Sparse Approximation via Orthogonal Projections: Beyond Greed and Convexity

Luis Mancera*†, Jose A. Guerrero-Colon† and Javier Portilla♠

Abstract

We present a simple and robust optimization method for, giving an overcomplete Parseval frame, an ℓ_p-norm value, and an image, obtaining the representation vector, up to that ℓ_p-norm, that minimizes the mean square error (MSE) of the image reconstruction. It is based on alternated orthogonal projections onto the ℓ_p-ball of given radius and the set of representation vectors providing perfect reconstruction of the image. This method, which we call ℓ_p-AP, yields the global optimum when $p \geq 1$, and a local optimum when $0 \leq p < 1$. We focus on the $p = 0$ and $p = 1$ cases, and apply them to the sparse approximation problem. We show that, even being suboptimal, ℓ_0-AP clearly outperforms ℓ_1-AP (which is equivalent to *Basis Pursuit*). This result strongly questions the equivalence of minimizing both norms when using natural images, typical dictionaries and useful sparseness levels. We also show that ℓ_0-AP outperforms greedy heuristics and iterative methods based on applying fixed thresholds. Finally, we show an example of application of ℓ_0-AP to the removal of spatial quantization artifacts.

Index Terms

Sparse approximation, overcomplete dictionaries, alternated orthogonal projections, hard-thresholding, image de-quantizing

I. INTRODUCTION

The structure of the objects in the world produces images with distinct statistical features. It is fundamental for image processing to have a good *a priori* description of natural images (e.g., [1], [2], [3], [4]). It has long been observed that the power of this description is considerably increased by representing the pixels into a new domain, where we express the image as a linear combination of basis functions from a given dictionary [5]. Overcomplete dictionaries (those with a greater number of basis functions than the number of pixels) allow for a more powerful image analysis and processing compared to critically sampled ones, because they favor the extraction of relevant local features without breaking the invariance to translation, rotation, phase, etc. [6], [7], [3].

Using an overcomplete representation, there are infinite ways of expressing an image with a given dictionary. Traditionally, the minimum energy solution has been chosen, because it is linear and thus, easy to compute, though it has the disadvantage of introducing a lot of statistical dependency among coefficients. This makes it inefficient to store and process. However, there are non-linear solutions for this problem which result into a much smaller statistical dependency among coefficients. This also has the effect of improving the performance of image processing tasks based on simple models assuming independent coefficients. And they have additional advantages in terms of compression and computational efficiency at the reconstruction stage.

Typically, overcomplete linear representations, such as multi-scale and multi-orientation pyramids (*x*-lets), provide *sparse* solutions, meaning that a big proportion of the total energy of the signal is concentrated in a small proportion of the coefficients [5], [8], [9], [10]. However, signal transformations decreasing the statistical dependency among coefficients also significantly increase the sparseness of the representations, when the involved signal are leptokurtotic [11]. This is also the case for some non-linear solutions to the representation problem. Furthermore, if we allow for a certain degree of error in the image reconstruction, those representations can be made strictly sparse, i.e., they can use a small proportion of non-zero coefficients, without this new constraint causing a serious damage to the image reconstruction. Some authors have formulated the problem of finding such representations as minimizing the number of basis functions used, up to some reconstruction error. This is called the *sparse approximation problem*. Unfortunately, this is a NP-hard problem, whose only known optimal solution is combinatorial. However, there exist suboptimal solutions with tractable complexity. We have classified them in three main groups: greedy heuristics, convex relaxation methods and iterative thresholding methods.

Greedy heuristics started with *Matching Pursuit* (MP) [9]. MP accumulatively selects at each step the

basis function that best represents the approximation error. As a refinement, *Orthogonal Matching Pursuit* (OMP) [12] optimizes the selected coefficients at each step to minimize the mean square error (MSE) of the reconstruction. Although MP and OMP have been used for several applications, such as denoising (e.g., [13]) or video coding (e.g. [14], [15], [16]), their strict implementation is very slow, because they select a single basis function at a time. Stagewise OMP (StOMP) [17] allows for selecting several basis functions at a time. Some heuristics to improve the OMP performance have been developed [18], [19], [20], but they are also very computationally demanding. In [21], we used a simple method based on non-iteratively selecting the largest amplitude coefficients at one stage, and then iteratively optimizing the associated coefficients to minimize the mean square reconstruction error.

The number of non-zero coefficients of a vector is measured by its ℓ_0-norm, a strongly non-convex function. It has long been proposed to use instead more tractable norms [22]. The ℓ_1-norm, as the lowest convex norm, has been widely used as a substitute for the ℓ_0 norm, giving raise to the convex relaxation problem [23]. Some techniques [24], [25], [26], [27] for solving this problem have become a reference in the signal and image processing field. Among them, the most commonly used is *Basis Pursuit* (BP) [25]. However, when applied to images (i) there are no robust and efficient implementations of these methods; and (ii), as we demonstrate here, their sparse approximation results are generally poor. The ℓ_1-AP method presented here (though inferior to ℓ_0-AP, also presented here) provides equivalent results to Basis Pursuit, but in a more easily applicable way when using common tight frames for image representation.

The use of convex relaxation methods for sparse approximation is theoretically supported on the equivalence conditions of the ℓ_1 and the ℓ_0-norm minimizations. However, sufficient conditions given in [28] require the proportion of non-zero coefficients to be extremely small, having thus little practical application [29]. More recently, the equivalency condition has been found to be a number of non-zeros proportional to N [29]. In [27] equivalency conditions in the presence of noise are stated. However, the proportionality factor in [29] is difficult to calculate in general, and the conditions in [27] are highly restrictive. Here, we prove through a large set of careful experiments that equivalency does not hold when using typical images and representations in useful sparseness levels. In fact, this has been suggested by some authors besides us, who obtained better practical results using hard instead of soft thresholding in several sparse-based applications (see, e.g., [30], [31]).

The minimization of the ℓ_1-norm has also been applied to signal and image restoration (e.g., [25], [27] for denoising). Note that, interpreted as a maximum a posteriori (MAP) estimation, it is in better agreement with empirical image statistics to use a smaller norm $p < 1$ (e.g., [1], [32]). But the problem of addressing in a classical Bayesian way the estimation is deeper here: using any prior model based

on observed linear responses (*analysis* coefficients) to build a prior on *synthesis* coefficients, although widespread, is conceptually inconsistent. In this respect, we consider more correct to use the qualitative assumption that typical images can be represented with little error using a small proportion of active synthesis coefficients than to use a quantitative statistical model in the linear representation space and then apply it to the non-linear synthesis coefficients.

There is also a group of heuristic algorithms which reduce the ℓ_0-norm by iterative hard-thresholding the vector representation and imposing perfect reconstruction [33], [34], [35], [36], [37], [38]. Although they do not formulate any criterion to optimize, some of them provide excellent results in practice [35], [38], establishing the state-of-the-art in terms of signal energy compaction. In the case of ℓ_1 minimization using iterative soft-thresholding there is a theoretically grounded method [39]. Although the referred work provides a very powerful frame for many applications, it provides poor results when applied to the sparse approximation problem. Heuristics based on iterative soft-thresholding have also been developed (e.g., [40]), with good results for many image processing applications. However, we have experienced that the same heuristics provide significantly better performance when using instead iterative hard-thresholding (which would correspond to a ℓ_0 instead of a ℓ_1-norm minimization).

Here we present a robust and simple optimization method that minimizes the MSE of the reconstruction from a representation vector (using a Parseval frame) up to some ℓ_p-norm, which we call the ℓ_p-AP method. We systematically test the sparse approximation performance of this method compared to previous approaches, using a set of five natural images and two widely used image representations. We also show an example of application of ℓ_0-AP to the removal of spatial quantization artifacts. A previous instantiation of this work was published in [31].

The structure of this paper is as follows. Section II first motivates the need of a non-linear mechanism for improving the sparseness of the image representation, and then it formally introduces the sparse approximation problem. Section III describes the ℓ_p-AP method and explains in detail the $p = 0$ and $p = 1$ cases. We also explore here briefly the extension to image restoration. Section IV describes some implementation details necessary to replicate the experiments presented in Section V, where results are discussed. Last section concludes the paper

II. SPARSE APPROXIMATION

Figure 1 shows the approximation quality obtained in the image reconstruction using a limited number of non-zero coefficients, for several representations. This is expressed as Peak Signal-to-Noise Ratio (PSNR), defined as $10 \cdot log_{10}(\frac{255^2}{MSE})$, where MSE is the average mean square error for five standard
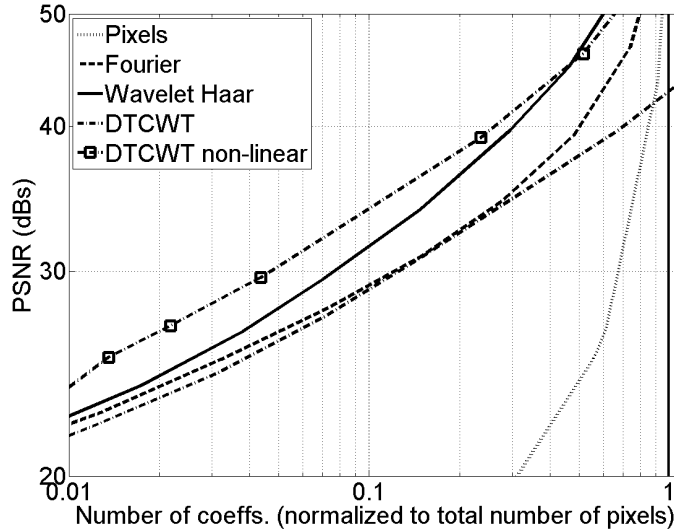
Fig. 1. Comparison of sparse approximation results obtained for several image representations. See text for details.

256×256 images[1]. Three of the representations used are critically sampled (pixels, Fourier, Haar wavelet), whereas the fourth is a redundant frame (the Dual Tree Complex Wavelet, DT-CWT [41]). For all these linear representations we take the K largest coefficients in amplitude. We can see how the quality of the reconstruction for a given number of basis functions is improved when we go from pixels to Fourier and from Fourier to critically sampled wavelets. However, when we do the same direct coefficient selection on an overcomplete representation, performance drops. This is due to multiple coefficients responding to the same image features. In order to increase the sparse approximation performance with respect to critically sampled wavelets, it is clear that we need a non-linear selection mechanism. We have also plotted here the results obtained with the ℓ_0-AP method proposed in this paper (DT-CWT, non-linear) (see Section III), showing a great improvement over the linear representations. Figure 2 illustrates the effect of using such a non-linear selection mechanism: coefficients of a non-linear solution to the representation problem are more sparsely distributed (central panel vs. left panel), thus avoiding simultaneous responses and representing more efficiently the local features of the image. Such a non-linear representation makes possible to keep a small proportion of the total number of coefficients (less than 4% in this case), whereas keeping the reconstruction quality high (35.7 dB, in this example, see right panel).

