IMAGE RESTORATION THROUGH L0 ANALYSIS-BASED SPARSE OPTIMIZATION IN TIGHT FRAMES

Javier Portilla

Imaging and Vision Department
Instituto de Óptica
Consejo Superior de Investigaciones Científicas (CSIC), Madrid

ABSTRACT

Sparse optimization in overcomplete frames has been widely applied in recent years to ill-conditioned inverse problems. In particular, analysis-based sparse optimization consists of achieving a certain trade-off between fidelity to the observation and sparsity in a given linear representation, typically measured by some ℓ_p quasi-norm. Whereas most popular choice for p is 1 (convex optimization case), there is an increasing evidence on both the computational feasibility and higher performance potential of non-convex approaches ($0 \leq p < 1$). The extreme $p = 0$ case is especial, because analysis coefficients of typical images obtained using typical pyramidal frames are not strictly sparse, but rather *compressible*. Here we model the analysis coefficients as a strictly sparse vector plus a Gaussian correction term. This statistical formulation allows for an elegant iterated marginal optimization. We also show that it provides state-of-the-art performance, in a least-squares error sense, in standard deconvolution tests.

Index Terms— Image deconvolution, image restoration, optimization, regularization, wavelets, sparsity, tight frames.

1. INTRODUCTION

Sparse optimization in overcomplete frames has been widely applied in recent years to ill-conditioned inverse problems. Two related approaches have been followed [5]. The most common, that we term here synthesis-based sparsity (SbS) has been to search for sparse solutions in a generative sense, that is, solutions expressible as linear combination of few basis functions from a dictionary (e.g. [19]). The other approach, analysis-based sparsity (AbS), consists of searching for solutions whose linear responses in a given representation are *sparse* (e.g. [2]). These solutions are *not* equivalent, except for the especial case when the matrix is square and non-singular (like the Fourier transform, as e.g. [3]), or when the prior term is quadratic, for overcomplete frames [5]. Another criterion to classify sparse optimization methods is the function used to promote the sparsity, which typically is a certain ℓ_p quasi-norm. Most popular choice for p is 1, because it is the smallest value giving raise to a convex optimization. However, there is an increasing amount of evidence on both the computational feasibility (beyond inefficient classical greedy and pursuit approaches) and higher potential of different non-convex solutions using $0 \leq p < 1$ (e.g. [13, 10, 18, 1, 4, 15]).

Focusing on sparse-optimization-based deconvolution methods, most of them follow an ℓ_1 SbS approach (e.g. [6, 7]), and only a few are ℓ_0 SbS (e.g. [14]). Whereas AbS methods have been used for other image processing tasks (such as inpainting, superresolution or image morphological decomposition [2, 18, 15]), to the best of our knowledge, they have never been applied to image deconvolution. In addition, the ℓ_0 AbS case is, indeed, especial, in the sense that typical images, when represented with usual frames (such as *x*-lets, that is, overcomplete wavelet-like transforms), give raise to analysis coefficients that are not strictly sparse, but just approximately so. I.e., they are *compressible*. Thus, direct ℓ_0 minimization of *analysis* coefficients is not justified. Here we model the analysis coefficients as a strictly sparse vector plus a Gaussian correction term. We had previously used this model for inverting *a posteriori deterministic* degradations, such as missing pixels, bits, color components, etc. [18, 15]. When dealing with noisy blurred images, in contrast, the involved densities are not degenerated, but otherwise the general approach is the same. In addition, we recently presented an ℓ_0 SbS precursor of the method described here [16]. As shown in the results, the new method's performance is state-of-the-art, at least when applied to images with a moderate amount of texture. The code used for the results shown here is publicly available in [17].

2. IMAGE MODELING AND PROBLEM FORMULATION

We assume here that our observation, $\mathbf{y}$, comes from convolving an N-pixel image, $\mathbf{x}$, with a known PSF kernel, $\mathbf{h}$, and adding zero-mean white Gaussian noise, $\mathbf{w}$, of known variance σ^2:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}, \tag{1}$$

where $\mathbf{H}$ is the corresponding block-circulant matrix associated to h. Therefore, our likelihood term is $p(\mathbf{y}|\mathbf{x}) \propto \exp(-\frac{1}{2\sigma^2}\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2)$.

