NON-CONVEX SPARSE OPTIMIZATION THROUGH DETERMINISTIC ANNEALING AND APPLICATIONS

Luis Mancera[*]

Dept. of Comp. Science and A.I.
Universidad de Granada
Granada, Spain
mancera@decsai.ugr.es

Javier Portilla

Instituto de Óptica
CSIC
Madrid, Spain
portilla@io.cfmac.csic.es

We propose a new formulation to the sparse approximation problem for the case of tight frames which allows to minimize the cost function using gradient descent. We obtain a generalized version of the iterative hard thresholding (IHT) algorithm, which provides locally optimal solutions. In addition, to avoid non-favorable minima we use an annealing technique consisting of gradually de-smoothing a previously smoothed version of the cost function. This results in decreasing the threshold through the iterations, as some authors have already proposed as a heuristic. We have adapted and applied our method to restore images having localized information losses, such as missing pixels. We present high-performance in-painting *results.*

Index Terms— Sparse approximation, ℓ_0-norm minimization, in-painting.

1. INTRODUCTION

Given an observed vector and a set of vectors defining a redundant dictionary, the sparse approximation problem can be stated as minimizing the vector's approximation error using a linear combination of a given number of vectors from the dictionary. This problem seems to be very relevant for both living beings and artificial systems that analyze and process stimuli. It very likely plays a role in verbal communication, vision, and, in general, in tasks involving mixed source identification and efficient coding/synthesis. Most degradation sources decrease the sparseness of the wavelet coefficients, and thus we can compensate for part of the degradation by finding sparse approximations to the observations (e.g., [1, 2]).

The exact solution to this problem requires a combinatorial search. Some authors have explored more tractable variants. Greedy methods approximate the image by incrementally selecting those vectors best describing the part not yet represented (e.g., [3, 4]). Also, if we minimize the sum of the absolute values of the coefficients (ℓ_1-norm) instead of the number of active vectors (ℓ_0-norm), then the optimization problem becomes convex (e.g., [5, 6]). Finally, iterative algorithms have been proposed, both for convex relaxation (as iterative soft thresholding method, IST, e.g. [7, 8]) and using hard thresholding, (iterative hard thresholding method, IHT, e.g. [9, 10]). They can be improved by some heuristics, like using decreasing thresholds.

Here, we show that it is possible to conjugate classical optimization methods with competitive results without using convex or greedy approximations. We derive the IHT method through gradient descent on a continuous cost function equivalent to that of the sparse approximation problem. We then use a deterministic annealing-like technique, through a homotopy [11] to avoid non-favorable local minima. We end up with a method already used as a heuristic (e.g., [9, 8]) but, up to our knowledge, we are first to propose a theoretically justified derivation. We have already shown [12] outstanding energy compaction performance of this method. Here we apply it to restoration, obtaining high-performance *in-painting* results.

[*]Both authors funded by grant TEC2006/13845/TCM from the Ministerio de Ciencia y Tecnología, Spain.

2. THE SPARSE APPROXIMATION PROBLEM

Let $\mathbf{\Phi}$ be a $N \times M$ matrix with $M > N$ and $rank(\mathbf{\Phi}) = N$. Then, for an image $\mathbf{x} \in \mathbb{R}^N$, the problem $\mathbf{\Phi a} = \mathbf{x}$ has infinite solutions in $\mathbf{a} \in \mathbb{R}^M$. We look for compressible solutions, that is, vectors that can be represented as a sum of one vector having a small proportion of non-zero coefficients (sparse) plus a small correction vector. The following optimization is a popular way to obtain these solutions:

$$\hat{\mathbf{a}}^0(\lambda) = \arg\min_{\mathbf{a}}\{\|\mathbf{a}\|_0 + \lambda\|\mathbf{\Phi a} - \mathbf{x}\|_2^2\}, \qquad (1)$$

where $\|\mathbf{a}\|_0$ means the ℓ_0-norm of $\mathbf{a}$, (number of non-zero elements) and $\lambda \in \mathbb{R}^*$ balances the accuracy vs. the sparseness of the solution.