We formally present next the sparse approximation problem. Let $\boldsymbol{\Phi}$ be a $N \times M$ matrix with $M > N$

---

[1](*House, Boat, Barbara, Peppers, Lena*) *Boat* and *Barbara* are cropped to 256×256 starting at row 200, column 100 for *Boat* and row 150, column 50 for *Barbara*.
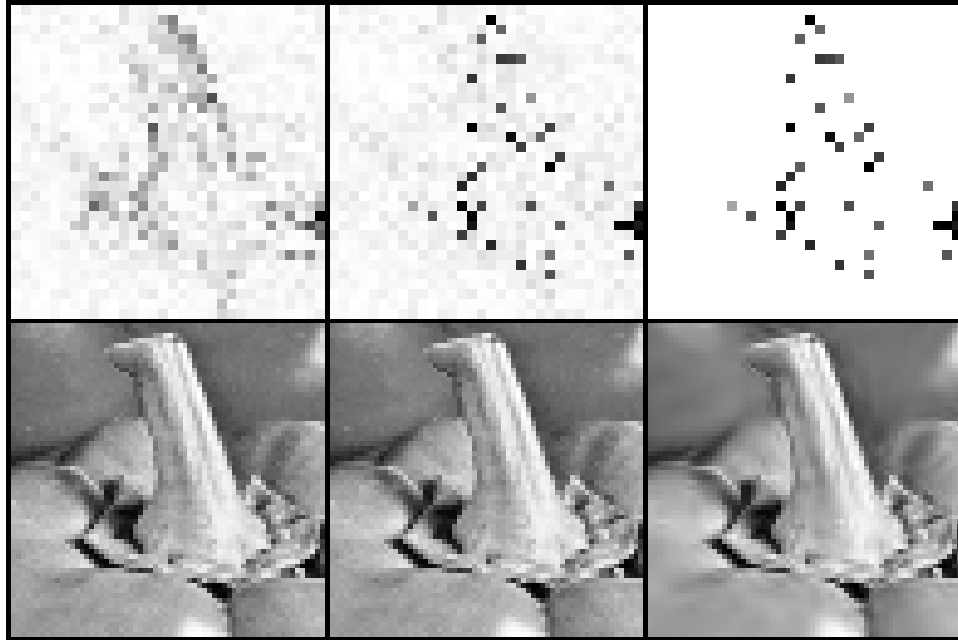
Fig. 2. **Top-left** Higher frequency scale of the analysis pyramid of *Peppers* image using 8-scales DT-CWT. Light and dark pixels indicate respectively small and large amplitude of the coefficients. Size of the subband has been duplicated by pixel replication to match the image size and then cropped to 64×64 starting at row 111, column 91 for visibility. **Top-center and top-right** Same subband from the non-linear solutions provided by minimizing the approximation error with our method using $0.143N$ coefficients. Center image is the representation obtained, and right image is the sparse approximation. **Bottom-left**, original image. **Bottom-center** Perfect reconstruction from the non-linear representation obtained with our method **Bottom-right** Reconstruction from the approximation result of our method (35.67 dB).

and $rank(\mathbf{\Phi}) = N$, representing the synthesis operator of an overcomplete Parseval frame. N is the number of pixels of the image and M the number of coefficients of the representation vector. Then, for an observed image $\mathbf{x} \in \mathbb{R}^N$, the system:

$$\mathbf{\Phi}\mathbf{a} = \mathbf{x} \tag{1}$$

has infinite solutions in $\mathbf{a} \in \mathbb{R}^M$. The minimum ℓ_2-norm solution, $\mathbf{a}^{LS} = \mathbf{\Phi}^T\mathbf{x}$, is often chosen because it is linear and, thus, simple to compute. Here, $\mathbf{\Phi}^T$ is the analysis operator of the Parseval frame. We aim to solve the sparse approximation problem, which is formulated as follows:

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

where $\|\mathbf{a}\|_0$ means the ℓ_0-norm of $\mathbf{a}$, i.e., the number of its non-zero elements, and $\lambda \in \mathbb{R}^*$ controls the accuracy of the approximation (the larger is λ, the more accurate the solution will be, at the price of reducing the sparseness). Unfortunately, as we illustrated in Figure 1, direct thresholding of $\mathbf{a}^{LS}$ does

not provide any result close to satisfy Eq. (2) for any λ. Thus, more sophisticated non-linear mechanisms than a single thresholding are required.

III. ENFORCING SPARSENESS THROUGH ALTERNATE PROJECTIONS

A. The ℓ_p-AP method

We can generalize Eq. (2) by:

$$\hat{\mathbf{a}}_p(\lambda) = \arg\min_{\mathbf{a}}\{\|\mathbf{a}\|_p + \lambda\|\mathbf{\Phi}\mathbf{a} - \mathbf{x}\|_2^2\}, \tag{3}$$

where $\|\mathbf{a}\|_p = (\sum_{i=1}^{M}|a_i|^p)^{\frac{1}{p}}$ denotes the ℓ_p-norm of $\mathbf{a}$ for some norm $p > 0$. Note that, for a fixed p and for each λ value, $\hat{\mathbf{a}}_p(\lambda)$ will have some ℓ_p-norm, R. Then, to solve Eq. (3) for a given λ is equivalent to minimize the approximation error for a given $\|\hat{\mathbf{a}}_p(\lambda)\|_p = R$. Being the ℓ_p-ball $B_p(R) = \{\mathbf{a} \in \mathbb{R}^M : \|\mathbf{a}\|_p \leq R\}$:

$$\hat{\mathbf{a}}_p(\lambda) = \hat{\mathbf{a}}_p(R) = \arg\min_{\mathbf{a} \in B_p(R)} \|\mathbf{\Phi}\mathbf{a} - \mathbf{x}\|_2. \tag{4}$$

Because of its simplicity, we have chosen to solve this problem via alternating orthogonal projections [42] onto two sets until convergence is reached. The first involved set is the set of representation vectors (synthesis coefficients) which are solutions to Eq. (1), defined as $S(\mathbf{\Phi}, \mathbf{x}) = \{\mathbf{a} \in \mathbb{R}^M : \mathbf{\Phi}\mathbf{a} = \mathbf{x}\}$. It is an affine subspace, and thus, it is convex. The second set is the ℓ_p-ball $B_p(R)$, for given p and radius R. This is convex only if $p \geq 1$. When both sets are convex, the projections converge on $B_p(R)$ to the global minimum of the distance to $S(\mathbf{\Phi}, \mathbf{x})$ in that ball. Otherwise, the procedure converges to a local minimum. See [43] for a more complete discussion on the convergence properties when non-convex sets are used. Here we assume that the starting vector for the iterations has an ℓ_p-norm larger than desired (as it happens in practice), thus ensuring that the achieved solution will lie on the boundary of the ℓ_p-ball and that, as a consequence, Eq. (4) will hold.

We denote $P_C^{\perp}(\mathbf{v})$ the orthogonal projection of the vector $\mathbf{v}$ onto a given set C. Then, projecting $\mathbf{a}$ onto the affine space $S(\mathbf{\Phi}, \mathbf{x})$ of perfect reconstruction yields the well-known result:

$$P_{S(\mathbf{\Phi},\mathbf{x})}^{\perp}(\mathbf{a}) = \mathbf{a} + \mathbf{\Phi}^T(\mathbf{x} - \mathbf{\Phi}\mathbf{a}), \tag{5}$$

which translates into adding to the vector the difference between the minimum LS-solution ($\mathbf{a}_L S = \mathbf{\Phi}^T \mathbf{x}$) and the analysis of the reconstruction from that vector ($\mathbf{\Phi}^T \mathbf{\Phi} \mathbf{a}$). The expression of the $P_{B_p(R)}^{\perp}(\mathbf{a})$ depends on p. We explore in detail the cases $p = 0$ and $p = 1$ in the next subsections. Use of intermediate norms is also a very interesting issue, which we will consider in future works.

Then, for two given p and R values, the ℓ_p-AP method is implemented through the following iterations:

$$\hat{\mathbf{a}}_p(R)^{(0)} = P^\perp_{B_p(R)}(\mathbf{a}^{LS}),$$

$$\hat{\mathbf{a}}_p(R)^{(n+1)} = P^\perp_{B_p(R)}(P^\perp_{S(\mathbf{\Phi},\mathbf{x})}(\hat{\mathbf{a}}_p(R)^{(n)})). \tag{6}$$

We choose to stop the iterations when $\|\hat{\mathbf{a}}_p(R)^{(n+1)} - \hat{\mathbf{a}}_p(R)^{(n)}\|_2 < \delta$ for some $\delta > 0$ (see implementation details in IV). Now we study the $p = 1$ and $p = 0$ cases in detail.

1) ℓ_0-AP: In this case it is straightforward to see that $P^\perp_{B_0(R)}(\mathbf{a})$ is a hard-thresholding preserving the R largest coefficients in amplitude:

$$P^\perp_{B_0(R)}(\mathbf{a}) = \mathbf{a}^h,$$

$$a_i^h = \begin{cases} a_i, & |a_i| > \tau_h(\mathbf{a}, R) \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

Here, $\tau_h(\mathbf{a}, R)$ is the smallest threshold among those preserving the $R - n_0$ largest amplitudes, being n_0 the smallest non-negative integer that guarantees that a solution exists ($n_0 = 0$ if there are no repeated amplitude values in the interval of interest). According to the previous definition, in most cases the threshold is set to the $R + 1$-th largest coefficient amplitude in $\mathbf{a}$. This threshold is found in our implementation through a golden search. This method can be also seen as a particular case of the method described in [35], but with the difference that in that work the proposed procedure was not described and justified as a formal optimization method, unlike ours. Left panel of Figure 3 shows an illustration of ℓ_0-AP with "toy dimensions" that allow to visualize its behavior ($N = 2$, $M = 3$, $R = 1$).

Next we prove that this method leads to a local optimum in the image domain. Note that, from Eq. (5):

$$\|\mathbf{a} - P^\perp_{S(\mathbf{\Phi},\mathbf{x})}(\mathbf{a})\|_2 = \|\mathbf{\Phi}^T(\mathbf{x} - \mathbf{\Phi}\mathbf{a})\|_2 = \|\mathbf{x} - \mathbf{\Phi}\mathbf{a}\|_2, \tag{8}$$

where the last step holds because $\mathbf{\Phi}^T$ is a Parseval frame. Provided that $\hat{\mathbf{a}}_0(R)$ is a local minimum in $B_0(R)$ of the distance to $S(\mathbf{\Phi}, \mathbf{x})$, then $\exists$ a $\delta > 0$ such that $\forall \mathbf{a} \in B_0(R)$, if $\|\mathbf{a} - \hat{\mathbf{a}}_0(R)\|_2 < \delta$ then $\|\mathbf{a} - P^\perp_{S(\mathbf{\Phi},\mathbf{x})}(\mathbf{a})\|_2 \geq \|\hat{\mathbf{a}}_0(R) - P^\perp_{S(\mathbf{\Phi},\mathbf{x})}(\hat{\mathbf{a}}_0(R))\|_2$. Using (8) we obtain that $\|\mathbf{x} - \mathbf{\Phi}\mathbf{a}\|_2 \geq \|\mathbf{x} - \mathbf{\Phi}\hat{\mathbf{a}}_0(R)\|_2$. That is, $\mathbf{\Phi}\hat{\mathbf{a}}_0(R)$ is a local minimum for $\mathbf{a} \in B_0(R)$ of the Euclidean distance to $\mathbf{x}$. Regarding the convergence properties, the method quickly evolves towards the solution at the first few iterations, and then it progressively slows down, as shown in Figure 4. The convergence speed also depends on the degree of sparseness being imposed (the higher sparseness, the faster), as it is also illustrated in this figure. In this work we have been interested in exploring the absolute performance limits of the studied methods. This has required making a few hundreds iterations for each sparse approximation experiment.

In a practical implementation, though, much fewer iterations could have been made with comparable results. See details about the stopping criteria in section IV-B.

2) ℓ_1-AP method: We first show that $P^{\perp}_{B_1(R)}(\mathbf{a})$ is a soft-thresholding operation. Given and observation $\mathbf{a}^o \in \mathbb{R}^M$, the orthogonal projection onto $B_1(R)$ is:

$$P^{\perp}_{B_1(R)}(\mathbf{a}^o) = \arg \min_{\mathbf{a} \in B_1(R)} \|\mathbf{a} - \mathbf{a}^o\|_2,$$

which was previously stated in Eq. (4) as equivalent to:

$$P^{\perp}_{B_1(R)}(\mathbf{a}^o) = \arg \min_{\mathbf{a} \in \mathbb{R}^M} \|\mathbf{a}\|_1 + \lambda \|\mathbf{a} - \mathbf{a}^o\|_2^2 = \arg \min_{\mathbf{a} \in \mathbb{R}^M} \|\mathbf{a} - \mathbf{a}^o\|_2^2 + \frac{1}{\lambda} \|\mathbf{a}\|_1,$$

for some λ value depending on R. In [39] it is proven that the solution is a soft-thresholding with threshold $\tau_s(\mathbf{a}, R) = \frac{1}{\lambda}$. That is:

$$P^{\perp}_{B_1(R)}(\mathbf{a}) = \mathbf{a}^s,$$

$$a_i^s = \begin{cases} sign(a_i) \cdot (|a_i| - \tau_s(\mathbf{a}, R)), & |a_i| > \tau_s(\mathbf{a}, R) \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

Note that it is straightforward to compute R (ℓ_1-norm of the soft-thresholded vector) if we know either λ or $\tau_s(\mathbf{a}, R)$. But our aim here is the opposite, namely, to compute the threshold $\tau_s(\mathbf{a}, R)$ that will yield the desired ℓ_1-norm R of the projected vector. In appendix A we derive an efficient iterative method for this computation, and prove its convergence.