For the prior modelling we consider a given over-complete $M \times N$ frame Φ^T, typically a multi-scale, multi-orientation, x-let representation, for which typical images produce *compressible* coefficient vectors, i.e., those that can be approximated using a highly sparse vector $\mathbf{a}$ plus a residual:

$$\Phi^T\mathbf{x} = \mathbf{a} + \mathbf{r}, \tag{2}$$

where $\mathbf{r}$ is the residual term. We model $\mathbf{r}$ as white and Gaussian, $p(\mathbf{r}) \propto \exp(-\frac{1}{2\sigma_r^2}\|\mathbf{r}\|_2^2)$, and $\mathbf{a}$ using the improper prior $p(\mathbf{a}) \propto \exp(-\frac{1}{\alpha}\|\mathbf{a}\|_0)$. Although $\mathbf{a}$ and $\mathbf{r}$ are actually not independent variables, as they live in a N-dimensional hyperplane in R^M, here we consider them as such, for simplicity sake. Summarizing our statistical modelling, the likelihood function and the two prior densities

This research has been supported by grant TEC2006-13485/TCM, from Spanish Government.

(sparse component and residual) are:

$$\begin{aligned} p(\mathbf{y}|\mathbf{x}) &\propto \exp(-\frac{1}{2\sigma^2}\|\mathbf{y}-\mathbf{H}\mathbf{x}\|_2^2) \\ p(\mathbf{x}|\mathbf{a}) &\propto \exp(-\frac{1}{2\sigma_r^2}\|\Phi^T\mathbf{x}-\mathbf{a}\|_2^2) \\ p(\mathbf{a}) &\propto \exp(-\frac{1}{\alpha}\|\mathbf{a}\|_0). \end{aligned} \quad (3)$$

Note that the resulting prior $p(\mathbf{x}) = \int_{\mathbf{a}} p(\mathbf{x}|\mathbf{a})p(\mathbf{a})d\mathbf{a}$ is the convolution of an ℓ_0 prior with a Gaussian. By noting that $p(\mathbf{y}|\mathbf{x},\mathbf{a}) = p(\mathbf{y}|\mathbf{x})$, we can write $p(\mathbf{x},\mathbf{a},\mathbf{y}) = p(\mathbf{y}|\mathbf{x})p(\mathbf{x}|\mathbf{a})p(\mathbf{a})$, whose maximization in the unknown vectors $(\mathbf{x},\mathbf{a})$ is the same as finding:

$$(\hat{\mathbf{x}},\hat{\mathbf{a}}) = \arg\min_{\mathbf{x},\mathbf{a}} \|\mathbf{a}\|_0 + \lambda\|\Phi^T\mathbf{x}-\mathbf{a}\|_2^2 + \mu\|\mathbf{y}-\mathbf{H}\mathbf{x}\|_2^2 \quad (4)$$

where $\lambda = \frac{\alpha}{2\sigma_r^2}$ and $\mu = \frac{\alpha}{2\sigma^2}$. Note that this solution is *not* equivalent to the MAP solution (maximization of $p(\mathbf{x}|\mathbf{y})$), which would be difficult to compute, as it involves integrating on vector $\mathbf{a}$.

3. ITERATED MARGINAL MINIMIZATION

We have attacked Eq.(4) by iteratively marginally maximizing $p(\mathbf{x},\mathbf{a},\mathbf{y})$ in $\mathbf{a}$ and x (which is equivalent to maximizing the conditional densities $p(\mathbf{x}|\mathbf{a},\mathbf{y})$ and $p(\mathbf{a}|\mathbf{x},\mathbf{y})$). At convergence (which we do not prove here) we can only claim that a joint local maximum in $(\mathbf{x},\mathbf{a})$ of $p(\mathbf{x},\mathbf{a},\mathbf{y})$ has been reached, even if, as in this case, each optimization step achieves the (marginal) global optimum. For a given $\mathbf{a}$ the minimization in $\mathbf{x}$ of the criterion in Eq.(4) is quadratic, and, thus, easy to solve. For a given $\mathbf{x}$ the minimization in $\mathbf{a}$ is also easy, because it is separable in each component a_i, resulting in a single hard thresholding of the coefficient vector $\Phi^T\mathbf{x}$. These operations result in the following estimation loop:

$$\begin{aligned} Step\,0: \quad \mathbf{x}^{(0)} &= \mathbf{y} \\ Step\,1: \quad \mathbf{b}^{(n)} &= \Phi^T\mathbf{x}^{(n)} \\ Step\,2: \quad \mathbf{a}^{(n)} &= \Theta_{hard}(\mathbf{b}^{(n)},\lambda^{-1/2}) \\ Step\,3: \quad \mathbf{z}^{(n)} &= \Phi\mathbf{a}^{(n)} \\ Step\,4: \quad \mathbf{x}^{(n+1)} &= (\Phi\Phi^T+\nu\mathbf{H}^*\mathbf{H})^{-1}(\mathbf{z}^{(n)}+\nu\mathbf{H}^*\mathbf{y}) \\ back\,to\,Step\,1 \end{aligned}$$

Step 1 is just applying the analysis frame (e.g., an over-complete wavelet-like transform) to the current image estimation. Step 2 is the hard thresholding operation resulting from marginally minimizing in $\mathbf{a}$ the criterion of Eq.(4). If Φ^T is a self invertible transform (a Parseval frame), then Step 3 is simply applying the reconstruction operation to the thresholded vector coefficient. Finally, Step 4 is the result of marginally minimizing in $\mathbf{x}$ the criterion of Eq.(4), where $\nu = \mu/\lambda = \sigma_r^2/\sigma^2$. Note first that, again, if our representation is a Parseval frame then $\Phi\Phi^T = \mathbf{I}$, which greatly simplifies the computations. Now Step 4 can easily be done in the Fourier domain:

$$X(u,v) = \frac{Z(u,v)+\nu H^*(u,v)Y(u,v)}{1+\nu|H(u,v)|^2},$$

with $Z(u,v) = \mathcal{F}\left\{z^{(n)}\right\}$ and $\mathbf{x}^{(n+1)} = \mathcal{F}^{-1}\{X(u,v)\}$. Note that both $H^*(u,v)Y(u,v)$ and $|H(u,v)|^2$ are computed only once, outside of the loop. Thus, if our representation is a Parseval frame and we have a fast implementation of the analysis and reconstruction operations (Steps 1 and 3), then the computational cost per iteration of the algorithm is low and, typically, $O(N\log N)$.

4. IMPLEMENTATION AND TESTS

4.1. Image Representation

We have considered as in some previous works [19, 18] a combined Parseval frame, simply obtained by concatenating the corresponding vectors of two individual Parseval frames, each divided by $\sqrt{2}$ to preserve the Euclidean norm. In this case we have used the 2-D Dual Tree Complex Wavelet Transform (DT-CWT [12]) and the Translation Invariant Haar Pyramid (TIHP [11]), both with 7 dyadic scales. This combination seems to work particularly well for typical (not too texture-rich) images. We have obtained times per iteration of 0.13, 0.52, and 0.64 seconds using 256×256 images, for our combined Matlab(R)+C implementation in a Pentium(R) 4 CPU 3.20 Ghz, using the TIHP alone, the DT-CWT alone, or the combination of both, respectively. As shown in Figure 1, very high quality results can still be obtained in favorable cases by using just the TIHP with few iterations, which translates into a deconvolution time really short (around 1 second in our platform). For the results shown here (50 iterations, combined representation) running time was 32 s.

4.2. Experiments

For this preliminary work we have used just the standard 256×256 images *House* and *Cameraman*. We have applied to each of these two images a total of five different combinations previously used in the restoration literature of 4 different point spread functions (PSF) and different amounts of white Gaussian additive noise (see Table 1 in next section). *PSF1* is $h_{i,j} = (1+i^2+j^2)^{-1}$, for $i,j = -7...7$. *PSF2* is a 9×9 uniform kernel. *PSF3* is a 5×5 separable kernel made as the outer product of [1,4,6,4,1]/16 in x and y. *PSF4* is Gaussian with spatial 2^{nd} order moment $\sigma_b^2 = \sigma_x^2+\sigma_y^2$, with $\sigma_x = \sigma_y = 1.6$.

