

UNIVERSIDAD DE GRANADA  
ESCUELA TÉCNICA SUPERIOR DE INGENIERÍAS  
INFORMÁTICA Y DE TELECOMUNICACIÓN

**Sparse approximation to natural  
images based on non-convex  
optimisation and applications to  
restoration**

**Ph.D. Dissertation**

Luis Mancera Pascual  
Computer Science Engineer

2008



DEPARTMENT OF COMPUTER SCIENCE AND  
ARTIFICIAL INTELLIGENCE  
ESCUELA TÉCNICA SUPERIOR DE INGENIERÍAS INFORMÁTICA  
Y DE TELECOMUNICACIÓN

**Sparse approximation to natural  
images based on non-convex  
optimisation and applications to  
restoration**

Ph.D. Dissertation

**Author:** Luis Mancera Pascual  
*Computer Science Engineer*

**Supervisor:** Francisco Javier de la Portilla Muelas  
*Telecommunications Engineer, Ph. D.*

2008  
Last update: 4 March



Title:  
Sparse approximation to natural images based on non-convex  
optimisation and applications to restoration

Author:  
Luis Mancera Pascual

**Committee:**

President : Rafael Molina Soriano

Members : Nick G. Kingsbury  
Carlos García Puntónet  
Jesús Malo Lpez

Secretary : Jesús Chamorro Martínez

Substitutes : Luis Salgado Álvarez de Sotomayor  
Gabriel Cristóbal Prez

Agree to evaluate this dissertation as: Sobresaliente Cum Laude

Granada, 11th February 2008



*To my father*





At page 59, vol. I, we find this sentence – "He was advancing by the only road that was ever traveled by the stranger as he approached the Hut; or, he came up the valley." This is merely a vagueness of speech. [...] The whole would be clearer thus – "He was advancing by the valley – the only road traveled by a stranger approaching the Hut." We have here sixteen words, instead of Mr. Cooper's twenty-five.

*E.A. Poe – Comments on F. Cooper's "Wyandotte".*

... and this round gold is but the image of the rounder globe, which, like a magician's glass, to each and every man in turn but mirrors back his own mysterious self. Great pains, small gains for those who ask the world to solve them; it cannot solve itself.

*H. Melville – Moby Dick or the whale, ch. 99, "The Doubloon"*



# Acknowledgments

First, I want to effusively thank Javier Portilla for his great commitment with this Thesis, and his huge effort to make a scientist out of me. Nobody better fitted for this task. I want to thank also Rafael Molina, who I admire as a professor, researcher and person, and whose proximity and advices I value most.

Jose A. Guerrero-Colón has been my *brother in arms* during these years. Without Jesús Chamorro this Thesis never could had come to light. I owe him the passion for image processing and, together with Javier, the opportunity to work on this field. Also, Joaquín Valdivia and Juan Luis Castro have always been an example to me, and their teaching and advices I treasure as if they were gold dust. I thank the Visual Information Processing group and the whole Dep. of Computer Science and Artificial Intelligence of the University of Granada for the indelible years I have passed in it.

I also thank Nick Kingsbury for hosting me in the University of Cambridge with open arms, and the Signal Processing group of the Engineering Department, where I have passed such a good and profitable time. I also want to thank Mario Figueiredo, Nick Kingsbury and Michael Elad their comments (and criticism) on our work.

Thanks to Ignacio Requena and Andrés Cano for helping me with the bureaucracy involved in writing and submitting a Thesis.

Thanks, above all else, to Sonia and Luis. They give a meaning to everything I do. Thanks also to my family.

*This work has been funded by programs TIC-2003-01504 and TEC2006/13845/TCM from the Spanish Ministerio de Educación y Ciencia.*



# Summary

In the modern world, current needs have provoked a fast increase in the interest in image processing applications. In the beginning, these applications were approximated with heuristics, using techniques particularly adapted to each case. Nowadays, it is increasingly important to develop good image models allowing a generic application to a wide variety of tasks. Our brain is able to discriminate the relevant information in a distorted image, because the objects of the world have a typical structure. This structure is reflected in natural images (those representing the real world we are living in), whilst random images do not have, in general, any structure at all. If we want to replicate this behaviour in a machine, it is very important to have a good *a priori* knowledge about the typical structure of natural images.