Figure 3 (right panel) illustrates the behavior of the ℓ_1-AP method with $N = 2$, $M = 3$, and $R = 1$. Only a face of $B_1(1)$ is shown here for clarity sake. It is easy to show that ℓ_1-AP provides the global optimal solution in the image domain. First note that $\hat{\mathbf{a}}_1(R)$ is the global minimum in $B_1(R)$ of the Euclidean distance to $S(\mathbf{\Phi}, \mathbf{x})$ (because the two involved sets in the projections are convex). Then, $\forall \mathbf{a} \in B_1(R)$ $\|\mathbf{a} - P^{\perp}_{S(\mathbf{\Phi},\mathbf{x})}(\mathbf{a})\|_2 \geq \|\hat{\mathbf{a}}_1(R) - P^{\perp}_{S(\mathbf{\Phi},\mathbf{x})}(\hat{\mathbf{a}}_1(R))\|_2$. Applying Eq. (5) and being $\mathbf{\Phi}^T$ a Parseval frame, we obtain that $\|\mathbf{x} - \mathbf{\Phi}\mathbf{a}\|_2 \geq \|\mathbf{x} - \mathbf{\Phi}\hat{\mathbf{a}}_1(R)\|_2$. That is, $\mathbf{\Phi}\hat{\mathbf{a}}_1(R)$ is the global minimum for $\mathbf{a} \in B_1(R)$ of the Euclidean distance to $\mathbf{x}$.

Figure 5 illustrates convergence properties of ℓ_1-AP. Interpretation is similar to Figure 4. As there is no local solutions, the convergence is more regular than with ℓ_0-AP and the number of iterations needed to converge is much smaller. Note that we have included an example (*House* image in right panel) where perfect reconstruction is achieved.

B. Minimizing the Mean Square Error for a given basis selection

In the ℓ_0-AP method, as the iterations runs, the selection of coefficients of ℓ_0-AP becomes stable, and solving Eq. (2) for a given selection of basis function implies that the final solution becomes optimal in
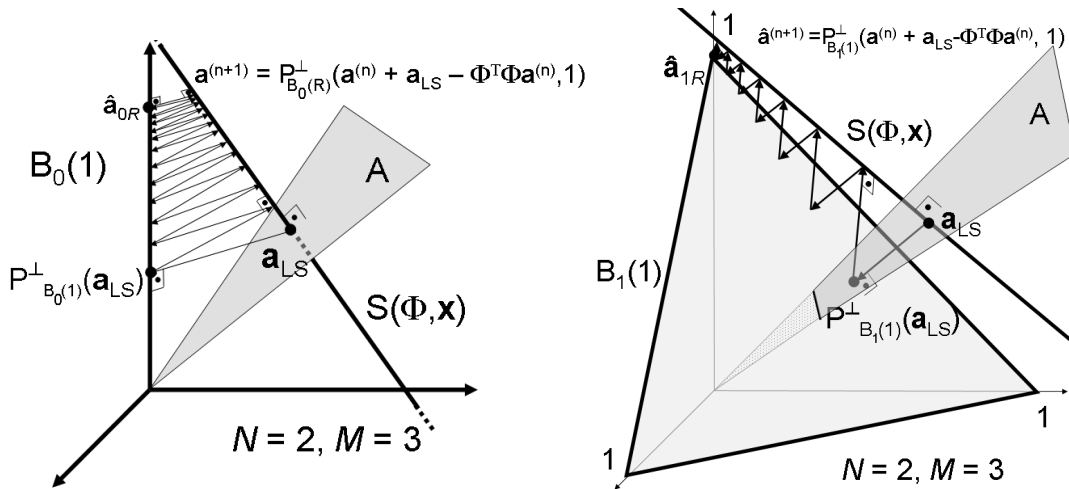
Fig. 3. **Left**, ℓ_0-norm minimization through alternate projections (ℓ_0-AP). **Right**, ℓ_1-norm minimization through alternate projections (ℓ_1-AP).
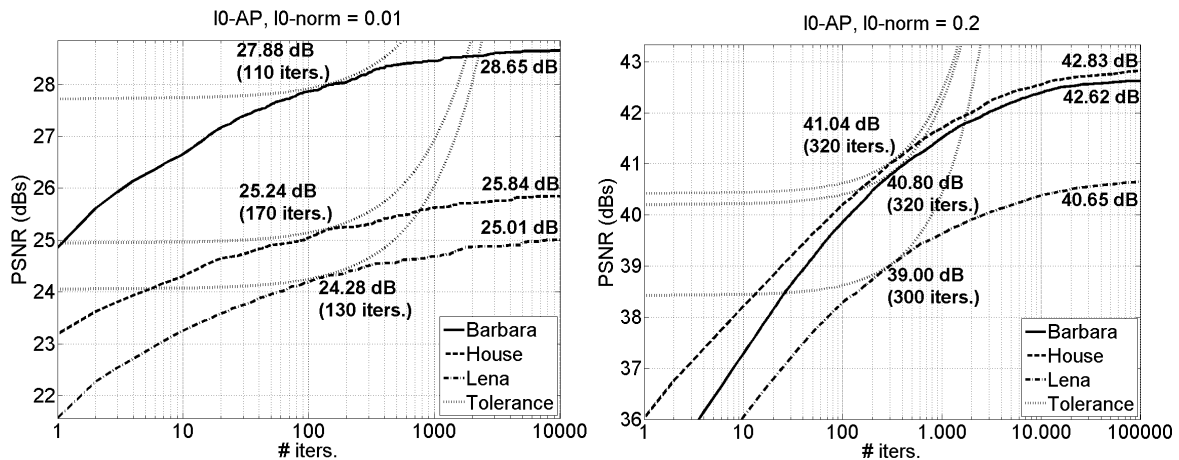
Fig. 4. Semi-logarithmic plot of PSNR vs. iterations for ℓ_0-AP using three images and two sparseness levels. Representation used is DT-CWT. Number at the end of curves is the PSNR obtained at convergence. Dotted line is the curve corresponding to the difference in PSNR used as a tolerance (a line if it was not in logarithmic scale). Iterations end when the slope of the bold line is lower than that indicated by the dotted line. Number accompanying this tangency point is the PSNR given by the method when iterations end, and the number of iterations taken. ℓ_0-norm is measured as the number of active coefficients normalized by the number of pixels of the image.

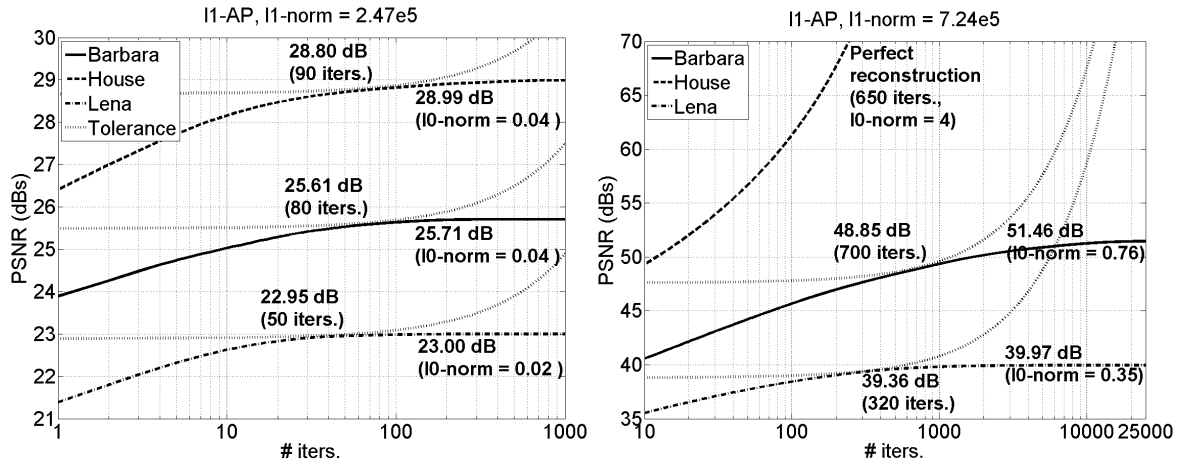Fig. 5. Semi-logarithmic convergence plots for ℓ_1-AP for three images and two ℓ_1-norm levels. Details are similar to those of Figure 4. It is also indicated the ℓ_0-norm of the solution at convergence.

a MSE sense for that selection [1]. As it was noted in [35], in the limit when fixing the number of active coefficients the two involved sets (vectorial space spanned by the selected basis functions, on the one hand, and the perfect reconstruction affine space, on the other hand) are convex, and iterations converge to the global optimum at a linear rate. However, this is not the case for a generic ℓ_p-norm, because the projection onto the ℓ_p-ball is not in general a hard-thresholding. However, as our final goal is sparse approximation, nothing prevents us from using the referred alternated projections method as a means to improve the quality of the approximation for a given set of selected basis functions, no matter how we have arrived to that selection.

Given a set I of R indices extracted from $\{1, ..., M\}$, we define $\mathbf{\Phi}_I$ as a $N \times R$ matrix formed by the columns ϕ_i of $\mathbf{\Phi}$ such that $i \in I$. We want to find:

$$\hat{\mathbf{a}}_I = \arg \min_{\mathbf{a}_I \in \mathbb{R}^R} \|\mathbf{\Phi}_I \mathbf{a}_I - \mathbf{x}\|_2, \tag{10}$$

which translates into $\hat{\mathbf{a}}_I = \mathbf{\Phi}_I^\sharp \mathbf{x}$, where $\mathbf{\Phi}_I^\sharp$ is the pseudo-inverse of $\mathbf{\Phi}_I$. Note that, when dealing with images, the size of $\mathbf{\Phi}_I$ makes impractical to compute its pseudo-inverse directly. However, we can apply instead the following iterative computation [35], [21]:

$$\mathbf{a}^{(0)} = \mathbf{D}_I \mathbf{\Phi}^T \mathbf{x},$$

$$\mathbf{a}^{(n+1)} = \mathbf{D}_I [\mathbf{a}^{(n)} + \mathbf{\Phi}^T (\mathbf{x} - \mathbf{\Phi} \mathbf{a}^{(n)})]. \tag{11}$$

---

[1]Note that the method is still suboptimal because the basis selection itself is not optimal in general.

where $\mathbf{D}_I$ is the diagonal $M \times M$ matrix such that $d_{ii} = 1$ if $i \in I$ and 0 otherwise. We prove in Appendix B that this method effectively solves for $\hat{\mathbf{a}}_I = \mathbf{\Phi}_I^\sharp \mathbf{x}$.

C. Estimation using ℓ_p-AP

Sparseness has been widely used as a prior information for solving inverse problems (e.g., [25], [39], [13]). In our context, we can choose to limit the ℓ_p-norm of the solution to a certain value R. If the degradation can be expressed (at least after the observation) as a deterministic function of the original (e.g., loss of certain identifiable pixels, bits or color components), we can define the *consistency set* as the one formed by the original coefficient vectors that would yield the observation after the observed degradation is applied. Then, for given p and R, and a certain representation, we search for the vector in the consistency set closest to $B_p(R)$. We can use the ℓ_p-AP method to solve this problem just changing $S(\mathbf{\Phi}, \mathbf{x})$ by the consistency set. Finding the orthogonal projection onto the consistency set is trivial for a wide class of "deteministic a posteriori" degradations. In general, this method has the drawback of requiring the choice of R, and this choice will vary from one application to another (see, e.g., application to in-painting in [31]). An elegant and robust choice of R that has proven to be useful for some applications such as removing quantization artifacts (see Section V) is taking the smallest R value for which the ℓ_p-ball and the consistency set still intersect.