4.3. Parameters

In future work we would like to estimate the model parameters α and σ_r either from a training set of typical images, or directly from the corrupted observations. For this work, however, we have just tried to evaluate how robust the method behaves for two different images and a wide range of image degradations. For that purpose we have, first, hand-optimized the two parameters for using them with the five degradations and the two images (10 experiments in total), obtaining $\alpha = 4.8$ and $\sigma_r = 7.22$. For comparison purposes, we have also optimized the two parameters for each individual experiment, as discussed in next section.

5. RESULTS AND DISCUSSION

Table 1 shows our results (double row labelled "L0-AbS", from ℓ_0 Analysis-based Sparsity), above by using the same model parameters for all the experiments, and below (bracketed), by hand-optimizing the two model parameters for each individual experiment. The results are compared to what to the best of our knowledge are the deblurring state-of-the-art methods, in terms of their proven performance with standard tests: Figueiredo&Nowak 2005 [8], Foi *et al.* 2006 [9], and Guerrero *et al.* 2008 [11]. The new proposed method clearly outperforms both [11] and [8], and it provides similar average performance as [9] (half of the best scores, highlighted, correspond to [9], and the other half to our method).

Figure 2 shows a visual comparison of our method to [9] for using Cameraman with *PSF2* and $\sigma^2 = 0.308$ (40 dB BSNR). Whereas both results recover well the original image, it is quite noticeable how our result suffers from less artifacts (see, for instance,

ringing around the camera handle in [9]). This translates into more than half a decibel SNR improvement. In Figure 3 we compare to [8] for the last experiment (with *PSF4*, a Gaussian kernel). Note how our method recovers the original sharpness of some edges (e.g., vertical edges in door and window, oblique edge at the bottom of the roof) that in [8] are still quite blurry.

On the other hand, bracketed (*oracle*) results shown in Table 1 demonstrate how, at least for the set of experiments considered, the model and method presented here are quite robust, so as to provide a performance close to these upper bounds, by using the same model parameters for different images and degradations. However, this does not mean that top performance would be obtained with all kind of images and degradations without adapting the image representation and/or the model parameters to the observation. For instance, we have experienced in preliminary tests, a significant performance drop when using this set of parameters applied to very texture-rich images, such as *Barbara*, or when the blurring is negligible compared to the amount of added noise. We believe this is due, in the first case, to the relative inefficiency of the used representation to deal with heavily textured images (e.g., local DCT could do better for those cases). In the second case, when noise clearly dominates to blur, better results are obtained by increasing the prior's sparsity, through decreasing both α and σ_r^2.

6. CONCLUSIONS

We have presented a new deconvolution method based on ℓ_0 analysis-based sparsity. Though it makes some unrealistic assumptions (such as the independence of the sparse and the residual terms), it is fully model-based, statistically motivated and conceptually simple. In addition, it translates into an efficient algorithm. Its performance through ten standard simulation experiments is either state-of-the-art (in half of the studied cases) or highly competitive (for the rest). We find remarkable how such a simple model and its derived method, with no ad-hoc manipulations, and with just two free parameters, compares favorably to other much more sophisticated models. It is also noteworthy how our applied iterative marginal optimization leads by itself to a good and reasonably fast convergence despite the strong non-convexity of the associated cost function, without requiring any heuristical optimization strategy (like dynamically decreasing the threshold [15], greedy pursuits [14], etc.) This combination of simplicity and effectiveness is rare in the context of difficult inverse problems.

We have also experienced a high robustness of the method's performance with the model parameters. However, we have not solved yet for a general procedure to choose the free parameters individually for each observation. In general, regarding the possibility of model parameter estimation, AbS has the conceptual and practical advantage of the representation coefficients being observable, and, as such, they are susceptible of direct statistical modelling, unlike what happens in the SbS case, where we can only indirectly infer, if at all, the statistics of the synthesis coefficients.