Inspired on what we know about the processing of visual stimuli in our brain, we want to represent images with as few samples as possible, making easier not only to describe them statistically, but also their capture, processing and storage. The ability to express images with few elements can be considerably increased if we transform the pixels to a new redundant domain, where there are more coefficients than pixels in the image. Given a vector and a set of other vectors defining a redundant domain, the sparse approximation problem is formulated as minimizing a certain measurement of the error when expressing the vector as a linear combination of a given number of (unknown) vectors from the given set. Because of the inherent complexity of this problem, most approximations have been traditionally based on greedy heuristics or convex relaxation of the cost function. In addition, much effort has been made to find the theoretical conditions under which these two types of approximations find the global optimum to the sparse approximation problem. A third class of methods exists based on iterative thresholding. They have been shown to be more efficient and provide better results than greedy heuristics or classic optimization methods. However, some of their more successful variants are still not well grounded in theory, and their performance when applied to restoration has not been yet extensively studied. In this Thesis, two methods of this kind

are examined. They had been already proposed as heuristics, but never derived as solutions to an optimization problem. This Thesis shows that it is possible to apply classical optimization tools to obtain useful (though suboptimal) solutions directly to the sparse approximation problem.

The first proposed method minimizes the approximation error given a value of a determined  $\ell_p$ -norm of the representation. Its convergence is proven by describing it as based on alternated orthogonal projections between two sets. We call this method  $\ell_p$ -AP. We focus on the cases  $p = 0$ , where sub-optimal solutions are found, and  $p = 1$ , where the global optimum is achieved. The emphasis of our experiments is on analysing the behaviour of the methods in practical image processing conditions, by means of intensive experiments. We show that  $\ell_0$ -AP is superior to  $\ell_1$ -AP and greedy methods. The second method is derived as a gradient descent in successively decreasingly smoothed versions of a continuous, constrained function which is equivalent to the (discontinuous, unconstrained) cost function of the sparse approximation problem. We call this method  $\ell_0$ -GM. We show that  $\ell_0$ -GM outperforms  $\ell_0$ -AP, with performance close to the state-of-the-art in sparse approximation. We also show a convex version of this method, which we call  $\ell_1$ -GM.

Last but not least, we have adapted the proposed methods to be applied to several restoration problems. We propose to use  $\ell_0$ -AP for removing spatial quantization artifacts and  $\ell_0$ -GM for interpolation of lost pixels or missing regions of the image, for the interpolation of color images from mosaics and finally for increasing the detail level or super-resolving images. We show that the proposed methods provide good results to these problems, being superior or similar to other methods chosen as representatives from current state-of-the-art.

# Contents

Figure index	xix
Table index	xxvii
<b>1 Introduction</b>	<b>1</b>
1.1 Introduction and objectives . . . . .	1
1.2 Contribution of this Thesis . . . . .	7
<b>2 The Sparse Approximation Problem</b>	<b>11</b>
2.1 Analysis-based sparseness vs. Synthesis-based sparseness . .	11
2.2 Formulation of the sparse approximation problem . . . . .	14
2.3 The sparse approximation problem in the literature . . . . .	16
2.3.1 Greedy heuristics . . . . .	17
2.3.2 Convex relaxation problem and <i>Basis Pursuit</i> . . . . .	19
2.3.3 Iterative shrinkage . . . . .	20
2.4 Equivalence conditions when minimising $\ell_1$ and $\ell_0$ -norms . .	23
<b>3 Sparse approximation using alternating projections</b>	<b>27</b>
3.1 $\ell_p$ -AP method . . . . .	28
3.1.1 $\ell_0$ -AP . . . . .	29
3.1.2 $\ell_1$ -AP . . . . .	32
3.2 Mean square error minimisation for a given selection of coefficients . . . . .	37
3.3 Implementation . . . . .	38
3.3.1 Representations . . . . .	38
3.3.2 Convergence and stopping criterion . . . . .	39
3.4 Results and discussion . . . . .	40
3.4.1 Some previous methods . . . . .	40
3.4.2 Comparison of $\ell_0$ -AP, $\ell_1$ -AP and previous methods .	41
3.4.3 Computational load . . . . .	46
3.5 Conclusions . . . . .	47