IV. Implementation

A. Representation

Initially we tested four different tight-frames (DT-CWT [41], Curvelets [44], Steerable Pyramid [6] and an overcomplete version of Haar wavelets [45]) and chose from them the two ones providing the best sparse approximation performance on our set of five test images. Those were DT-CWT and Curvelets [2]. For an homogenous treatment of all coefficient vectors, complex coefficients of DT-CWT are separated in their real and imaginary parts. The DT-CWT redundancy factor is 4. Redundancy factor of Curvelets is ≈ 7.2. DT-CWT MATLAB ® code is available at (*http://www-sigproc.eng.cam.ac.uk/%7Engk/*). We used CurveLab 2.0 as Curvelet implementation for MATLAB ® (*http://www.curvelet.org*). In order to optimize the sparse approximation for extremely high sparseness rates, in both representations we have inserted an extra single coefficient capturing the global mean of the image. As a consequence, the best

---

[2]We have carried out most of the experiments on all the representations, and results with the other representations were qualitatively similar to the ones we present here.

approximation using a single coefficient is almost always, for natural images, a constant image with the global mean.

B. Convergence and stopping criteria

The convergence criterium of ℓ_p-AP methods is translated in our implementation into the use of two tolerance constants. The first one controls the PSNR of the estimation to decide if the convergence has been reached. We have chosen to stop when, after 10 iterations, the increase of PSNR is below the tolerance. We have empirically established this value to 0.02 dB. This stopping criteria has been drawn as dotted curves in Figures 4 and 5. Note that these curves would be straight lines tangent to the convergence curves if the abscissa axis was linear. We have tested that this criterion typically provides, for ℓ_0-AP, differences with respect to the PSNR at convergence smaller than 1 dB at high sparseness range and than 2 dB at small range. This differences are even smaller for ℓ_1-AP (thus favoring this latter method in the comparison). However, if the radius of the ℓ_p-ball used is large (specifically, greater or equal than N, for ℓ_0-AP), the method may achieve perfect reconstruction. In this case, the increase in PSNR is, according to the theory, linear. For detecting this situation we have used a second tolerance, which controls the increase of PSNR every 10 iterations and stops the execution when the difference of the last two increases in PSNR detected are below a constant (10^{-6} for ℓ_0-AP and 10^{-4} for ℓ_1-AP). In addition, the threshold search requires to set an extra tolerance parameter. In our golden search implementation for the ℓ_0-AP method, that parameter controls the search interval length. For ℓ_1-AP, we have set the tolerance on the difference of the desired and achieved radius of the ℓ_1-ball. As a good compromise between accuracy and computational load, we have set both tolerances to 0.1.

V. RESULTS AND DISCUSSION

The aim of our experiments is to determine the sparse approximation performance (Eq. (2)) of a wide class of different methods, including our new ones, implemented on widely used tight frames, for a wide range of R, and for a set of representative test images. Reconstruction quality is measured by mean square error, then averaged for all images, and finally expressed in PSNR terms. We use the same set of standard images as in Figure 1. The ℓ_0-norm has been normalized by N (number of pixels in the image). Note that we have used in our figures a logarithmic scale for the ordinate axis, no matter the PSNR is already a logarithmic measurement. We believe that this, although unusual, is justified in this case because it greatly improves the visualization of the resulting curves. Each marker in the curves correspond to a measurement. We have linearly interpolated intermediate values.

A. Some previous methods

Our first experiments compare commonly used sparse approximation strategies. Here we have set two goals: a) compare the use of hard vs. soft-thresholding in iterative thresholding algorithms; and b) compare accumulative vs. direct selection of basis functions in greedy algorithms.

Regarding the first goal, we have implemented a thresholding method that iterates between hard/soft-thresholding with fixed threshold and projects back onto the set $S(\mathbf{\Phi}, \mathbf{x})$ of vectors with perfect reconstruction. This is similar to [34] when using hard-thresholding and to [39] when using soft-thresholding. We have used the same stopping criterium as with ℓ_p-AP. We label these methods ℓ_0-FT and ℓ_1-FT respectively. Note that our implementation of these methods only differs from ℓ_0-AP and ℓ_1-AP in that we use a fixed radius of the ℓ_p-ball whereas they use a fixed threshold.

To compare greedy heuristical methods, we have implemented StOMP [17] and the method presented in [21], which we call here DT+OP (from Direct Thresholding plus LS-Optimization). To define the threshold values used by StOMP, we have set a sampling over the number of coefficients used, and we have set the number of basis functions added to the selection of StOMP at each step as the difference with the next sample. DT+OP directly threshold the forward representation. Both methods use Eq. (13) to optimize the quality of the reconstruction. DT+OP apply them once after the thresholding, whereas StOMP apply them at each iteration (double loop). Here we have also used the stopping criteria described for ℓ_p-AP.

Figure 6 shows graphically some numerical results of this experiment. Left panel is for DT-CWT and right panel for Curvelets. This figure shows that performance of hard-thresholding is generally better than that of soft-thresholding. It also shows that results obtained with DT+OP outperform our implementation of StOMP, except in the high sparseness range (which has little practical importance). Among all these compared methods, ℓ_0-FT has the best performance. Other authors have pointed out before that hard-thresholding better favors the compaction of energy [32], [36], [37], [31], as compared to soft-thresholding.

B. Comparing ℓ_0-AP to ℓ_1-AP and to previous methods

Our second experiment compares ℓ_0-AP vs. ℓ_1-AP. We also compare to the method resulting from MSE-optimizing (using iterations (13)) the basis functions selected by ℓ_1-AP , which we term ℓ_1-AP+OP. Finally, we have included ℓ_0-FT and DT+OP too as the two best methods from from Figure 6. Figure 7 shows the result of this experiment. Left panel shows DT-CWT results and right one Curvelet results. We can see how ℓ_0-AP clearly outperforms ℓ_1-AP, even though the latter is optimally minimizing the ℓ_1-norm.
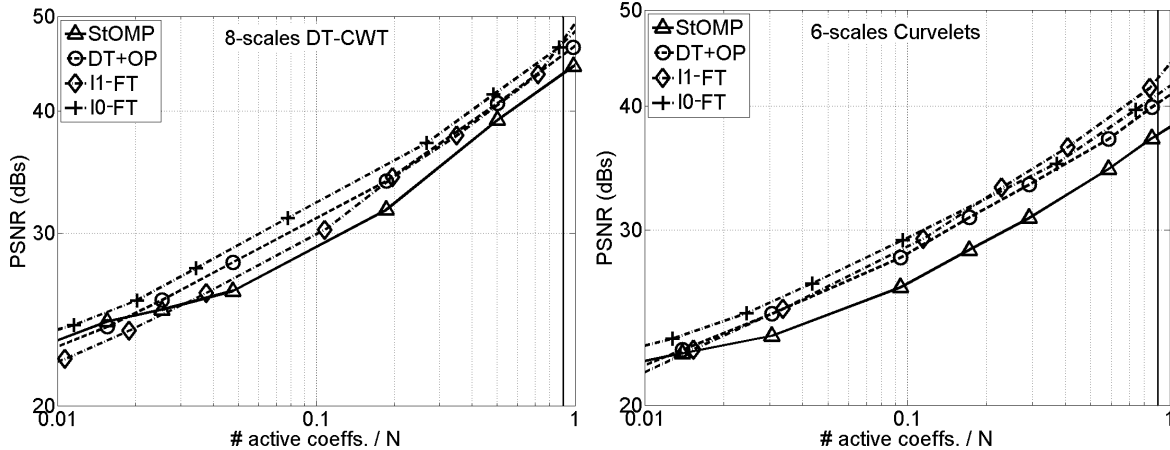
Fig. 6. Averaged compaction results in our test set of StOMP [17], DT+OP [21], ℓ_0-FT [34] and ℓ_1-FT [39] in our test set. **Left**, using DT-CWT [41]. **Right**, using Curvelets [44].
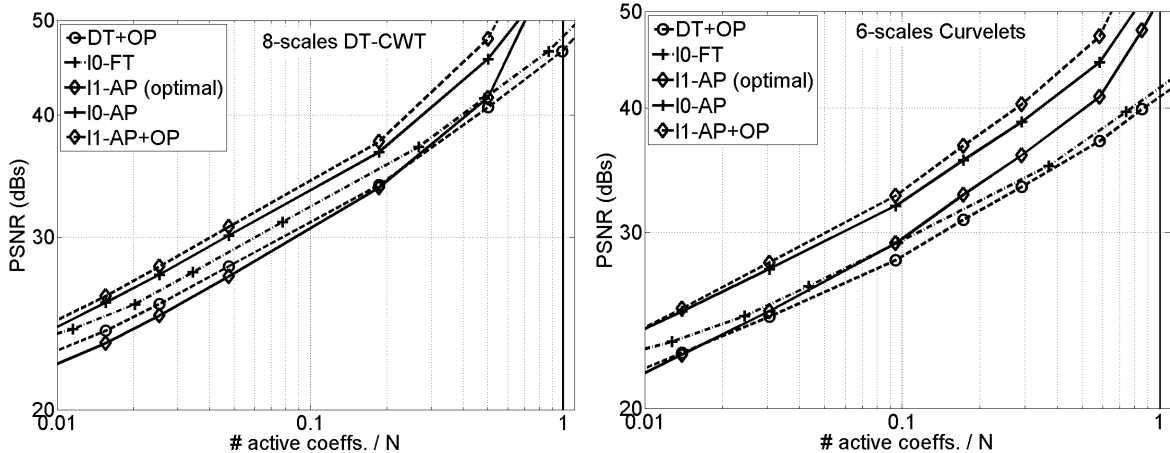
Fig. 7. Averaged compaction results in the test set comparing ℓ_0-AP w.r.t. ℓ_1-AP, ℓ_1-AP+OP, ℓ_0-FT [34] and DT+OP [21]. **Left**, using 8-scales DT-CWT. **Right**, using 6-scales Curvelets.

We also observe how ℓ_1-AP+OP improves drastically the results from ℓ_1-AP, slightly outperforming ℓ_0-AP. This shows that selection of coefficients made by ℓ_1-AP is generally better than the one made by ℓ_0-AP, specially in the low sparseness range. This seems a consequence of ℓ_0-AP getting trapped in local optima, which rapidly increase their number as the sparseness level decreases. Note how ℓ_0-AP performs also significantly better than DT+OP and ℓ_0-FT. It is interesting how fixing the radius of the ℓ_p-ball proves to be much better than fixing the threshold at each iterations. Tables I and II show numerical results for the curves of Figure 7.

Figure 8 visually compares the methods using *Einstein* image and $0.0765N$ Curvelets coefficients. From

top to bottom, left column shows the original image, result from ℓ_1-AP (30.85 dB) and ℓ_1-AP+OP (33.52 dB). Note the great improvement in visual quality obtained by post-optimizing the selected coefficients. Right column shows DT+OP (30.21 dB), ℓ_0-FT (30.65 dB) and ℓ_0-AP (32.98 dB). Though more than a half dB below in PSNR, there is not significant visual difference between ℓ_0-AP and ℓ_1-AP+OP.

As we have pointed out before, the ℓ_0-AP method is equivalent to [35] when using a fixed number of coefficients at each iteration and no additional heuristics are used in the referred algorithm. However, the referred authors threshold the magnitude of each complex coefficient, whereas, in order to keep a consistent procedure for all Parseval frames used in this paper, we separate the real and imaginary parts Thus, to properly compare their results with ours we double the number of selected coefficients given in [35] (so obtaining the same number of real coefficients), and we used DT-CWT with 5-scales to match representations. Using a fixed number of $24,000$ selected coefficients with *Lena* 512×512^3 we get an improvement over his result of 2.02 dB (39.09 vs. 37.07 dB). In [35] is also presented a dynamic version of the method which increases the number of complex coefficients used at each iteration (from $2,400$ to $12,000$ in 30 iterations). Their result is still 0.41 dB below ours (38.68 dB). However, it is easy to test that this difference is due to the extra flexibility of our scheme that allows to independently choose for the real or imaginary part of every complex coefficient. This factor seems more important in this case than the use of dynamic thresholding. Actually, if complex coefficients are not separated in our ℓ_0-AP implementation, our result goes 1.31 dB below theirs.