7. REFERENCES

[1] 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.

[2] J. Bobin, J.L. Starck, J. Fadili, Y. Moudden, and D.L. Donoho, "Morphological Component Analysis: an adaptive threshold-

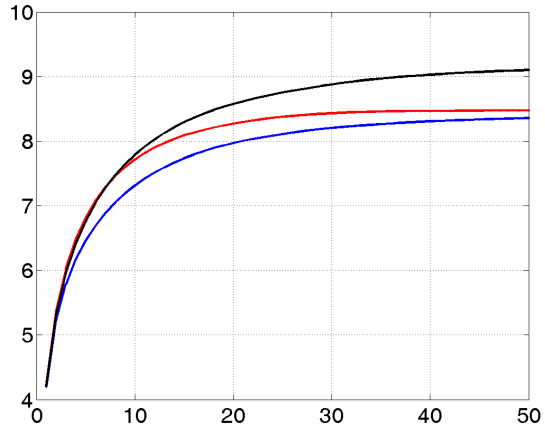

Fig. 1. Increment in signal-to-noise ratio (ISNR) as a function of the number of iterations, obtained with *Cameraman* blurred by *PSF2*, and noise variance $\sigma^2 = 0.308$. Curves correspond, from down to top, to using the image representations DT-CWT, TIHP, and their combination, respectively.

Blur	*PSF 1*		*PSF 2*	*PSF 3*	*PSF 4*
$\sigma \rightarrow$	$\sqrt{2}$	$\sqrt{8}$	$\sqrt{0.308}$	7	2
Method	*HOUSE*				
[8]	8.47	6.63	10.71	4.22	4.49
[9]	**9.05**	**7.64**	10.71	**5.10**	4.03
[11]	8.64	7.03	9.04	4.30	4.11
L0-AbS	8.40	7.12	**10.74**	4.55	**4.80**
oracle $\rightarrow$	(8.80)	(7.13)	(11.12)	(4.57)	(4.85)
Method	*CAMERAMAN*				
[8]	7.46	5.24	8.16	2.84	3.18
[9]	**8.25**	**6.34**	8.57	2.56	3.05
[11]	7.45	5.55	7.33	2.73	3.25
L0-AbS	7.70	5.55	**9.10**	**2.94**	**3.49**
oracle $\rightarrow$	(7.71)	(5.88)	(9.11)	(3.00)	(3.54)

Table 1. Performance comparison with previous state-of-the-art methods in terms of Increment of Signal-to-Noise ratio (ISNR), in dBs. First row shows the blurring kernel used (see text for details), and the second denotes noise standard deviation. Best results are highlighted. Bracketed results have been obtained by choosing α and σ_r individually for each experiment (see text for details).

Fig. 2. From left to right and from top to bottom: (a) Crop from *Cameraman*; (b) Simulated observation using a 9×9 uniform kernel, with noise variance 0.308 (3rd. experiment); (c) Previous best result, from [9] (8.57 dB ISNR); (d) Our result (9.10 dB).