<b>4</b>	<b>Sparse approximation using gradient descent</b>	<b>49</b>
4.1	An alternative formulation with a continuous cost function .	50
4.2	Local minimisation with $\ell_0$ -norm: IHT . . . . .	52
4.3	Global minimisation with $\ell_0$ -norm: $\ell_0$ -GM . . . . .	53
4.3.1	Using a single solution for all the sparseness levels . .	58
4.4	Implementation . . . . .	61
4.5	Results and discussion for $\ell_0$ -GM . . . . .	61
4.6	Gradient descent for minimisation of $\ell_1$ -norm: IST & $\ell_1$ -GM	63
4.6.1	Alternative formulation of the convex cost function .	63
4.6.2	Cost function minimisation with a fixed threshold: IST	67
4.6.3	A more efficient convex minimisation: $\ell_1$ -GM . . . .	69
4.6.4	Practical advantages of $\ell_1$ -GM . . . . .	69
4.7	Conclusions . . . . .	70
<b>5</b>	<b>Application to image restoration</b>	<b>73</b>
5.1	Consistency with an observation . . . . .	75
5.2	Formulation using synthesis-sense sparseness . . . . .	76
5.3	Estimation using $\ell_p$ -AP and synthesis-sense sparseness . . .	77
5.4	Formulation using analysis-sense sparseness . . . . .	79
5.5	Estimation using $\ell_p$ -GM and analysis-sense sparseness . . . .	79
<b>6</b>	<b>Some applications</b>	<b>81</b>
6.1	Removing quantisation artifacts . . . . .	81
6.1.1	Introduction . . . . .	81
6.1.2	Consistency set . . . . .	82
6.1.3	Implementation . . . . .	83
6.1.4	Results and discussion . . . . .	84
6.1.5	Conclusions . . . . .	90
6.2	Interpolation of missing pixels . . . . .	91
6.2.1	Introduction . . . . .	91
6.2.2	Consistency set . . . . .	91
6.2.3	$\ell_0$ -AP: new strategy for searching the radius . . . . .	92
6.2.4	Implementation . . . . .	93
6.2.5	Results and discussion . . . . .	94
6.2.6	Conclusions . . . . .	98
6.3	Spatial-chromatic interpolation in digital camera mosaics . .	100
6.3.1	Introduction . . . . .	100
6.3.2	Consistency set . . . . .	101
6.3.3	Additional constraint increasing the spatial-chromatic correlation . . . . .	101
6.3.4	Implementation . . . . .	102
6.3.5	Results and discussion . . . . .	103



---

6.3.6	Conclusions . . . . .	104
6.4	Detail increase . . . . .	105
6.4.1	Introduction . . . . .	105
6.4.2	Consistency set . . . . .	109
6.4.3	Implementation . . . . .	109
6.4.4	Results and discussion . . . . .	109
6.4.5	Conclusions . . . . .	110
<b>7</b>	<b>Conclusions and future work</b>	<b>113</b>
7.1	Conclusions . . . . .	113
7.2	Future work . . . . .	115
	<b>Appendices</b>	<b>117</b>
<b>A</b>	<b>Test images set</b>	<b>119</b>
<b>B</b>	<b>Minimisation of the quadratic error of the reconstruction with a given support.</b>	<b>121</b>
B.1	First case: incomplete subset . . . . .	121
B.2	Second case: complete subset . . . . .	122
<b>C</b>	<b>A Parseval frame formed concatenating two Parseval frames</b>	<b>125</b>
<b>D</b>	<b>Publications</b>	<b>127</b>
	<b>Bibliography</b>	<b>129</b>



# List of Figures

1.1	<b>Top left</b> , <i>Einstein</i> standard image. <b>Top right</b> , random image (white Gaussian noise). <b>Bottom</b> , sum of the two images above. . . . .	3
1.2	<b>Left</b> , image obtained using the 10% of pixels with largest deviation with respect to the global mean of the <i>Einstein</i> image. <b>Right</b> , image built using the 10% of the coefficients with largest amplitude in the Fourier representation of the same image. . . . .	4
1.3	<b>Left</b> , sub-band of the representation of a natural image under a wavelet-like (DT-CWT) filter bank. Dark pixels represent high amplitude coefficients and light ones those with low amplitude. <b>Right</b> , same sub-band of a sparse approximation to the natural image. . . . .	5
2.1	Sparse approximation results comparison for several representations. Data have been obtained using the largest in amplitude coefficients of the linear responses and then averaging in our test set. Performance is measured in terms of PSNR given a number of coefficients. This number is normalised by the total number of pixels in the image. . . .	13

2.2	<b>Top-left</b> , highest-frequency sub-band of analysis vector of <i>Peppers</i> image using DT-CWT with 8 scales. Light and dark points correspond, respectively, with low and high amplitudes of the coefficients. The size of the sub-band has been doubled in both directions through replication of rows and columns in order to match the size of the image, and then it has been cropped to $64 \times 64$ for visibility. <b>Top-centre</b> , same sub-band, but this time non-linearly obtained using the $\ell_0$ -AP method (see Chapter 3). <b>Top-right</b> , result of applying a threshold in amplitude to the result in central panel (preserving 7 times less coefficients than pixels in the image). <b>Bottom-left</b> , original image, which is perfectly reconstructed by the representations corresponding to the left and central panels of the top row. <b>Bottom-right</b> , approximation obtained with the sparse coefficients corresponding to the top-right panel (35.67 dB). . . . .	14
2.3	One-dimensional $p$ -th power of the $\ell_p$ -norm for different $p$ values. . . . .	16
3.1	<b>Top</b> , graphical explanation of the $\ell_0$ -AP method. <b>Bottom</b> , same for $\ell_1$ -AP. Only a face of the ball is shown for clarity. .	31
3.2	Logarithmic plot of the approximation quality (PSNR, in dB) vs. the number of iterations for $\ell_0$ -AP for three images and two sparseness levels. The representation used here is DT-CWT. The number at the end of the curves is the PSNR at convergence. The numbers accompanying the tangency point (indicated by the intersection with the dotted curves) are the PSNR and the number of iterations obtained when the stopping criterion is reached. . . . .	32
3.3	Convergence curves in semi-logarithmic scale for $\ell_1$ -AP, using three images and two sparseness levels. Details are similar to Figure 3.2. It is also indicated the $\ell_0$ -norm, normalised by $N$ , of the solution at convergence. . . . .	37
3.4	Averaged compaction results (fidelity of the approximation as PSNR, in dB) for our test set using StOMP, DT+OP, IHT and IST. <b>Top</b> , using 8-scale DT-CWT. <b>Bottom</b> , using 6-scale Curvelets. . . . .	42
3.5	Compaction results, averaged in our test set, of methods $\ell_0$ -AP, $\ell_1$ -AP, $\ell_1$ -AP+OP, IHT and DT+OP. <b>Top</b> , using 8-scale DT-CWT. <b>Bottom</b> , using 6-scale Curvelets. . . . .	43