3. A GRADIENT-DESCENT APPROACH

3.1. Alternative continuous formulation

Here we assume that $\mathbf{\Phi}^T$ is a Parseval frame, so $\mathbf{\Phi\Phi}^T = \mathbf{I}$ and, thus, $||\mathbf{\Phi}^T\mathbf{x}||_2 = ||\mathbf{x}||_2$, for all $\mathbf{x} \in \mathbb{R}^N$. Under such condition, we prove next that Eq. (1) can be re-written as:

$$(\hat{\mathbf{a}}, \hat{\mathbf{b}}) = \arg\min_{\mathbf{a},\mathbf{b}}\{\|\mathbf{a}\|_0 + \lambda\|\mathbf{b} - \mathbf{a}\|_2^2 \text{ s.t. } \mathbf{\Phi b} = \mathbf{x}\}. \qquad (2)$$

We show that $\hat{\mathbf{a}} = \hat{\mathbf{a}}^0(\lambda)$. Firstly we express Eq. (2) as:

$$\hat{\mathbf{a}} = \arg\min_{\mathbf{a}}\{\|\mathbf{a}\|_0 + \lambda \min_{\mathbf{b}}\{\|\mathbf{b} - \mathbf{a}\|_2^2 \text{ s.t. } \mathbf{\Phi b} = \mathbf{x}\}\}.$$

For a given $\mathbf{a}$, the solution to the inner minimization yields $\tilde{\mathbf{b}}(\mathbf{a}) = \mathbf{a} + \mathbf{\Phi}^T(\mathbf{\Phi a} - \mathbf{x})$. Substituting in Eq. (2), and using the fact that $\mathbf{\Phi}^T$ is a Parseval frame, we obtain:

$$\hat{\mathbf{a}} = \arg\min_{\mathbf{a}}\{\|\mathbf{a}\|_0 + \lambda\|\mathbf{\Phi a} - \mathbf{x}\|_2^2\} = \hat{\mathbf{a}}^0(\lambda).$$

□. Now, we re-write Eq. (2) as:

$$\hat{\mathbf{b}} = \arg\min_{\mathbf{b}}\{\min_{\mathbf{a}}\{\|\mathbf{a}\|_0 + \lambda\|\mathbf{b} - \mathbf{a}\|_2^2\} \text{ s.t. } \mathbf{\Phi b} = \mathbf{x}\}. \qquad (3)$$

We see that minimizing this cost function in $\mathbf{a}$ for a given $\mathbf{b}$ can be done by minimizing independently for each index. We express the cost as $c(\mathbf{a},\mathbf{b}) = \sum_{i=1}^{M} c'(a_i, b_i)$, where:

$$c'(a,b) = \begin{cases} 1+\lambda(b-a)^2, & |a| > 0 \\ \lambda b^2, & |a| = 0. \end{cases}$$

It is easy to see that if the value $\tilde{a}_i(b_i)$ minimizing $c'(a_i, b_i)$ is not zero, then $\tilde{a}_i(b_i) = b_i$, and $c'(\tilde{a}_i(b_i), b_i) = 1$. If it is zero, then $c'(\tilde{a}_i(b_i), b_i) = \lambda b_i^2$. Then $c(\tilde{\mathbf{a}}(\mathbf{b}), \mathbf{b}) = \sum_{i=1}^{M} \min(1, \lambda b_i^2)$. Given some λ value, we note θ the b_i value producing the same cost in the ℓ_0 term as in the ℓ_2 term, which holds $\lambda\theta^2 = 1$. Therefore $\theta = \lambda^{-\frac{1}{2}}$, and we can write:

$$\tilde{a}_i(b_i) = \begin{cases} b_i, & |b_i| > \theta \\ 0, & |b_i| \leq \theta. \end{cases}$$

This is a hard-thresholding operation with threshold θ, which we note $\tilde{\mathbf{a}}(\mathbf{b}) = S_0(\mathbf{b}, \theta)$. Substituting $S_0(\mathbf{b}, \theta)$ for $\mathbf{a}$, in Eq. (3):

$$\hat{\mathbf{b}} = \arg\min_{\mathbf{b}} \{\|S_0(\mathbf{b},\theta)\|_0 + \lambda\|\mathbf{b} - S_0(\mathbf{b},\theta)\|_2^2 \text{ s.t. } \mathbf{\Phi b} = \mathbf{x}\}.$$

This result can be expressed as:

$$\begin{aligned} \hat{\mathbf{b}} &= \arg\min_{\mathbf{b}} \{C(\mathbf{b},\theta) \text{ s.t. } \mathbf{\Phi b} = \mathbf{x}\}, \qquad (4) \\ \hat{\mathbf{a}} &= S_0(\hat{\mathbf{b}}, \theta), \end{aligned}$$

where:

$$C(\mathbf{b},\theta) = \sum_{i=1}^{M} \min\left(1, \left(\frac{b_i}{\theta}\right)^2\right). \qquad (5)$$

