haar wavelet, and did not consider the low-pass residual during the CS sampling and recovery processes. Figures of merits were finally averaged over five runs of different realizations of the measurement random matrix Φ with an added Gaussian observation noise with standard deviation $\sigma_n = 0.3$. The proposed algorithm (LAPMS) was compared with: Bayesian CS using Laplace priors (LAP) [6]; Bayesian CS (BCS) [5]; Basis Pursuit (BP) [3]; GPSR [4]; and the Tree-Structured Wavelet CS (TSWCS) [11]. The performance of the algorithms was measured in terms of both the reconstruction error and the execution time. The reconstruction error was calculated by the relative error $\|\mathbf{w} - \hat{\mathbf{w}}\|_2 / \|\mathbf{w}\|_2$, and the execution time was obtained directly from the reported time of the MATLAB© implementation.

The reconstruction error achieved for each image using different number of measurements is shown in Figure 1. We observe that LAPMS outperforms the original single scale LAP algorithm, and it is overall only slightly worse than both BP and TSWCS, while outperforming both on the *Boat* image. Furthermore, LAPMS is clearly better than BCS and GPSR for all images.

We have also observed that both LAP and LAPMS perform comparably better with respect to the other methods as the input signals become sparser. Since the reduced dimensionality of the patches decreases the sparsity of their wavelet representations, it is expected that even better comparative results will be obtained on larger images.

In addition, it can be shown that the wavelet coefficients for the test images show different sparsity patterns across different scales and orientations. For instance, hard-thresholding of the coefficients demonstrate that diagonal details are much sparser than both horizontal and vertical ones. This diversity is effectively captured by LAPMS, appropriately estimating the parameters λ_b aimed to control the sparsity for each sub-band. Specifically, empirical results indicated that each estimated λ_b is inversely proportional to the sparsity

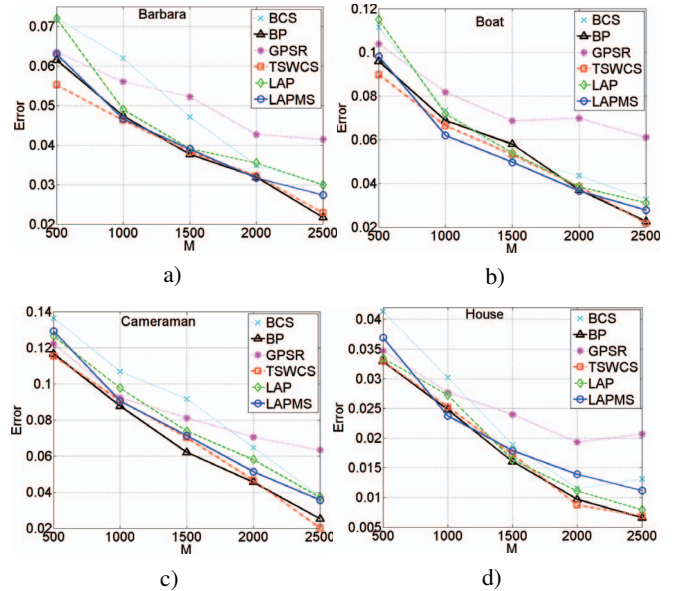


Fig. 1. Comparison on the Reconstruction Error of the CS algorithms for different images. a) *Barbara*; b) *Boat*; c) *Cameraman*; d) *House*.

level found on its corresponding original wavelet sub-band (not shown due to space limitations).

Figure 2 shows a visual example comparing the reconstruction of the CS methods tested using the *House* patch. Note that the methods providing multiscale adaptation, namely TSWCS and LAPMS, achieve fewer visual artifacts and a better reconstruction of the edge of the roof. Nevertheless, LAPMS only needed a third of the time required by TSWCS. Note also that BP, while providing a low error, shows strong artifacts, which is a fact observed throughout most of our experiments.

Table 1 (upper part) shows the averaged results for the reconstruction error obtained from the four plots in Figure 1, where LAPMS consistently outperforms the original LAP method, while keeping an acceptable performance compared to both BP and TSWCS. On the other hand, Table 1 (bottom part) shows the averaged execution time of the methods. Although on average LAPMS is only 3% better in reconstruction performance than the original LAP, it converges faster by a 10% factor. In addition LAPMS is at least twice as fast as TSWCS, and even four times faster than BP, with a slight decrease in the reconstruction accuracy on the order of 10%.