3.6	Visual comparison of the methods using $0.0765 \cdot N$ Curvelets coefficients and the <i>Einstein</i> image, , where $N$ is the number of pixels in the image. Results are cropped to $128 \times 128$ , starting at pixel (71, 41), to improved the visibility. <b>Left column</b> , from top to bottom: original image and results of $\ell_1$ -AP (30.85 dB) and $\ell_1$ -AP+OP (33.52 dB). <b>Right column</b> , from top to bottom: results from StOMP (28.66 dB), IHT (29.10 dB) and $\ell_0$ -AP (32.98 dB). . . . .	46
4.1	Bold line shows the minimum between $y(x) = 1$ (dashed) and $y(x) = x^2$ (dotted). . . . .	52
4.2	<b>Top</b> , IHT convergence curves using a low threshold ( $\theta = 5$ ) and three different $\alpha$ values. We have used <i>House</i> image and 8-scale DT-CWT. <b>Bottom</b> , same result for a higher threshold ( $\theta = 60$ ). . . . .	54
4.3	1-D smoothing function: an inverted parabola in the interval $[-1, 1]$ , centred at 0 and with maximum 1. Outside that interval is 0. . . . .	55
4.4	1-D Function with multiple minima progressively smoothed until obtaining only one. The black continuous line indicates the path joining the global optima through the scale of the smoothing kernel. We have used here as smoothing kernel a normalised (in area) version of $h(x)$ (See Figure 4.3). . . . .	58
4.5	Fidelity-sparseness results of $\ell_0$ -GM, using $\beta = 0.9$ (circles, $1.5 \cdot 10^2$ iterations) and $\beta = 0.99$ (solid, $1.5 \cdot 10^3$ iterations), compared to IHT, using several thresholds (dashed, $10^5$ iterations). We use <i>House</i> image and DT-CWT with 8-scales. . . . .	59
4.6	<b>Top</b> , sparse approximation fidelity averaged in our test set using $\ell_0$ -GM with $\alpha = 1.85(\theta^2/2)$ , three different $\beta$ values and 8-scale DT-CWT. <b>Bottom</b> , quality of the reconstruction from the highest amplitude coefficients of the vector obtained using $\ell_0$ -GM for a very high $\lambda$ value (very low sparseness), and the same $\beta$ values. Dotted curves correspond to that of top panel. The vertical axis has been re-scaled to improve visibility. . . . .	60
4.7	$\ell_0$ -GM sparse approximation results averaged in our test set compared to other methods previously seen (StOMP, IHT, $\ell_0$ -AP and $\ell_1$ -AP+OP). . . . .	62