C. Computation

Iteration time is dominated in all methods by one image analysis and one image synthesis operation at each iteration. However, searching for the threshold also takes a significant part. Both ℓ_1-AP and ℓ_0-AP require additional search of a threshold, whereas other methods like DT+OP and ℓ_0-FT do not, and, thus, they are relatively faster. Nevertheless, the time consumed by the methods depends more on the average number of iterations until the convergence stop criterion is met. In Table III we show that ℓ_0-AP takes more iterations to converge. Note that this is in part due to the tolerance used for detecting perfect reconstruction is comparatively smaller (see subsection IV-B), so the number of iterations taken when using low sparseness levels is much higher. It is important to point, as we did in section III, that most applications will not require so many iterations. Nevertheless, it is fair to recognize that dynamic thresholding strategies [35], [40], [46], [38] have proven to be intrinsically faster than those based on

---

[3]We thank Prof. Nick Kingsbury for helping us to replicate his experiment.

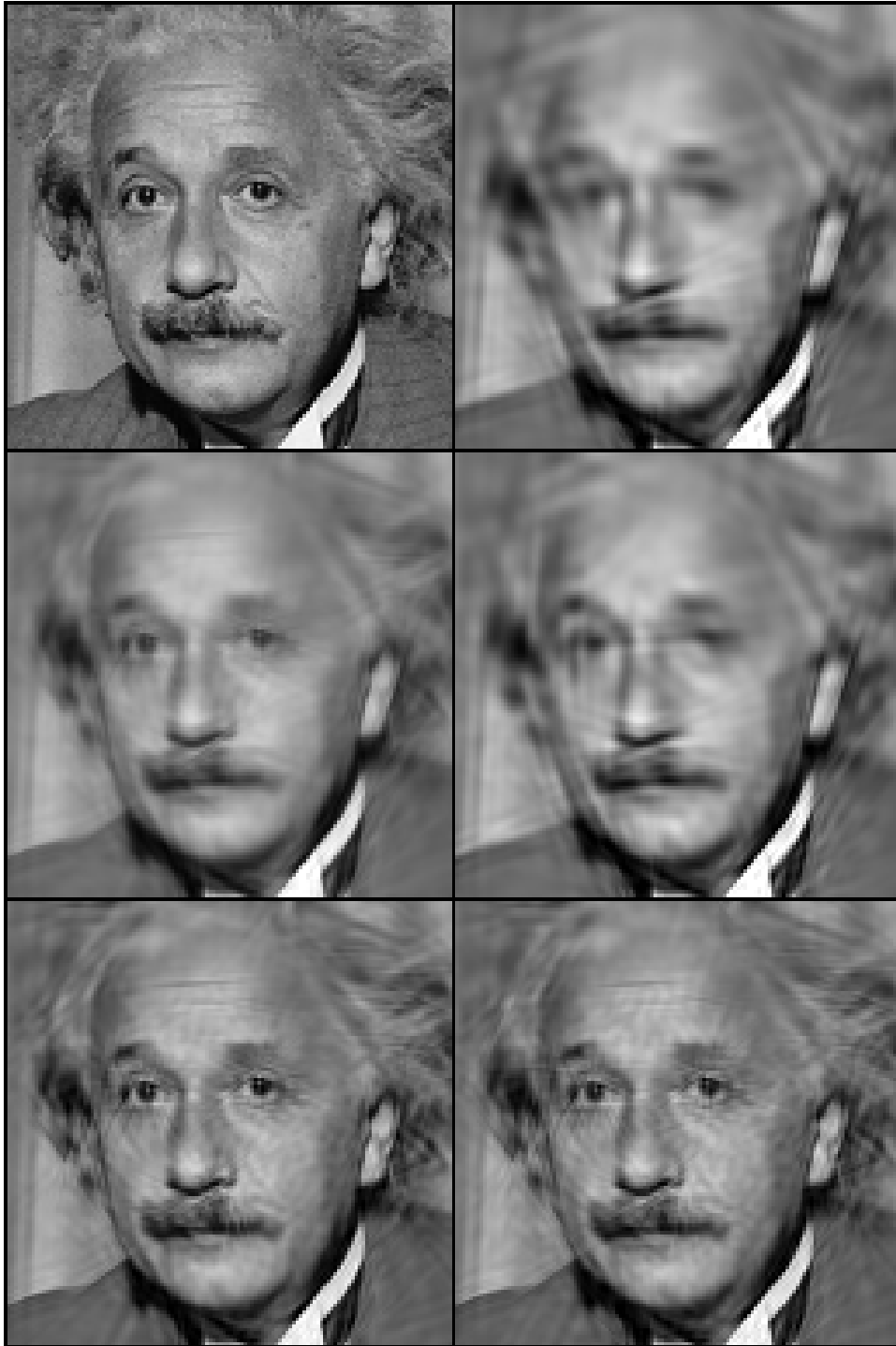Fig. 8.   Visual comparison of the method using $7.65e - 2 \cdot N$ Curvelets coefficients for the *Einstein* image. Results are cropped to 128×128 crop, starting at (71,41), for visibility. **Left column**, from top to bottom: original image, ℓ_1-AP (30.85 dB), ℓ_1-AP+OP (33.52 dB). **Right column**, from top to bottom, DT+OP (30.21 dB, ℓ_0-FT (30.65 dB) ℓ_0-AP (32.98 dB).

fixing the selection mask, the threshold or the number of selected basis functions.

As an illustrative example of running times with the (very demanding) stopping criteria described here, using a Pentium IV with 3.4 GHz and 1 GB RAM, over 256×256 imagex, ℓ_0-AP takes about 7 minutes to stop using DT-CWT and about 1 hour using Curvelets. On the other hand, ℓ_1-AP takes about 3 minutes using DTCWT and 30 minutes using Curvelets.

D. *An application example: de-quantizing*

We have performed another experiment to demonstrate the applicability of ℓ_0-AP to the removal of spatial quantization artifact as explained in Section III-C. Our observation is the *Einstein* image quantized using 3-bits. We compare the performance of ℓ_1-AP and ℓ_0-AP for removing its artifacts, using both DT-CWT and Curvelets. Figure 9 shows this comparison. We can see that ℓ_1-AP results are too smooth, even decreasing the PSNR with respect to the degraded observation. On the other hand, though showing artifacts inherent to the tight frames used, the performance of ℓ_0-AP is much more satisfactory. Restoration of other type of degraded images, such as filling-in missing pixels or texture separation require in general different strategies to choose the R parameter and have not been addressed here (e.g., [31]).

VI. CONCLUSIONS

We have proposed a new optimization method, called ℓ_p-AP, which minimizes the mean square error of the reconstruction of an image from a representation vector given a maximum ℓ_p-norm for that vector. It iteratively orthogonally projects onto the ℓ_p-ball of given radius R, given p and R; and onto the set of vectors providing perfect reconstruction. It achieves the global optimum when $p \geq 1$ and a local optimum when $0 < p \leq 1$. We have applied this method to solve the sparse approximation problem (minimizing the MSE using a given number of coefficients), focusing on the $p = 0$ and $p = 1$ cases. The ℓ_0-AP case translates into similar iterations than [35], in its simplest version.

We have shown that ℓ_0-AP clearly outperforms ℓ_1-AP, which is optimal for solving the commonly approached convex relaxation problem. This result demonstrates that equivalency conditions for minimizing ℓ_1 and ℓ_0-norm do not hold using natural images and typical image representations. This is an important practical issue for image processing. However, we have also tested that the selection of active coefficients resulting by the convex relaxation ($p = 1$) is generally better than the one made by ℓ_0-AP. Thus, slightly improved performance may be achieved by post-optimizing the amplitude of the non-zero coefficients in the ℓ_1-AP solution.
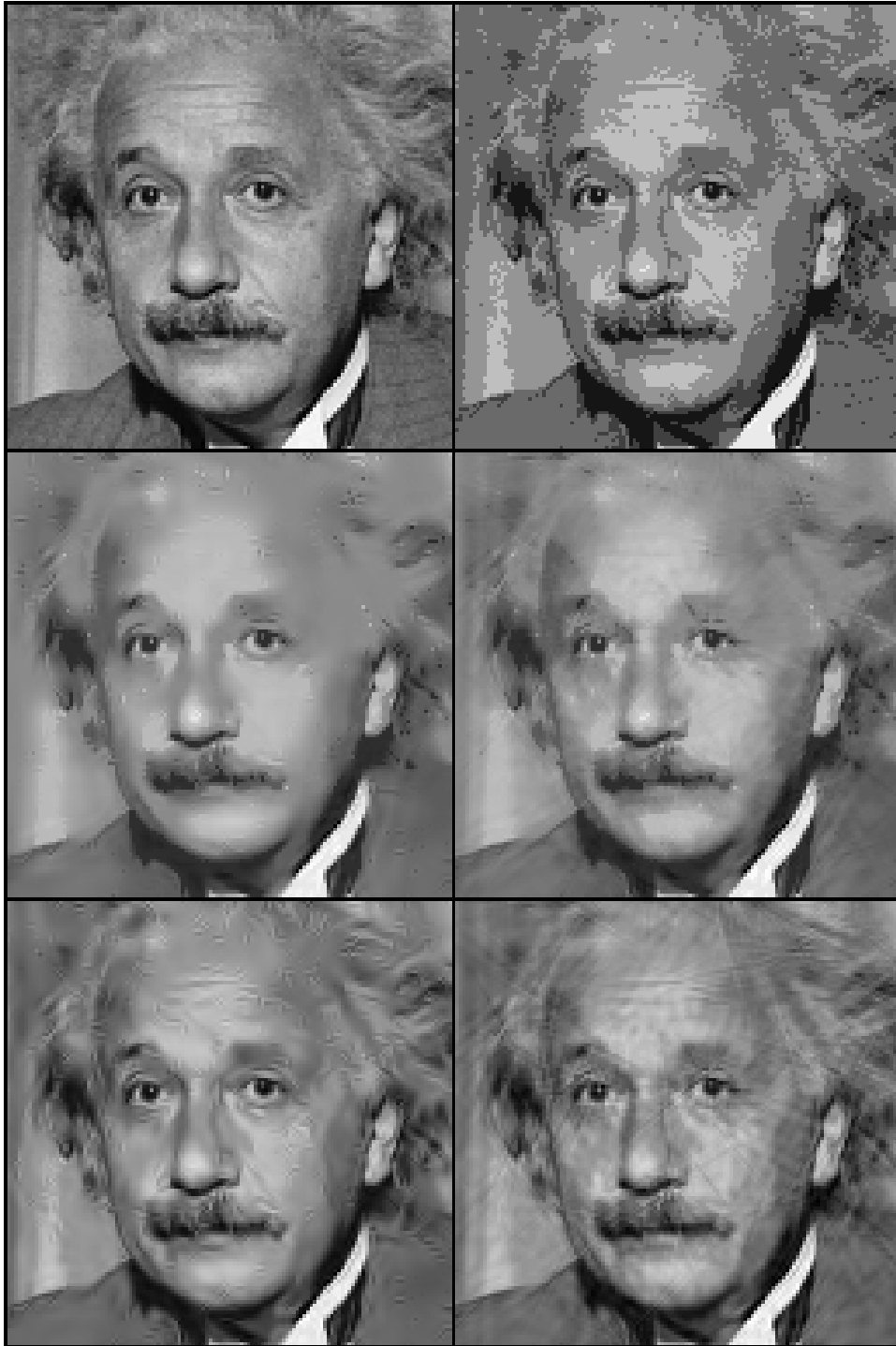
Fig. 9. Application of ℓ_1-AP and ℓ_0-AP to removing quantization artifacts. **First row**, *Einstein* image, cropped to 128×128 for artifacts visibility starting at (71,41); and 3-bits quantized observation (29.58 dB). **Second row**, ℓ_1-AP results using 8-scales DT-CWT (28.14 dB) and 6-scales Curvelets (28.61 dB). **Third row** ℓ_0-AP results using the same tight-frames (31.24 dB and 31.29 dB respectively).