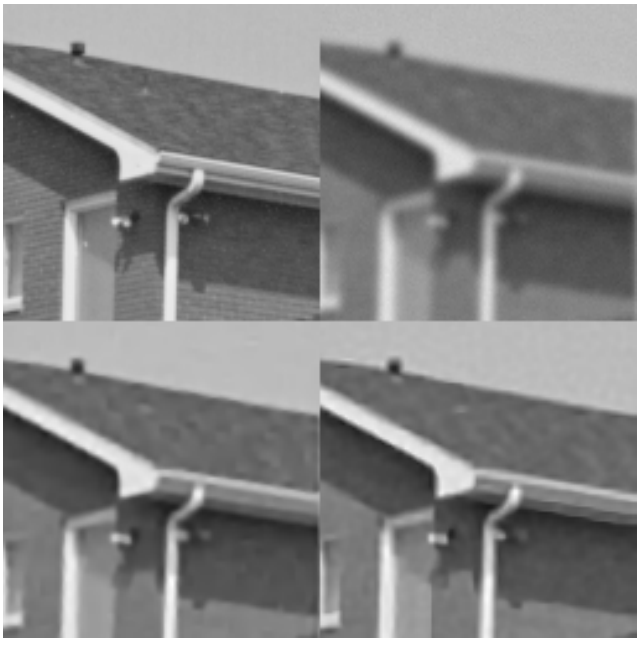

Fig. 3. From left to right and from top to bottom: (a) crop from *House* test image; (b) Simulated observation using a Gaussian kernel, with spatial 2nd order moment 5.12, and noise variance 4 (5th. experiment); (c) Previous best result, using [8] (4.49 dB ISNR); (d) Our result (4.80 dB).

ing strategy," *IEEE Trans. on Image Proc.*, **16**(11), 2675–2681, Nov. 2007.

[3] E. J. Candes, J. Romberg, T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, **52**(2), 489-509, Feb. 2006.

[4] R. Chartrand, "Exact reconstructions of sparse signals via nonconvex minimization," in *IEEE Signal Process. Lett.*, **14**(10), 707-710, Oct. 2007.

[5] M. Elad, P. Milanfar, R. Rubinstein, "Analysis versus synthesis in signal priors," *Inverse Problems*, **23**, p. 947, 2007.

[6] M. J. Fadili and J.-L. Starck, "Sparse representation-based image deconvolution by iterative thresholding," in *ADA IV, Elsevier*, France, 2006.

[7] M. A. Figueiredo, J. M. Bioucas-Dias, R. D. Nowak, "Majorization-Minimization Algorithms for Wavelet-Based Image Restoration," *IEEE Trans. on Image Proc.*, **16**(12), 2980-2991, Dec. 2007.

[8] M Figueiredo and R Nowak, "A bound optimization approach to wavelet-based image deconvolution," in *IEEE Int'l Conf. on Image Proc.*, **2**, 782–785, Oct. 2005.

[9] A Foi, K Dabov, V Katkovnik, and K Egiazarian, "Shape-adaptive dct for denoising and image reconstruction," in *SPIE Electronic Imaging 2006, Image Processing: Algorithms and Systems V*, pp. 6064A–18, 2006.

[10] O.G. Guleryuz, "Nonlinear approximation based image recovery using adaptive sparse reconstructions and iterated denoising-part I: Theory.," *IEEE Trans. on Image Proc.*, **15**(3), p. 539, Mar. 2006.

[11] J. A. Guerrero-Colón, L. Mancera, and J. Portilla, "Image restoration using space-variant Gaussian scale mixture in overcomplete pyramids," *IEEE Trans. Image Proc.*, **17**(1), 27 – 41, Jan. 2008.

[12] N. Kingsbury, "Complex Wavelets for shift invariant Analysis and Filtering of Signals," *Journal of Applied and Computational Harmonic Analysis*, **10**(3), 234-253, May 2001.

[13] N. Kingsbury and T. Reeves, "Redundant representation with Complex Wavelets: How to achieve sparsity," in *Int'l Conf. on Image Proc.*, (Barcelona, Spain), 14-18 Sep. 2003.

[14] J. Mairal, M. Elad, G. Sapiro, "Sparse representation for color image restoration," *IEEE Trans. on Image Proc.*, **18**(1), 53-69, Jan. 2008.

[15] L. Mancera and J. Portilla, "Non-convex sparse optimization through deterministic anneling and applications, " in *IEEE Intl. Conf. on Image Proc.*, San Diego, CA, Oct. 2008.

[16] J. Portilla and L. Mancera, "Dynamic iterated hard-thresholding for non-convex sparse optimisation," in *Workshop on Sparsity and its application to large inverse problems*, Cambridge, UK, 14-15 Dec. 2008.

[17] J. Portilla, "L0-AbS: Implementation and Available Code," in *http://www.io.csic.es/PagsPers/JPortilla/index.html*, available after publication of this work, 2009.

[18] 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.

[19] J.-L. Starck, M. Elad, D. Donoho, "Image decomposition via the combination of sparse representations and a variational approach," *IEEE Trans. Image Proc.*, **14**(10), 1570–1582, 2005.