- 4.8  $64 \times 64$  crop of the reconstruction of *Einstein* image using  $0.04 \cdot N$  (2605) active DT-CWT coefficients, for several sparse approximation methods. **Top-left**, result of StOMP, implemented as described in Section 3.4.1 (28.98 dB). **Top-right**, IHT (31.20 dB). **Centre-left**,  $\ell_1$ -AP (29.70 dB). **Centre-right**,  $\ell_0$ -AP (31.97 dB). **Bottom-left**,  $\ell_1$ -AP+OP (32.38 dB). **Bottom-right**,  $\ell_0$ -GM (33.28 dB). . . . . 65
- 4.9 **Top**, convergence curves for IST with a low threshold ( $\theta = 5$ ) and three different  $\alpha$  values. We have used *House* image and 8-scale DT-CWT. **Bottom**, same result for a higher threshold ( $\theta = 60$ ). . . . . 68
- 4.10 Fidelity-sparseness results of  $\ell_0$ -GM, using  $\beta = 0.9$  (circles,  $1.5 \cdot 10^2$  iterations) and  $\beta = 0.99$  (solid,  $1.5 \cdot 10^3$  iterations), compared to IHT, using several thresholds (dashed,  $10^3$  iterations). We use *House* image and DT-CWT with 8-scales. 70
- 4.11 Averaged sparse approximation results in the test set using  $\ell_1$ -GM with  $\alpha = 1.85\theta$ , different  $\beta$  values and using DT-CWT with 8 scales. We also show the result of  $\ell_1$ -AP. . . . . 71
- 4.12 Iterations needed to provide nearly optimal result for different sparseness level using  $\ell_1$ -AP and  $\ell_1$ -GM ( $\alpha = 1.85\theta$ ,  $\beta = 0.99$ ). We use *Barbara* image and 8-scale DT-CWT. . . 72
- 5.1 **Top-left**, *Peppers* crop, starting at row 111, column 91. **Bottom-left**, same crop of the high frequency sub-band of the linear response to *Peppers* using 8-scales DT-CWT, corresponding to orientation  $-45$ . We have previously doubled the size of this sub-band, through pixel replication, in order to match the image size. **Top-right**, degraded image by setting to zero, randomly, 40% of the pixels. **Bottom-right**, corresponding sub-band. . . . . 74
- 6.1 Example of application of  $\ell_1$ -AP and  $\ell_0$ -AP to de-quantizing. **Top-left**, original *Einstein* image, cropped to  $128 \times 128$  pixels. **Top-right**, 3-bits observed quantisation (PSNR: 27.98 dB). **Centre-left**,  $\ell_1$ -AP result using 8-scale DT-CWT (30.17 dB). **Centre-right**,  $\ell_1$ -AP result using 6-scale Curvelets (30.61 dB). **Bottom-left**,  $\ell_0$ -AP result using 8-scale DT-CWT (31.21 dB). **Bottom-right**,  $\ell_0$ -AP result using 6-scale Curvelets (31.38 dB). . . . . 86

- 6.2 Example of application of  $\ell_1$ -AP and  $\ell_0$ -AP to de-quantizing. **Top-left**, original *Peppers* image, cropped to  $128 \times 128$ . **Top-right**, 3-bits observed quantisation (PSNR: 28.81 dB). **Centre-left**,  $\ell_1$ -AP result using 8-scale DT-CWT (29.08 dB). **Centre-right**,  $\ell_1$ -AP result using 6-scale Curvelets (29.50 dB). **Bottom-left**,  $\ell_0$ -AP result using 8-scale DT-CWT (31.06 dB). **Bottom-right**,  $\ell_0$ -AP result using 6-scale Curvelets (30.85 dB). . . . . 87
- 6.3 **Top-left**, *Einstein* quantised with 3 bits and cropped to  $128 \times 128$ , (PSNR: 27.98 dB). **Top-right**, RCIR result (30.39 dB). **Centre-left**, CD result (30.44 dB). **Centre-right**, DT+OP result using DT-CWT with 8 scales (30.72 dB). **Bottom-left**,  $\ell_0$ -AP result using DT-CWT with 8 scales (31.21 dB). **Bottom-right**,  $\ell_0$ -AP result using jointly 8-scale DT-CWT and 6-scale Curvelets, with equal scale factor,  $\sqrt{\frac{1}{2}}$  (31.93 dB). . . . . 88
- 6.4 **Top-left**, *Peppers* quantised with 3 bits and cropped to  $128 \times 128$ , (PSNR: 28.81 dB). **Top-right**, RCIR result (29.65 dB). **Centre-left**, CD result (29.85 dB). **Centre-right**, DT+OP result using DT-CWT with 8 scales (30.38 dB). **Bottom-left**,  $\ell_0$ -AP result using 8-scale DT-CWT (31.07 dB). **Bottom-right**,  $\ell_0$ -AP result using jointly 8-scale DT-CWT and 6-scale Curvelets, with equal scale factor,  $\sqrt{\frac{1}{2}}$  (31.46 dB). . . . . 89
- 6.5 **Left**, detail of the sky of a photographic 8-bits image with contrast amplified approximately 40 times. **Right**, same detail after processing with  $\ell_0$ -AP. . . . . 90
- 6.6 **Bold line - left axis**, Mean Square Error of the estimated pixels normalised to the minimum value of this curve, in ordinates, and to the value for which this minimum occurs, in abscissas. **Dashed line - right axis**, Normalised Mean Square Value of the estimated pixels. **Dotted line**, typical deviation for each value in the horizontal axis. **Dashed - dotted line**, Mean Square Value of the original pixels in the missing positions. All curves are averaged in our test set, using a random mask where approximately 40% of the pixels are lost. . . . . 93