5. CONCLUSIONS

In this work we presented a novel algorithm for image reconstruction from compressive measurements of wavelet coefficients. In order to handle *a priori* information on the typically variant structure of wavelet coefficients of natural images, a multiscale Laplacian prior was incorporated into a hierarchi-

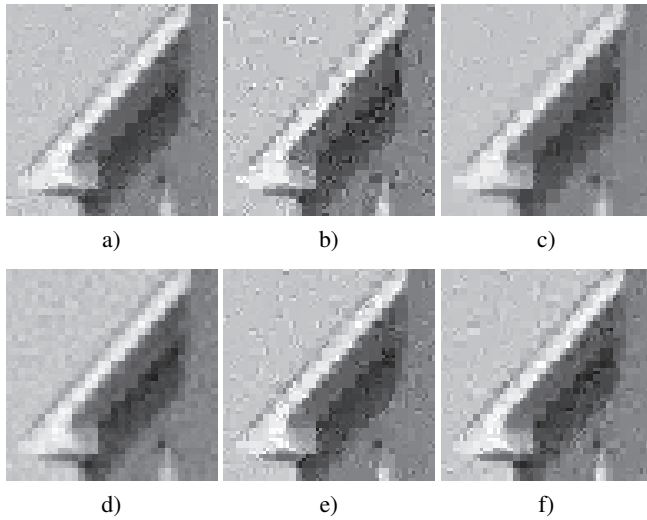


Fig. 2. Reconstructed *House* image from 1000 measurements using the following CS methods: a) BCS (error: 0.030, time: 21s.); b) BP (0.025, 74s.); c) GPSR (0.028, 80s.); 4) TSWCS (0.025, 91s.); e) LAP (0.027, 35s.); f) LAPMS (0.024, 30s.)

cal Bayesian model leading to an efficient constructive implementation. Experimental results show that the reconstruction performance is significantly improved due to the utilization of multiscale signal priors. The proposed method also provides a competitive performance compared to other state-of-the-art methods in terms of reconstruction error. Moreover, it is also at least two to four times faster than existing algorithms with lower reconstruction error. These results reinforce the importance of using *a priori* information regarding the structural nature of the data to be recovered, and also point out the favorable trade off between execution time and reconstruction error of the proposed algorithm. Future work will include more advanced prior models for the wavelet coefficients.

6. REFERENCES

- [1] E.J. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. on Inf. Theory*, **52**, 2, 489–509, Feb. 2006.
- [2] D.L. Donoho, "Compressed sensing," *IEEE Trans. on Inf. Theory*, **52**, 4, 1289–306, Apr. 2006.
- [3] D. Donoho and Y. Tsaig, "Fast solution of l_1 -norm minimization problems when the solution may be sparse," *Tech. Report*, 2006.
- [4] M.A.T. Figueiredo, R.D. Nowak and S.J. Wright, "Gradient projection for sparse reconstruction: Application to Compressed Sensing and other inverse problems," *IEEE J. Sel. Topics in Sig. Proc.*, **1**, 4, 586–98, 2007.
- [5] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Trans. on Sig. Proc.*, **56**, 6, 2346–56, Jun. 2008.
- [6] S.D. Babacan, R. Molina, and A.K. Katsaggelos, "Bayesian compressive sensing using Laplace priors," *IEEE Trans. on Image Proc.*, (to appear), 2009.
- [7] R.M. Figueras and E.P. Simoncelli, "Statistically driven sparse image approximation," in *IEEE Int. Conf. on Image Proc. (ICIP)*, 2007.
- [8] M.F. Duarte, M.B. Wakin, and R.G. Baraniuk, "Wavelet-domain compressive signal reconstruction using a hidden Markov tree model," in *IEEE Int. Conf. on Acoust., Speech, and Signal Processing (ICASSP)*, 2008.
- [9] R.G. Baraniuk, V. Cevher, M.F. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Trans. on Inf. Theory*, (submitted) 2008.
- [10] J.K. Romberg, H. Choi, and R.G. Baraniuk, "Bayesian tree-structured image modeling using wavelet-domain hidden markov models," *IEEE Trans. on Image Proc.*, **10**, 7, 1056–68, Jul. 2001.
- [11] L. He and L. Carin, "Exploiting structure in wavelet-based Bayesian compressive sensing," *IEEE Trans. on Image Proc.*, (submitted) 2008.
- [12] S.D. Babacan, R. Molina, and A.K. Katsaggelos, "Fast Bayesian compressive sensing using Laplace priors," in *IEEE Int. Conf. on Acoust., Speech, and Sig. Proc. (ICASSP)*, 2009.
- [13] M. Tipping, "Sparse Bayesian learning and the relevance vector machine," *Journal of Machine Learning Research*, vol. 1, pp. 211–244, Jun. 2001.

		Avg. Reconst. Error				
# meas.		500	1000	1500	2000	2500
BCS		0.0903	0.0681	0.0521	0.0387	0.0284
GPSR		0.0810	0.0644	0.0565	0.0505	0.0466
TSWCS		0.0733	0.0572	0.0447	0.0315	0.0179
BP		0.0767	0.0572	0.0434	0.0310	0.0190
LAP		0.0867	0.0614	0.0458	0.0357	0.0265
LAPMS		0.0818	0.0557	0.0445	0.0334	0.0255
		Time (s.)				
# meas.		500	1000	1500	2000	2500
BCS		4.28	16.76	49.75	77.61	89.66
GPSR		0.63	0.81	0.89	1.27	1.70
TSWCS		71.75	93.95	113.08	170.48	156.62
BP		21.15	76.47	169.81	293.92	463.86
LAP		6.54	35.55	55.17	71.62	84.01
LAPMS		6.47	32.89	40.84	59.46	68.33

Table 1. Performance comparison on different CS recovery algorithms in terms of averaged reconstruction error and execution time for a range of number of measurements.