3.2. Local minimization

The gradient[1] of the unconstrained cost function of Eq. (5) is:

$$\nabla C(\mathbf{b},\theta) = \frac{2}{\theta^2}\left(\mathbf{b} - S_0(\mathbf{b},\theta)\right).$$

In order to do gradient descent under the perfect reconstruction (PR) constraint, we start from a vector $\mathbf{b}^{(0)}$ holding the constraint, and project the gradient onto the corresponding PR affine subspace, which yields $\nabla^{PR}C(\mathbf{b},\theta) = (\mathbf{I} - \mathbf{\Phi}^T\mathbf{\Phi})\nabla C(\mathbf{b},\theta)$. This is the null component of the gradient with respect to $\mathbf{\Phi}$, so $\mathbf{b} - \alpha\nabla^{PR}C(\mathbf{b},\theta)$ will provide PR, no matter which α value we use. A necessary (and, in our case, sufficient) condition for achieving a local minimum of the constrained cost function is $\nabla^{PR}C(\mathbf{b}^*,\theta) = \mathbf{0}$. This is the fixed point of the gradient descent iterations:

$$\mathbf{b}^{(k+1)} = \mathbf{b}^{(k)} - 2\alpha\theta^{-2}(\mathbf{I} - \mathbf{\Phi}^T\mathbf{\Phi})\left(\mathbf{b}^{(k)} - S_0(\mathbf{b}^{(k)},\theta)\right).$$

The choice $\alpha = \frac{\theta^2}{2}$ minimizes for a single step descent the unconstrained cost function of Eq. (5), and it results in the IHT algorithm. Recently, [10] has proved, in a parallel and independent work, convergence of IHT to a local minimum of Eq. (1).

[1]The discontinuity for $b_i = \theta_i$ does not cause problems in our methods.

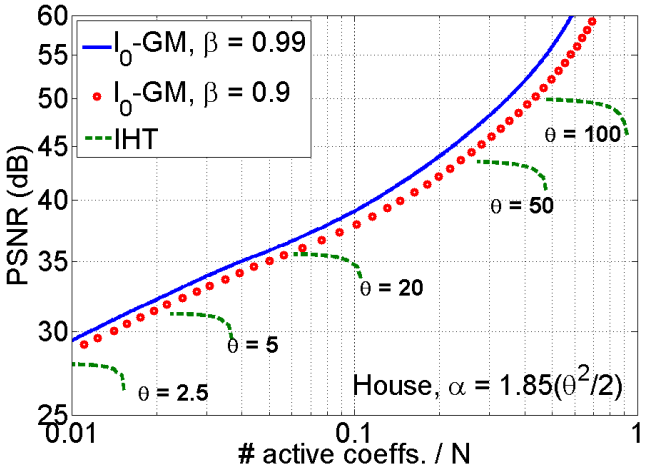

Fig. 1. Fidelity-sparseness results of ℓ_0-GM, using $\beta = 0.9$ (circles, 150 iters.) and $\beta = 0.99$ (solid, 1,500 iters.), w.r.t. IHT, using several thresholds (dashed, 100,000 iters.). We used here the image *House* and Kingsbury's DT-CWT with 8-scales.

3.3. Global minimization: a deterministic *annealing*

IHT is expensive in computational terms and requires a previous knowledge of λ. Besides that, it gets trapped into non-favourable local optima. In this section we try to overcome these problems. First note that the cost function in Eq. (5) can be written as:

$$C(\mathbf{b},\theta) = \sum_{i=1}^{M}(1 - h(b_i/\theta)),$$

where $h(x) = \max(1 - x^2, 0)$. According to this, we can write:

$$\begin{aligned} \hat{\mathbf{b}}(\theta) &= \arg\max_{\mathbf{b}} C'(\mathbf{b},\theta), \\ C'(\mathbf{b},\theta) &= \sum_{i=1}^{M} h(b_i/\theta) = M - C(\mathbf{b},\theta). \end{aligned}$$