- 6.7 Visual interpolation example of randomly missing pixels. **Top-left**, *Barbara* image, cropped to  $128 \times 128$ . **Top-right**, missing  $\approx 80\%$  of the pixels and filling them with the global mean (PSNR: 14.75 dB). **Centre-left**, interpolation made by  $\ell_1$ -GM (23.26 dB). **Centre-right**, result from *Fast-inpainting* (24.84 dB). **Bottom-left**,  $\ell_0$ -GM result using Curvelets with 6 scales (25.19 dB) **Bottom-right**, interpolation made by  $\ell_0$ -GM combining 6-scale Curvelets and LDCT with block size  $32 \times 32$ , and equal scale factors,  $\sqrt{0.5}$  (25.65 dB). . . . . 96
- 6.8 **Top-left**, *Barbara* image where value of missing pixels is the global mean of the observed ones (PSNR: 24.19 dB). **Top-right**, *Fast-inpainting* result (32.71 dB). **Bottom - left**, *EM-inpainting* result using 6-scale Curvelets and LDCT with block size  $32 \times 32$  (34.14 dB). **Bottom-right**,  $\ell_0$ -GM result using 6-scale Curvelets (34.92 dB). . . . . 97
- 6.9 **Left**, detail of the result of *EM-inpainting* shown in Figure 6.8 (PSNR: 34.38 dB). **Right**, same for the method  $\ell_0$ -GM (35.13 dB). . . . . 98
- 6.10 **Top-left**, damaged photographic image. **Top-right**, *Fast-inpainting* result. **Bottom-left**, *EM-inpainting* result using Curvelets and LDCT. **Bottom-right**, our result using  $\ell_0$ -GM with 6-scale Curvelets and LDCT using  $32 \times 32$  blocks and both scale factors to  $\sqrt{0.5}$ . . . . . 99
- 6.11 **Left**, detail of the result of *EM-inpainting* shown in Figure 6.10. **Right**, same for  $\ell_0$ -GM. . . . . 99
- 6.12 Visual example of the comparison between  $\ell_1$ -GM and  $\ell_0$ -GM applied to de-mosaicing. **Top-left**,  $64 \times 64$  detail of image 15 of *Eastman Kodak* database. **Top-right**, result of de-mosaicing a Bayer mosaic with patter 'GB' using  $\ell_1$ -GM with DT-CWT. PSNR for channels R, G and B is 37.24 dB, 39.87 dB and 37.17 dB. **Bottom-left**,  $\ell_0$ -GM result using Curvelets (39.29, 42.27 and 38.20 dB). **Top-right**,  $\ell_0$ -GM result using DT-CWT (39.59, 41.99 and 39.07 dB). . . . . 104
- 6.13 Visual comparison between de-mosaicing methods. **Top-left**, result of method in [145] with the image 15 of *Eastman Kodak* database. We have cropped the image to  $64 \times 64$  to improve the visibility of artifacts. PSNR of R,G and B channels is, respectively, 39.81, 41.57 and 39.09 dB. **Top-right**, [144] result (39.89, 43.87 and 39.26 dB). **Bottom-left**, [146] result (40.39, 44.14 and 39.73 dB). **Bottom-right**,  $\ell_0$ -GM result (39.59, 41.99, 39.07 dB). . . . . 108



---

6.14	<b>Top-left</b> , nearest neighbour interpolation (replication) of the resizing of the sub-sampling of the $2 \times 2$ averaged blocks <i>House</i> image (PSNR: 30.52 dB). We have cropped to $128 \times 128$ to improve visibility. <b>Top-right</b> , result of bilinear interpolation (29.65 dB). <b>Bottom-left</b> , result of $\ell_1$ -GM (33.50 dB). <b>Bottom-right</b> , result of $\ell_0$ -GM (33.53 dB).	111
A.1	Test set of images used in this Thesis. . . . .	120



# List of Tables

3.1	.....	44
3.2	.....	45
3.3	.....	47
4.1	.....	63
4.2	.....	64
6.1	.....	85
6.2	.....	95
6.3	.....	95
6.4	.....	105
6.5	.....	106
6.6	.....	107
6.7	.....	110



# Chapter 1

## Introduction

### 1.1 Introduction and objectives

Human brain has adapted throughout a long evolution, as well as during our personal development, to efficiently deal with visual stimuli [1]. So, there is a strong connection between the physical origin of those stimuli and the human visual system structure. In addition, vision is arguably the most powerful among human senses, in terms of the amount of information that it can acquire and process by time unit.

On the other hand, we have always had the need to transmit information to other people. To do it, we count on some tools limited to the precise place and moment where the message is emitted, as phonetic language. However, we also need to communicate to larger number of people, even though they are not present at the moment of emitting the message. This achievement was made first, and not by chance, through visual stimuli, as idiographic language or paintings. Some examples are Palaeolithic art and symbolic writings.