We have also compared with greedy and iterative thresholding heuristical methods, showing that ℓ_0-AP also outperforms the iterative fixed thresholding (either soft or hard) and the StOMP algorithm version implemented for this purpose. More exhaustive tests would be necessary to state the superiority of ℓ_0-AP over greedy (MP-like) algorithms in general but we have not tried that because of the huge computational cost of the strict implementations of theses techniques. Among these methods, iterative fixed thresholding provided clearly the second best energy compaction performance. Although not compared here in detail, methods based on dynamic adjustment of a hard threshold through iterations (e.g. [33], [34], [35], [38]) clearly have currently the highest compaction potential. However, those methods, unlike ours, have not been mathematically formulated as optimization techniques. It is easy to adapt ℓ_0-AP to iteratively increase the number of coefficients (as in [35]), but we have focused here on a fully justified optimization model. An additional fact is that, for some restoration tasks (like image de-quantizing), we have experienced that better energy compaction does not always imply better restoration performance.

We have shown here how the ℓ_p-AP method can be easily adapted to image restoration. The resulting method has the drawback of requiring a criterion for choosing the desired ℓ_p-ball radius of the synthesis vector. We have shown a successful example of application to the removal of spatial quantization artifacts, where the radius choice has been made to maximize the degree of sparseness compatible with the observation. In this application, convex relaxation has yielded relatively poor results.

In the future we pretend to study the use of intermediate norms. This is justified as a way to explore a better compromise between the good performance of ℓ_0-AP and the ability of ℓ_1-AP to avoid local minima. Apart from that, in terms of is application to restoration, the solution of the problem at hand is equivalent to a MAP estimation under white Gaussian noise when the marginal prior is a generalized Gaussian [39], as pointed before. It has been observed that typically the marginal pdf is better adjusted with intermediate norms (≈ 0.7) (see, e.g. [1], [39]). However, using empirical priors adjusted for analysis coefficients is justified only if we apply them to analysis (but not to synthesis) coefficients. It is easy to adapt our method in that direction by using an extra projection in the iteration loop, projection that does not break the consistency of the theoretical frame [47]. Image processing based on sparsifying the analysis (linear) coefficients has been successfully used to separate mixed images (MCA [40]) and to fill in missing pixels [46] [4]. We are also currently working on a practical and theoretically grounded optimization method using dynamic iterative thresholding, with a much higher performance than ℓ_0-AP, both in sparse approximation power and in computational load terms [47].

---

[4]See [48] for an interesting discussion on the analysis-sense vs. the synthesis sense sparseness applied to estimation.

Appendix A

Method to find the threshold given the radius of the ℓ_1-ball to project onto

In [39] was demonstrated that the orthogonal projection operator of a vector $\mathbf{a} \in \mathbb{R}^M$ onto $B_1(R)$, the ℓ_1-ball of radius R, is a soft-thresholding (Eq. (9), in Section III). The remaining problem in our context is, given the vector $\mathbf{a}$, finding the threshold $\tau_s(\mathbf{a}, R)$ (or, equivalently, the λ of Eq. (3), as $\tau_s = 1/\lambda$) that yields the desired ℓ_1-norm of the thresholded vector.

First, we express the ℓ_1-norm (i.e., the radius R) of the soft-thresholded version of a vector $\mathbf{a}$ with a given threshold [5] τ_s. Naming $\Upsilon(\mathbf{a}, \tau) = \{i \in \{1, ..., M\} : |a_i| > \tau\}$ the set of coefficient indices whose amplitude is strictly above the threshold we can write:

$$
\begin{aligned}
R &= \sum_{\Upsilon(\mathbf{a}, \tau_s)} (|a_i| - \tau_s) \\
&= \left(\sum_{\Upsilon(\mathbf{a}, \tau_s)} |a_i| \right) - card(\Upsilon(\mathbf{a}, \tau_s)) \cdot \tau_s,
\end{aligned}
$$

which yields:

$$
\tau_s = \frac{\left(\sum_{\Upsilon(\mathbf{a}, \tau_s)} |a_i| \right) - R}{card(\Upsilon(\mathbf{a}, \tau_s))}, \tag{12}
$$

where $card(\cdot)$ indicates cardinality of a set. The term on the right depends on τ_s, but we can solve this equation iteratively by using:

$$
\tau^{(0)} = 0,
$$

$$
\tau^{(n+1)} = \frac{\left(\sum_{\Upsilon(\mathbf{a}, \tau^{(n)})} |a_i| \right) - R}{card(\Upsilon(\mathbf{a}, \tau^{(n)}))}.
$$

Iterations end when $\|\tau^{(n+1)} - \tau^{(n)}\|_2$ is below a threshold (see subsection IV-B).

We assume here that $\|\mathbf{a}\|_1 > R$, because otherwise the projection onto $B_1(R)$ is trivial (the identity). Next, we demonstrate that the iterations converge to τ_s, by noting first that $R(\tau)$ is a strictly monotonically decreasing function, and, thus, so it is $\tau(R)$. This implies that Eq. (12) has a unique solution in τ_s. Note that if we take $\tau^{(n+1)} = \tau^{(n)}$ then Eq. (12) is satisfied, so we know that if iterations converge they have to do it to the unique solution τ_s. Now, to prove convergence to τ_s it is sufficient to demonstrate that the succession $\tau^{(n)}$ converges. This can be done by showing that being $\tau^{(n)}$ monotonically increasing, it never gets above τ_s, as we do next. We start by noting that $\tau^{(0)} = 0 < \tau_s$. Assuming $\tau^{(n)} \le \tau_s$, then:

$$
\sum_{\Theta(\mathbf{a}, \tau^{(n)}, \tau_s)} |a_i| \le \sum_{\Theta(\mathbf{a}, \tau^{(n)}, \tau_s)} \tau_s,
$$

[5]For notation clarity, we drop here the dependence of τ_s on $(\mathbf{a}, R)$.

where $\Theta(\mathbf{a}, \tau_1, \tau_2) = \{i \in \{1, ..., M\} : \tau_1 < |a_i| \leq \tau_2\}$. From here we can derive the following:

$$\sum_{\Upsilon(\mathbf{a},\tau^{(n)})} |a_i| - \sum_{\Upsilon(\mathbf{a},\tau_s)} |a_i| \leq \sum_{\Upsilon(\mathbf{a},\tau^{(n)})} \tau_s - \sum_{\Upsilon(\mathbf{a},\tau_s)} \tau_s,$$

$$\sum_{\Upsilon(\mathbf{a},\tau^{(n)})} |a_i| - R \leq \sum_{\Upsilon(\mathbf{a},\tau^{(n)})} \tau_s,$$

$$\sum_{\Upsilon(\mathbf{a},\tau^{(n)})} |a_i| - R \leq card(\Upsilon(\mathbf{a}, \tau^{(n)})) \cdot \tau_s,$$

$$\frac{\sum_{\Upsilon(\mathbf{a},\tau^{(n)})} |a_i| - R}{card(\Upsilon(\mathbf{a}, \tau^{(n)}))} \leq \tau_s,$$

$$\tau^{(n+1)} \leq \tau_s.$$

So the succession is bounded above by τ_s. Then, as $\sum_{\Theta(\mathbf{a},\tau^{(n)},\tau_s)} |a_i| \geq \sum_{\Theta(\mathbf{a},\tau^{(n)},\tau_s)} \tau^{(n)}$, and also $\sum_{\Upsilon(\mathbf{a},\tau_s)} \tau_s \geq \sum_{\Upsilon(\mathbf{a},\tau_s)} \tau^{(n)}$, then

$$\sum_{\Theta(\mathbf{a},\tau^{(n)},\tau_s)} |a_i| + \sum_{\Upsilon(\mathbf{a},\tau_s)} \tau_s \geq \sum_{\Upsilon(\mathbf{a},\tau^{(n)})} \tau^{(n)},$$

and we can derive the following inequalities:

$$\sum_{\Upsilon(\mathbf{a},\tau^{(n)})} |a_i| - \sum_{\Upsilon(\mathbf{a},\tau_s)} |a_i| + \sum_{\Upsilon(\mathbf{a},\tau_s)} \tau_s \geq \sum_{\Upsilon(\mathbf{a},\tau^{(n)})} \tau^{(n)},$$

$$\sum_{\Upsilon(\mathbf{a},\tau^{(n)})} |a_i| - R \geq \sum_{\Upsilon(\mathbf{a},\tau^{(n)})} \tau^{(n)},$$

$$\sum_{\Upsilon(\mathbf{a},\tau^{(n)})} |a_i| - R \geq card(\Upsilon(\mathbf{a}, \tau^{(n)})) \cdot \tau^{(n)},$$

$$\frac{\sum_{\Upsilon(\mathbf{a},\tau^{(n)})} |a_i| - R}{card(\Upsilon(\mathbf{a}, \tau^{(n)}))} \geq \tau^{(n)},$$

$$\tau^{(n+1)} \geq \tau^{(n)}.$$

As a consequence the succession is monotonically increasing, and the proof is complete.

An alternative demonstration of this method and its convergence comes from noting that the orthogonal projection onto $B_1(R)$ can be implemented as the orthogonal projection of the amplitudes a_i's on the intersection of the convex sets $C_1 = \{\mathbf{b} \in \mathbb{R}^M : \sum_i b_i \leq R\}$, and $C_2 = \{\mathbf{b} \in \mathbb{R}^M : b_i \geq 0\}$, applying alternated orthogonal projections and preserving the signs of the coefficients not being set to zero.

APPENDIX B

MEAN SQUARE ERROR APPROXIMATION GIVEN A SUBSET OF ACTIVE COEFFICIENTS BY ALTERNATED PROJECTIONS.

We prove here that Eq. (11) provide the pseudoinverse solution to the approximation problem for a given set of active coefficients. That is, they provide the mean square error (MSE) solution to Eq. (10) when there is no exact representation of the image, and the minimum ℓ_2 norm solution otherwise. Given an image $\mathbf{x} \in \mathbb{R}^N$, a subset I of R indices extracted from $\{1, ..., M\}$, and an $N \times R$ matrix $\mathbf{\Phi}_I$ formed by the columns ϕ_i from $\mathbf{\Phi}$ such that $i \in I$, we want to solve:

$$\hat{\mathbf{a}}_I = \arg \min_{\mathbf{a}_I} \|\mathbf{\Phi}_I \mathbf{a}_I - \mathbf{x}\|_2, \tag{13}$$

that can be expressed as:

$$\hat{\mathbf{a}}_I = \mathbf{\Phi}_I^\sharp \mathbf{x},$$

where $\mathbf{\Phi}_I^\sharp$ is the pseudoinverse of $\mathbf{\Phi}_I$. We study the two cases of interest: (1) $range(\mathbf{\Phi}_I) = R \leq N$, and (2) $R > range(\mathbf{\Phi}_I) = N$.