It is easy to show that $C'(\mathbf{b},\theta) \propto C_\delta(\mathbf{b}) * H(\mathbf{b}/\theta)$, where $H(\mathbf{b}) = \prod_{i=1}^{M} h(b_i)$ is a smoothing kernel and $C_\delta(\mathbf{b}) = \sum_{i=1}^{M}\delta(b_i)$ is an infinitely sharp function [2]. Thus, the role of θ is to smooth a sharp cost function: the higher the threshold, the smoother the cost, and, thus, the easier is to find a favorable local optimum. Starting by the highest possible $\theta^{(0)}$, we obtain a reliable optimum, and then we obtain $\theta^{(k+1)}$ slightly decreasing $\theta^{(k)}$ (i.e., increase $\lambda^{(k)}$), finding new optima for each step until reaching the λ reference value. We call this method ℓ_0-GM (from *Gradual Minimization*). A simplified version of this consists of doing a single gradient descent step each time, slowly increasing $\lambda^{(k)}$ at each iteration. This idea is closely related to other schemes, such as GNC [14]. Alternatively [13] minimizes a continuous function which gets closer and closer to the ℓ_0-norm.

Figure 1 shows the convergence trajectories of IHT for several thresholds (dashed), and two trajectories (solid and circles) of the exponential decay of the threshold $\theta^{(k)} = \theta^{(0)}\beta^k$ for two different β values. We used $\alpha^{(k)} = 1.85(\theta^{(k)})^2/2$. Note that ℓ_0-GM not only greatly reduces the number of iterations w.r.t. IHT, but it also provides significantly higher fidelity at each sparseness level.

[2]The proportionality factor is easy to compute, but irrelevant here.

4. APPLICATION TO IMAGE RESTORATION

Consider that our observation, $\mathbf{y} \in \mathbb{R}^N$, has lost some localized information. As a starting point, we assume that we can replicate, from $\mathbf{y}$, the associated degradation, $f_{\mathbf{y}}(\mathbf{x})$. We define $R(\mathbf{y})$ as the set of images consistent with $\mathbf{y}$: $R(\mathbf{y}) = \{\mathbf{x} \in R^N : f_{\mathbf{y}}(\mathbf{x}) = \mathbf{y}\}$.

4.1. Analysis-sense sparseness-based restoration

Usual Bayesian approaches to image restoration build image priors reflecting the typical behavior of signals in many previous observations. But, whereas the responses to a signal of a set of linear kernels (analysis coefficients) are directly observable, the synthesis coefficients of an optimal sparse representation are not. Here we differentiate between synthesis-sense sparseness (SsS) and analysis-sense sparseness (AsS). The theoretical properties and the different uses we can give to both concepts are addressed in [15]. Although SsS methods are very powerful, AsS methods are easier to justify from an empirical Bayesian perspective, because they use statistical priors based on linear observations. Recently some authors have obtained very positive results using AsS methods in image processing (e.g. [8, 1]).

4.2. Estimation using ℓ_0-GM

Strict sparseness of analysis coefficients is, in general, not attainable. We consider instead that typical responses to natural images are compressible. Thus, we model the linear response as a strictly sparse vector representing the highest amplitude responses ($\mathbf{a}$), plus a small correction term ($\mathbf{r}$). We look for:

$$(\hat{\mathbf{a}}, \hat{\mathbf{r}}) = \arg\min_{\mathbf{a},\mathbf{r}} \{\|\mathbf{a}\|_0 + \lambda\|\mathbf{r}\|_2^2 \text{ s.t. } (\mathbf{a}+\mathbf{r}) \in S_A(\mathbf{y})\}. \quad (6)$$

Here $S_A(\mathbf{y})$ is the set of linear responses to images that are consistent with the observation, $S_A(\mathbf{y}) = \{\mathbf{b} \in R^M : \exists \mathbf{x} \in R(\mathbf{y}), \mathbf{\Phi}^T\mathbf{x} = \mathbf{b}\}$. Using $\mathbf{b} = \mathbf{a} + \mathbf{r}$, we follow a completely parallel path to the previous case and we end up with an expression depending only on $\mathbf{b}$, analogous to Eq. (4) but substituting the constraint set by $S_A(\mathbf{y})$. Note that now the constraint set is no longer affine, in general, and, thus, we have to consider its curvature by projecting the gradient onto the constraint tangent hyperplane at every boundary point $\mathbf{b}$:

$$\nabla^{S_A(\mathbf{y})} C(\mathbf{b}) = \lim_{\nu \to 0} \frac{P^{\perp}_{S_A(\mathbf{y})}(\nu \nabla C(\mathbf{b}))}{\nu},$$

where $P^{\perp}_{S_A(\mathbf{y})}$ is the orthogonal projector on $S_A(\mathbf{y})$. The gradient descent method is $\mathbf{b}^{(k+1)} = \mathbf{b}^{(k)} - \alpha \nabla^{S_A(\mathbf{y})} C(\mathbf{b})$. In practice, though, it is more convenient to use a simpler computation in the estimation loop:

$$\mathbf{b}^{(k+1)} = P^{\perp}_{S_A(\mathbf{y})} \left(\mathbf{b}^{(k)} - \alpha \nabla C(\mathbf{b}) \right),$$

which is equivalent to the previous one if the projection is linear. Given that $\mathbf{\Phi}^T$ is a Parseval frame we can write the previous projection in terms of the constraint projection in the image domain:

$$P^{\perp}_{S_A(\mathbf{y})}(\mathbf{b}) = \mathbf{\Phi}^T P^{\perp}_{R(\mathbf{y})} (\mathbf{\Phi}\mathbf{b}).$$

Because of the similar structure of Eqs. (2) and (6), the same results on global minimization strategies apply now. This means that significantly better restoration performance is achieved by using an exponentially decaying threshold until reaching the desired λ, than using the fixed threshold corresponding to that λ. We have experienced (as in [8, 1]) that the optimal final θ in our restoration applications is usually close to zero. Thus, an arbitrarily small value can be used in practice, avoiding, as a consequence, the λ selection problem.

5. APPLICATION TO *IN-PAINTING*

To estimate missing pixels is a common problem when dealing with digital images. Here we compare the performance of our method with two other ones: *EM-inpainting* [1], based on convex relaxation of the sparseness condition, and *Fast-inpainting* [16], a simple but effective PDE-based method. Given a set of indices, I, extracted from $\{1, ..., N\}$, and given $\mathbf{y} = f_{\mathbf{y}}(\mathbf{x})$, where all y_i with $i \in I$ holds that $y_i = x_i$, we define the consistency set, $R_I(\mathbf{y})$, as:

$$R_I(\mathbf{y}) = \{\mathbf{x} \in \mathbb{R}^N : \forall i \in I,\ x_i = y_i\}.$$

Given an image $\mathbf{x} \in \mathbb{R}^N$ and a $N \times N$ diagonal matrix $\mathbf{D}$, where each element d_{ii} is 1 if $i \in I$ and 0 otherwise, the orthogonal projection onto $R_I(\mathbf{y})$ results in $P^{\perp}_{R_I(\mathbf{y})}(\mathbf{x}) = \mathbf{D}\mathbf{y} + (\mathbf{I} - \mathbf{D})\mathbf{x}$, where $\mathbf{I}$ is the $N \times N$ identity matrix.

We have downloaded[3] our test images and the *EM-inpainting* results from *http://www.greyc.ensicaen.fr/∼jfadili*. For a fair comparison, we have forced the estimations to have the same observed pixel values in the three methods. In our implementation of *Fast-inpainting*, iterations end when the mean square difference of the estimation w.r.t. the previous one is less than 10^{-3}. Regarding ℓ_0-GM, we get a good compromise between performance and computational cost by using $\beta = 0.8$, $\alpha = \frac{\theta^2}{2}$. We stop the iterations when the threshold is below 0.1. To compare images, we use Peak Signal-to-Noise Ratio (PSNR), expressed as $10 \cdot log_{10}(255^2/MSE)$, where MSE is the Mean Square Error.

Top-left panel of Figure 2 is a simulated observation. Top-right panel corresponds to the result obtained using *Fast-inpainting* (32.71 dB), and bottom-left to *EM-inpainting* (34.14 dB). Last panel is the ℓ_0-GM result using 6-scales Candès' Curvelets (which is a Parseval frame) (34.92 dB). Note that ℓ_0-GM provides also the best visual performance. Figure 3 shows the case of a real degraded photo. Results have been arranged same as before. The downloaded image was deformed, so we have applied a 1.4 scale factor to the vertical axis. Here, ℓ_0-GM uses 6-scales Curvelets combined with LDCT. We can see again that ℓ_0-GM visually outperforms its competitors.

Times per iteration for *EM-inpainting* and ℓ_0-GM are very similar (≈ 3 s. for 256×256 images), but ours makes less iterations on average (40 vs. 100). *Fast-inpainting* takes around 0.5 s. per image.