To obtain an image from the world around consists of projecting the three dimensional space we are living in onto a two dimensional surface, reproducing the shape of the objects and their details. Under the term *natural image* we denote those images captured, typically through photography, from the real world. This means that they are similar to the visual information usually captured by our eyes. Surprisingly enough, for the message to be interpreted properly, it is not necessary that those images are perfect projections of the real world. Indeed, our visual system is able to detect and recognise the represented objects even if they are distorted, up to a certain degree, naturally. This is a characteristic skill of very advanced visual systems which exploit in a massive way this kind of information.

Nowadays, we are witnessing an unprecedented technological revolution.

We are increasingly processing more and more information, which has to be transmitted to more and more people. As a natural consequence of the dominant role of vision in our perception, one of the areas more affected by this revolution is that of digital images. In the last years, techniques for capturing, processing, transmitting and storing these images have developed beyond expectations. In fact, digital images have already substituted analogue ones as the main representation vehicle. Thanks to the vast possibilities of digital technology, the image manipulation tools have exponentially increased. We keep on demanding more quality without compromising speed, and, therefore, a new effort is needed for improving the capture and the posterior processing. In addition, because of the greater importance of digital communications, it is also increasingly important to save bandwidth, and so it is sought to maximise the visual quality for a given information support (number of bits).

In the beginning, problems such as image coding for compression, enhancement, removing noise and annoying artifacts, lost information recovery, pattern recognition, etc., were approached heuristically, using more or less *ad-hoc* techniques. However, now there is no doubt about the importance of developing good image models allowing a more generic application to a wide variety of tasks.

Most of these applications are related to human vision. These kinds of tasks are carried out continuously in our brain. Therefore, to develop a good model it is convenient to pose the following question. How can our visual system discriminate the relevant information in a distorted image?

Obviously, an arbitrary image does not represent objects of our world, which have a typical structure allowing us to recognise them. This structure is reflected in natural images [1, 2], which consist, typically, of localised oriented features (edges, lines, corners, etc.) and relatively large smoothly varying areas, possibly with some texture in them. Left superior panel in the Figure 1.1 is an example of a typical natural image. On the other hand, random images, as that seen in the right superior panel in the Figure 1.1, do not have, in general, any structure at all. Due to the huge amount of visual stimuli processed, and learned to process through evolution, our brain is able to distinguish very clearly the original image underlying a degraded version of itself (see inferior panel in the Figure 1.1, formed as an additive mixture of the two panels above). Thus, if we want to automate this behaviour, it is very important to have good *a priori* knowledge of the typical structure of natural images, as many authors have pointed before (e.g., [3, 1, 4, 5]).

One of the criteria usually considered when evaluating the efficiency of a neural system is the maximisation of the ratio between the amount of information and the number of neurons required to represent it [5]. In a