A. Case 1: $range(\mathbf{\Phi}_I) = R \leq N$

We can express in this case:

$$\hat{\mathbf{a}}_I = [\mathbf{\Phi}_I^T \mathbf{\Phi}_I]^{-1} \mathbf{\Phi}_I^T \mathbf{x}.$$

The inverse involved is potentially huge. Fortunately, we can use the Taylor expansion of the inverse of the matrix [49] so we have:

$$\hat{\mathbf{a}}_I = \sum_{k=0}^{\infty} (\mathbf{I} - \mathbf{\Phi}_I^T \mathbf{\Phi}_I)^k \mathbf{\Phi}_I^T \mathbf{x}.$$

As necessary convergence condition for the Taylor expansion, we check that for usual frames in image representation eigenvalues of $(\mathbf{I} - \mathbf{\Phi}_I^T \mathbf{\Phi}_I)$ are not above 1 in absolute value. From there we derive an iterative method to calculate $\hat{\mathbf{a}}_I$:

$$\mathbf{a}_I^{(n+1)} = \mathbf{a}_I^{(n)} + \mathbf{\Phi}_I^T (\mathbf{x} - \mathbf{\Phi}_I \mathbf{a}_I^{(n)}). \tag{14}$$

We now define $\mathbf{S}_I$ as the $R \times N$ matrix selecting the R coefficients indicated by I. Then, $\mathbf{S}_I^T$ is the operator that expands a $R \times 1$ vector into a $N \times 1$ vector reinserting each coefficient in its original position and setting the rest to zero. Noting that $\mathbf{\Phi}_I = \mathbf{\Phi} \mathbf{S}_I^T$ and $\mathbf{\Phi}_I^T = \mathbf{S}_I \mathbf{\Phi}^T$, we have:

$$\mathbf{a}_I^{(n+1)} = \mathbf{a}_I^{(n)} + \mathbf{S}_I \mathbf{\Phi}^T (\mathbf{x} - \mathbf{\Phi} \mathbf{S}_I^T \mathbf{a}_I^{(n)}).$$

Multiplying by $\mathbf{S}_I^T$ (which is an expansion matrix, so it does not destroy any information):

$$\mathbf{S}_I^T \mathbf{a}_I^{(n+1)} = \mathbf{S}_I^T \mathbf{a}_I^{(n)} + \mathbf{S}_I^T \mathbf{S}_I \mathbf{\Phi}^T (\mathbf{x} - \mathbf{\Phi} \mathbf{S}_I^T \mathbf{a}_I^{(n)}).$$

and as $\mathbf{a}_I = \mathbf{S}_I \mathbf{a}$, for some $\mathbf{a} \in \mathbb{R}^M$ we can write:

$$\mathbf{S}_I^T \mathbf{S}_I \mathbf{a}^{(n+1)} = \mathbf{S}_I^T \mathbf{S}_I \mathbf{a}^{(n)} + \mathbf{S}_I^T \mathbf{S}_I \mathbf{\Phi}^T (\mathbf{x} - \mathbf{\Phi} \mathbf{S}_I^T \mathbf{S}_I \mathbf{a}^{(n)}).$$

Let $\mathbf{D}_I$ be a diagonal $M \times M$ matrix where $d_{ii} = 1$ if $i \in I$ and 0 otherwise. Noting that $\mathbf{S}_I^T \mathbf{S}_I = \mathbf{D}_I$ and using the fact that $\mathbf{D}_I$ is idempotent, we get:

$$\mathbf{D}_I \mathbf{a}^{(n+1)} = \mathbf{D}_I [\mathbf{D}_I \mathbf{a}^{(n)} + \mathbf{\Phi}^T (\mathbf{x} - \mathbf{\Phi} \mathbf{D}_I \mathbf{a}^{(n)})].$$

As the right term only depends on $\mathbf{D}_I \mathbf{a}^{(n)}$, then these iterations are completely equivalent as those in Eq. (11), being $\mathbf{a}^{(n)}$ the intermediate result of Eq. (11).

B. Case 2: $R > range(\mathbf{\Phi}_I) = N$

Here, Eq. (13) has infinite solutions with perfect reconstruction of $\mathbf{x}$. The pseudoinverse gives the minimum Euclidean norm solution:

$$\hat{\mathbf{a}}_I = \mathbf{\Phi}_I^T [\mathbf{\Phi}_I \mathbf{\Phi}_I^T]^{-1} \mathbf{x}.$$

We can write $\hat{\mathbf{a}}_I = \mathbf{\Phi}_I^T \hat{\mathbf{z}}_I$, where $\hat{\mathbf{z}}_I = [\mathbf{\Phi}_I \mathbf{\Phi}_I^T]^{-1} \mathbf{x}$. Then:

$$\hat{\mathbf{z}}_I = \sum_{k=0}^{\infty} [\mathbf{I} - \mathbf{\Phi}_I \mathbf{\Phi}_I^T]^k \mathbf{x},$$

which can be computed through the iterative method:

$$\mathbf{z}_I^{(n+1)} = \mathbf{z}_I^{(n)} - \mathbf{\Phi}_I \mathbf{\Phi}_I^T \mathbf{z}_I^{(n)} + \mathbf{x}.$$

Multiplying by $\mathbf{\Phi}_I^T$:

$$\mathbf{\Phi}_I^T \mathbf{z}_I^{(n+1)} = \mathbf{\Phi}_I^T \mathbf{z}_I^{(n)} - \mathbf{\Phi}_I^T \mathbf{\Phi}_I \mathbf{\Phi}_I^T \mathbf{z}_I^{(n)} + \mathbf{\Phi}_I^T \mathbf{x},$$

and substituting $\mathbf{\Phi}_I^T \mathbf{z}_I^{(n)}$ by $\mathbf{a}_I^{(n)}$ we obtain Eq. (14) and, thus, the solution is achieved by the same iterative method as in the previous case.

REFERENCES

[1] E.P. Simoncelli, E.H. Adelson. Noise Removal via Bayesian Wavelet Coring. *Proc. 3rd IEEE Int'l. Conf. on Image Proc.*, **I**, pp. 379-382, September 1996.

[2] J. Portilla, V. Strela, M.J. Wainwright, E.P. Simoncelli. Image Denoising using Scale Mixtures of Gaussians in the Wavelet Domain. *IEEE Trans. on Image Proc.*, **12**, 11, pp. 1338-1351, November 2003.

[3] B.A. Olshausen, D.J. Field. Sparse Coding of Sensory Inputs. *Curr. Opinion in Neurobiology*, **14**, pp. 481-487, July, 2004.

[4] R. Neelamani, H. Choi, R. Baraniuk. ForWard: Fourier-Wavelet Regularized Deconvolution for Ill-Conditioned Systems. *IEEE Trans. on Signal Proc.*, **52**, 2, pp. 418-433, April 2004.

[5] S. Mallat. A Theory for multiresolution signal decomposition: The wavelet representation. *PAMI*, **11**, pp. 674-693, 1989.

[6] E.P. Simoncelli, W.T. Freeman, E.H. Adelson, D.J. Heeger. Shiftable Multi-Scale Transforms. *IEEE Trans. on Inf. Theory*, **38**, 2, 587-607, 1992.

[7] R.R. Coifman, D.L. Donoho. Translation Invariant De-noising. *Lecture Notes in Statistics*, **103**, 125-50, 1995.

[8] R.R. Coifman, M.V. Wickerhauser. Entropy-Based Algorithms for Best-Basis Selection. *IEEE Trans. on Inf. Theory*, **38**, 2, pp. 713-718, March 1992.

[9] S. Mallat, Z. Zhang. Matching Pursuit in Time-Frequency Dictionary. *IEEE Trans. on Signal Proc.*, **41**, 12, pp. 3397-3415, December 1993.

[10] B.A. Olshausen, D.J. Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, **381**, pp. 607-9. 1995.

[11] P. Comon. Independent Component Analysis. A New Concept?. *Signal Processing*, **36**, pp. 287-314, 1994.

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

[13] J. Mairal, G. Sapiro, M. Elad. Multiscale Sparse Image Representation with Learned Dictionaries. *Proc. 14th IEEE Int'l. Conf. on Image Proc.*, San Antonio, TX, 16-19 September 2007.

[14] C. Distante, A. Leone, L. My, M. Rizzello, P. Siciliano. A Video Compression Algorithm based on Matching Pursuit Integrated into a Wireless Embedded Sensor Node Compliant with IEEE 1451.1 Standard Architecture. *Proc. 2nd IASTED Int'l. Conf.: Comm. and Comp. Network*, Cambridge, MA, 8-10 November 2004.

[15] Adel Rahmoune, Pierre Vandergheynst, Pascal Frossard. MP3D: Highly Scalable Video Coding Scheme Based on Matching Pursuit. *Proc. IEEE Int'l. Conf. on Acoustic, Speech and Sig. Proc.*, **3**, pp. 133-136, Montreal, Canada, 2004.

[16] J.L. Lin, W.L. Hwang, S.C. Pei. Video Compression based on Orthonormal Matching Pursuits. *Proc. Int'l. Symp. on Circuits and Systems, 2006 (ISCAS 2006)*, pp. 4-8, 21-24 May 2006.

[17] D.L. Donoho, Y. Tsaig, I. Drori, J.L. Starck. Sparse Solution of Undetermined Linear Equations by Stagewise Orthogonal Matching Pursuit. *Technical report included in Documentation of SparseLab (http://sparselab.stanford.edu/)*, April 2006.

[18] M. Andrle, L. Rebollo-Neyra, E. Sagianos. Backward-Optimized Orthogonal Matching Pursuit Approach. *IEEE Signal Proc. Letters*, **11**, 9, pp. 705-708, September 2004.

[19] G.Z. Zarabulut, L. Moura, D. Panario, A. Yongaoglu. Flexible Tree-Search Based Orthogonal Matching Pursuit Algorithm. *30th IEEE Int'l. Conf. on Acoustic, Speech, and Signal Proc.*, **IV**, pp. 673-676, Philadelphia, PA, 18-23 March 2005.

[20] C. La, M.N. Do. Tree-Based Orthogonal Matching Pursuit Algorithm for Signal Reconstruction. *13th IEEE Int'l. Conf. on Image Proc.*, Atlanta, GE, 8-11 October 2006.

[21] L. Mancera, J. Portilla. Image Dequantizing via Enforcing Sparseness in Overcomplete Representations. *7th Int'l. Conf. Adv. Concepts in Int. Vis. Syst. (ACIVS'05)*, LNCS-3708, pp. 411-8, Antwerp, Belgium, 20-23 September 2005.

[22] J.F. Claerbout, F. Muir. Robust Modelling of Erratic Data. *Geophysics*, **38**, 5, pp. 826-844, October 1973.

[23] J.A. Tropp. Just Relax: Convex Programming Methods for Identifying Sparse Signals in Noise. *IEEE Trans. Inf. Theory*, **52**, 3, March 2006.

[24] N.N. Abdelmalek. An Efficient Method for the Discrete Linear L1 Approximation Problem. *Math. Comp.*, **29**, 131, pp. 844-850, July 1975.

[25] S.S. Chen, D.L. Donoho, M.A. Saunders. Atomic Decomposition by Basis Pursuit. *SIAM J. of Sci. Comp.*, **20**, 1, pp. 33-61, 1999.

[26] E.J. Candes, M. Rudelson, T. Tao, R. Vershynin. Error Correction in Linear Programming. *46th Sympusium on Foundations of Computer Science (FOCS 2005)*, Pittsburgh, PA, 22-25 October 2005.

[27] D.L. Donoho, M. Elad, V.N. Temlyakov. Stable Recovery of Sparse Overcomplete Representations in the Presence of Noise. *IEEE Trans. on Inf. Theory*, **52**, 1, pp. 6-18, January 2006.

[28] D.L. Donoho, M. Elad. Optimally Sparse Representation in General (Non-Orthogonal) Dictionaries via via ℓ_1-norm Minimization. *Natl. Acad. Sci.*, USA, **100**, 5, pp. 2197-3002, 2003.

[29] D.L. Donoho. For Most Large Undetermined Systems of Linear Equations the Minimal ℓ_1-norm Solution is also the Sparsest Solution. *Technical Report*, "http://www-stat.stanford.edu/ donoho/Reports/", September 2004.

[30] O.G. Guleryuz. On Missing Data Prediction using Sparse Signal Models: A Comparison of Atomic Decomposition with Iterated Denoising. *Proc. of the SPIE*, 5914, 1G, San Diego, CA, August 2005.

[31] L. Mancera, J. Portilla. L0-norm-based Representation through Alternate Projections. *Proc. 13th IEEE Int'l. Conf. on Image Proc.*, pp. 2089-2092, Atlanta, GE, 8-11 October 2006.