6. CONCLUSIONS

It is possible to find equivalent formulations to the original sparse approximation problem for the tight frame case, allowing the use of standard optimization techniques, like gradient descent. This result is against the common belief that only convex relaxation or heuristical approaches make the sparse approximation problem tractable (in theory and practice). We have proposed an original formal derivation of a previous heuristical algorithm, based on iterative hard thresholding (IHT) with decreasing thresholds. Firstly, we have re-expressed the sparse approximation problem using a continuous cost function under an affine constraint. Then, we have done gradient descent on it, deriving so IHT, and we have demonstrated that its fixed point corresponds to a local minimum. Next, we have re-written the new function as an infinitely sharp cost function convolved with a smoothing kernel, which has allowed us to use a deterministic annealing-like approach to justify the use of decreasing thresholds. We have shown that an analogous derivation can be used for restoring images whose degradation function can be identified from the observation, and we have applied it to *in-painting* with great success.

[3] We thank J. Fadili for making available their source code and images.

Fig. 2. **Top-left**, 120×120 cropped degraded *Barbara* (24.19 dB). **Top-right**, *Fast-inpainting* (32.71 dB). **Bottom - left**, *EM-inpainting* (Curvelets + Local-DCT, 34.14 dB). **Bottom-right**, ℓ_0-GM (6-scales Curvelets, 34.92 dB).

Fig. 3. **Top-left**, 100×100 cropped *Girls* image. **Top-right**, *Fast-inpainting*. **Bottom-left**, *EM-inpainting* (Curvelets). **Bottom-right**, ℓ_0-GM (6-scales Curvelets + LDCT, block size 32×32).

7. REFERENCES

[1] M.J. Fadili, J.L. Starck, "EM algorithm for sparse representation-based image inpainting," in *IEEE Int. Conf. on Image Proc.*, Genoa, Italy, 11-14 Sep. 2005.

[2] J. Mairal, M. Elad, G. Sapiro, "Sparse representation for color image restoration," *IEEE Trans. on Image Proc.*, **18**, 1, 53-69, Jan. 2007.

[3] S. Mallat, Z. Zhang, "Matching pursuit in time-frequency dictionary," *IEEE Trans. on Signal Proc.*, **41**, 12, 3397-3415, Dec. 1993.

[4] Y. C. Pati, R. Rezaiifar, P.S. Krishnaprasad, "Orthogonal Matching Pursuit: Recursive function approximation with application to wavelet decomposition," in *27th Asilomar Conf. in Sig., Syst. and Comp.*, 1-3 Nov., 1993.

[5] S.S. Chen, D.L. Donoho, M.A. Saunders, "Atomic decomposition by basis pursuit," *SIAM*, **20**, 1, 33-61, 1999.

[6] B. Efron, T. Hastie, I. Jonhstone, R. Tibshirani, "Least Angle Regression," *Annual Stat.*, **32**, 2, 407-499", 2004.

[7] M.A.T. Figueiredo, R.D. Nowak, "An EM algorithm for wavelet-based image restoration," *IEEE Trans. on Image Proc.*, **12**, 8, 906-916, Aug. 2003.

[8] J.L. Starck, "Morphological component analysis," in *Proc. of the SPIE*, San Diego, CA, Aug. 2005.

[9] T.H. Reeves, N.G. Kingsbury, "Overcomplete image coding using iterative projection-based noise shaping," in *IEEE Int. Conf. on Image Proc.*, Rochester, NY, 23-25 Sep. 2002.

[10] T. Blumensath, M. Yaghoobi, M.E. Davies, "Iterative hard thresholding and L0 regularisation," in *IEEE Int. Conf. on Acoust., Speech and Sig. Proc.*, 15-20 Apr. 2007.

[11] M. Osborne, B. Presnell, B. Turlach, "A new approach to variable selection in least squares problems," *IMA J. of Numerical Anal.*, **20**, 389404, 2000.

[12] J. Portilla and L. Mancera, "L0-based sparse approximation: Two alternative methods and some applications," in *Proc. of the SPIE*, San Diego, CA, 26-30 Aug. 2007.

[13] G.H. Mohimani, M. Babaie-Zadeh, Christian Jutten, "Fast sparse representation based on smoothed L0 norm," in *Int. ICA Conf.*, London, UK, 9-12 Sep. 2007.

[14] A. Blake, A. Zisserman, "Graduated non-convexity," in *Visual Reconstruction*, MA MIT Press, Cambridge, Ed. 1987.

[15] M. Elad, P. Milanfar, and R. Rubinstein, "Analysis versus synthesis in signal priors," *Inverse Prob.*, **23**, 3, 947-978, Jun. 2007.

[16] M.M. Oliveira, B. Bowen, R. McKenna, Y.S. Chang, "Fast digital image inpainting," in *Int. Conf. on Vis., Imaging and Image Proc.*, Marbella, Spain, 3-5 Sep. 2001.