[32] P. Moulin, J. Liu. Analysis of Multiresolution Image Denoising Schemes Using Generalized-Gaussian and Complexity Priors. *IEEE Trans. Image Theory*, **45**, pp. 909-919, 1999.

[33] S. Fischer, G. Cristobal. Minimum Entropy Transform using Gabor Wavelets for Image Compression. *11th Int'l. Conf. on Image Anal. and Proc. (ICIAP'01)*, Palermo, Italy, 26-28 September 2001.

[34] T.H. Reeves, N.G. Kingsbury. Overcomplete Image Coding using Iterative Projection-Based Noise Shaping. *Proc. 9th IEEE Int'l. Conf. on Image Proc.*, **3**, pp. 597-600, Rochester, NY, 23-25 September, 2002.

[35] N. Kingsbury, T. Reeves. Redundant Representation with Complex Wavelets: How to Achieve Sparsity, *10th IEEE Int'l. Conf. on Image Proc.*, **1**, pp. 45-48, Barcelona, Spain, 14-18 September 2003.

[36] O.G. Guleryuz: Nonlinear approximation based image recovery using adaptive sparse reconstructions and iterated denoising-part I: theory. *IEEE Transactions on Image Processing*, **15**, 3, pp. 539-554, March 2006.

[37] O.G. Guleryuz: Nonlinear approximation based image recovery using adaptive sparse reconstructions and iterated denoising-part II: adaptive algorithms. *IEEE Transactions on Image Processing*, **15**, 3, pp. 555-571, March 2006.

[38] R.M. Figueras i Ventura, E.P. Simoncelli. Statistically Driven Sparse Image Approximation. *14th IEEE Int'l. Conf. on Image Proc.*, San Antonio, TX, 16-19 September 2007.

[39] M.A.T. Figueiredo, J.M. Bioucas-Dias, R.D. Nowak. Majorization-Minimization Algorithms for Wavelet-Based Image Restoration. *Submitted*, 2007.

[40] J.L. Starck. Morphological Component Analysis. *Proc. of the SPIE*, **5914**, San Diego, CA, August 2005.

[41] N. Kingsbury. Complex Wavelets for Shift-Invariant Analysis and Filtering of Signals. *J. of App. and Comp. Harmonic Anal.*, **10**, 3, pp. 234-53, May 2001.

[42] D.C. Youla. Generalized Image Restoration by the Method of Alternating Orthogonal Projections. *IEEE Trans. on Circuits and Syst.*, **CAS-25**, 9, 1978.

[43] J.A. Tropp, I.S. Dhillon, R.W. Heath, T. Strohmer. Designing Structured Tight Frames Via An Alternating Projection Method. *IEEE Trans. on Inf. Theory*, **51**, 1, pp. 188-209, January 2005.

[44] E.J. Candes, D.L. Donoho. Curvelets: A Surprisingly Effective Nonadaptive Representation for Objects with Edges. *Curves and Surfaces IV*. P.J. Laurent ed., 1999.

[45] J.A. Guerrero-Colon, J. Portilla. Deblurring-by-Denoising using Spatially Adaptive Gaussian Scale Mixtures in Overcomplete Pyramids. *Proc. 13th IEEE Int'l. Conf. on Image Proc.*, Atlanta, GE, 8-11 October, 2007.

[46] M.J. Fadili, J.L. Starck. EM Algorithm for Sparse Representation-Based Image Inpainting. *Proc. 12th IEEE Int'l. Conf. on Image Proc.*, Genoa, Italy, 11-14 September, 2005.

[47] J. Portilla, L. Mancera. L0-based sparse approximations: Two alternative methods and some applications. *SPIE Optics & Photonics*, San Diego, CA, 26-30 August, 2007.

[48] M. Elad, P. Milanfar, R. Rubinstein. Analysis versus Synthesis in Signal Priors. *Inverse Problems*, **23**, pp. 947-968, 2007.

[49] G.H. Golub, C.F. Van Loan. Matrix Computations. *Johns Hopkins University Press*, Baltimore, MD, USA, pp. 694, 1996.

Luis Mancera received his MS Degree in Computer Science in 2001 from the Escuela Técnica Superior de Ingeniería Informática in the University of Granada, Spain. He worked for two years for the European Centre for Nuclear Research (CERN) in Geneva, Switzerland. Now he is carrying out his Ph.D. studies mainly centered in non-linear sparse image representation methods and their applications. He is a grant holder of a FPI contract at the Dept. of Computer Science and Artificial Intelligence, Universidad de Granada, Spain, funded by Ministerio de Educación y Ciencia, from the Spanish government.

Jose A. Guerrero-Colon was born in Málaga, Spain. He received his MS Degree in Computer Science in 2003 from the Escuela Técnica Superior de Ingeniería Informática, Universidad de Málaga, Spain. Currently his carrying out his Ph.D. studies and working in several projects related to his main topic: New visual-statistical models applied to the restoration of photographic images corrupted by blur and noise. He is a grant holder of FPU contract, funded by Ministerio de Educación y Ciencia, from the Spanish government, at the Dept. of Computer Science and Artificial Intelligence, Universidad de Granada, Spain.

Javier Portilla graduated (M.S.'94, Ph.D.'99) from E.T.S.I. Telecomunicación, in the Universidad Politécnica de Madrid. After a post-doctoral stay in New York University, he worked as an Assistant Professor at the Computer Science and Artificial Intelligence Department, in the Universidad de Granada, Spain. He obtained a "Ramón y Cajal" research position, which he still enjoys, now in the Instituto de Óptica, Consejo Superior de Investigaciones Científicas (C.S.I.C), back in Madrid. He is interested in modelling image statistics and their degradation sources, for designing Bayesian methods, and also on new non-linear image representation models inspired in visual neural science, and their application to image processing.

DT-CWT/Curv		Normalized ℓ_0-norm (K/N)		
Image	Method	3.05e-002	9.45e-002	2.91e-001
Barbara	DT+OP	27.08/25.58	30.83/29.41	37.82/34.96
	ℓ_0-FT	28.18/25.98	33.25/30.51	40.58/34.30
	ℓ_1-AP	26.39/25.75	30.21/31.02	38.65/38.38
	ℓ_0-AP	*29.23/28.61*	*33.38/34.29*	*41.76/41.51*
	ℓ_1-AP+OP	**29.95/29.10**	**34.12/34.94**	**43.09/43.08**
House	DT+OP	29.95/27.41	32.82/31.96	38.98/36.62
	ℓ_0-FT	31.19/28.64	34.76/32.69	40.25/37.77
	ℓ_1-AP	29.60/27.98	32.79/33.32	39.56/39.35
	ℓ_0-AP	*32.09/30.45*	*35.18/35.63*	*43.00/41.87*
	ℓ_1-AP+OP	**32.61/31.19**	**35.79/36.69**	**44.41/43.78**
Boat	DT+OP	24.86/23.42	28.20/26.38	34.65/31.59
	ℓ_0-FT	25.46/24.07	30.07/27.62	34.69/31.17
	ℓ_1-AP	24.10/23.41	27.53/27.00	34.56/33.62
	ℓ_0-AP	*26.48/26.15*	*30.43/29.97*	*38.00/36.73*
	ℓ_1-AP+OP	**26.92/26.19**	**31.05/30.24**	**39.08/38.02**
Lena	DT+OP	26.00/24.22	29.46/27.54	36.93/32.92
	ℓ_0-FT	27.09/24.86	*32.17*/28.63	39.38/32.54
	ℓ_1-AP	25.50/24.78	28.92/28.72	36.89/35.69
	ℓ_0-AP	*27.72/27.10*	31.63/*31.27*	*40.29/38.66*
	ℓ_1-AP+OP	**28.49/27.64**	**32.50/32.15**	**41.41/40.40**
Peppers	DT+OP	25.18/24.00	29.15/27.39	35.95/32.43
	ℓ_0-FT	25.79/24.53	*31.57*/28.41	38.43/32.28
	ℓ_1-AP	24.36/24.47	28.46/28.80	35.99/34.89
	ℓ_0-AP	*27.43/26.85*	31.47/*30.88*	*38.81/37.35*
	ℓ_1-AP+OP	**27.82/27.43**	**32.26/32.13**	**40.12/39.17**

TABLE I

DETAILED COMPARISON OF THE METHODS USING 8-SCALES DT-CWT AND 6-SCALES CURVELETS IN OUR SET OF STANDARD IMAGES. BOLD NUMBERS INDICATE THE METHOD PROVIDING THE BEST APPROXIMATION FOR EACH IMAGE AND SPARSENESS LEVEL. CURSIVE INDICATES THE SECOND BEST. ℓ_0-NORM IS EXPRESSED AS NUMBER OF NON-ZERO COEFFICIENTS NORMALIZED BY N. NORMALIZED COLUMNS CORRESPOND, RESPECTIVELY, TO 2001, 6189 AND 19096 ACTIVE COEFFICIENTS.

DT-CWT/Curv		Normalized ℓ_0-norm (K/N)		
Image	Method	5.85e-001	8.55e-001	1.49e+000
Barbara	DT+OP	43.62/39.23	45.89/41.99	52.42/46.13
	ℓ_0-FT	44.73/39.69	47.69/42.78	53.44/47.16
	ℓ_1-AP	45.39/44.16	50.24/50.20	>**100**/>**100**
	ℓ_0-AP	*48.31/47.37*	*51.76/53.00*	*61.73/64.86*
	ℓ_1-AP+OP	**52.03/51.33**	**56.88/61.48**	>**100**/>**100**
House	DT+OP	44.65/39.80	47.29/43.23	53.40/46.92
	ℓ_0-FT	45.85/42.02	50.01/43.95	55.98/47.94
	ℓ_1-AP	46.22/43.78	50.09/50.94	>**100**/>**100**
	ℓ_0-AP	*50.92/46.52*	*54.56/53.96*	*67.38/63.13*
	ℓ_1-AP+OP	**53.18/49.17**	**57.05/60.94**	>**100**/>**100**
Boat	DT+OP	39.27/35.00	42.26/37.51	48.59/41.82
	ℓ_0-FT	40.61/36.07	44.29/38.27	52.50/42.61
	ℓ_1-AP	40.71/38.73	45.86/45.29	>**100**/57.46
	ℓ_0-AP	*45.50/42.54*	*50.22/49.66*	*63.29/58.70*
	ℓ_1-AP+OP	**47.76/45.14**	**52.90/55.64**	>**100**/**71.97**
Lena	DT+OP	42.32/36.96	44.61/40.12	50.68/44.36
	ℓ_0-FT	43.65/38.07	46.15/41.61	51.57/45.47
	ℓ_1-AP	43.71/41.30	48.14/49.70	>**100**/>**100**
	ℓ_0-AP	*47.54/44.74*	*51.16/51.02*	*61.60/60.94*
	ℓ_1-AP+OP	**50.67/47.47**	**55.09/58.18**	>**100**/>**100**
Peppers	DT+OP	40.77/36.15	44.00/38.97	50.88/43.22
	ℓ_0-FT	42.38/37.22	45.51/40.31	51.60/44.74
	ℓ_1-AP	41.92/39.75	49.84/46.17	>**100**/58.62
	ℓ_0-AP	*45.76/43.04*	*52.13/50.40*	*63.78/61.45*
	ℓ_1-AP+OP	**48.12/45.66**	**56.05/56.30**	>**100**/**72.91**

TABLE II

CONTINUATION OF TABLE I. NORMALIZED COLUMNS CORRESPOND, RESPECTIVELY, TO 38342, 56048 AND 97471 ACTIVE

COEFFICIENTS.

	Iters.	
Meth.	DT-CWT	Curv.
ℓ_0-FT	180	231
DT+OP	188	174
ℓ_1-AP	263	360
ℓ_1-AP+OP	333	440
ℓ_0-AP	495	920

TABLE III

AVERAGED NUMBER OF ITERATIONS EXECUTED IN OUR TEST SET USING 8-SCALES DT-CWT AND 6-SCALES CURVELETS.