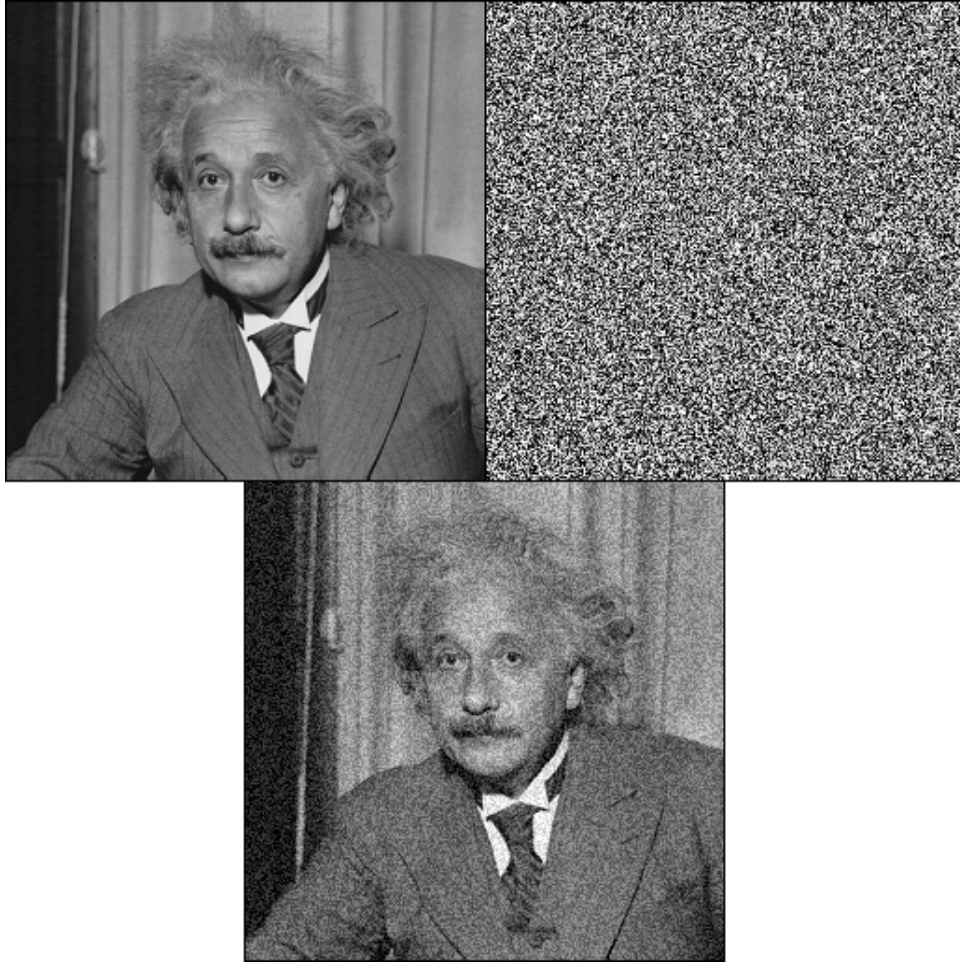


Figure 1.1: **Top left**, Einstein *standard image*. **Top right**, *random image (white Gaussian noise)*. **Bottom**, *sum of the two images above*.

similar way, we can pose the same goal when we deal with natural images. If we manage to represent them with as few numbers as possible, it will be easier not only to store them, but also to get more powerful statistical descriptions. As a consequence, we will also increase the performance in restoration tasks.

Unfortunately, despite their typical structure, the large size of the set of natural images and the strong statistical dependence between neighbouring pixels make the modeling too complex to be done in the pixel domain. The ability to accumulate information in few elements can be considerably powered by transforming the image from pixels to new domains. Let us come back to the example shown in the left panel of Figure 1.1. If we only take the 10% of the pixels with largest deviation with respect to the global

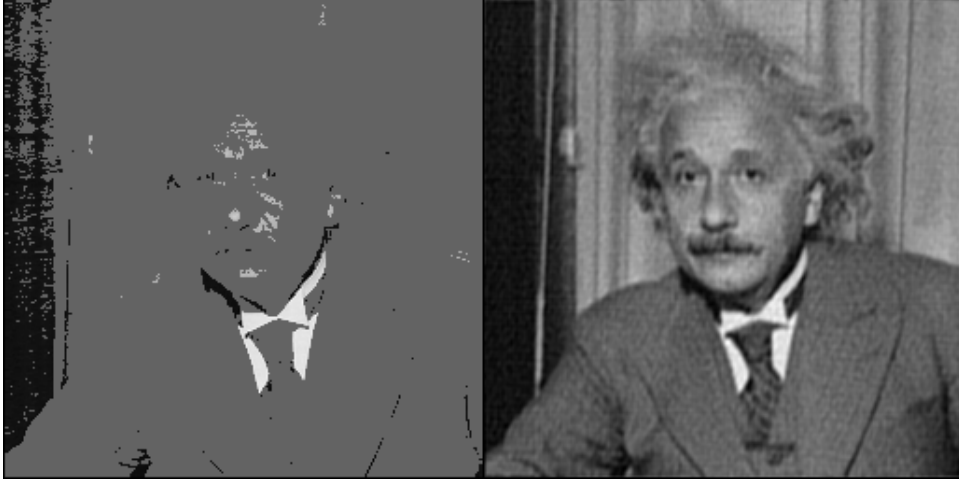


Figure 1.2: **Left**, image obtained using the 10% of pixels with largest deviation with respect to the global mean of the Einstein image. **Right**, image built using the 10% of the coefficients with largest amplitude in the Fourier representation of the same image.

mean, we obtain the image in the left panel of Figure 1.2, where, as we can see, most of the features of the original image have been removed. On the other hand, right panel of Figure 1.2 shows the image made by the 10% of the frequencies in the Fourier domain whose coefficients are largest in amplitude. Although obtained using the same number of coefficients, the latter image is much closer to the original than the former one, both in objective and subjective terms.

This property also makes easier, as we have already mentioned, the statistical description of natural images. For example, they typically have large areas with smooth texture, and, thus, the energy of the Fourier representation is usually concentrated in the low frequencies. When degrading an image, if every frequency is uniformly corrupted (e.g., white Gaussian noise) then dominant frequencies are relatively less affected. Intuitively, we see that, by removing the lower amplitude frequencies, relatively very affected but with little significance to reconstruct the original image, we would greatly reduce the amount of noise in the observation, while maintaining a high fidelity to the original image. As a consequence, when most of the energy is concentrated in few coefficients, removing those with low amplitude will result in more important reduction of noise in the image than when the energy is more uniformly distributed.

The Oxford English Dictionary defines *sparseness* as the property of being thinly dispersed. The Cambridge Advanced Learner's Dictionary defines it as the property of being small in numbers or amount, often scattered over a large area. In this Thesis, we have made an extended use of



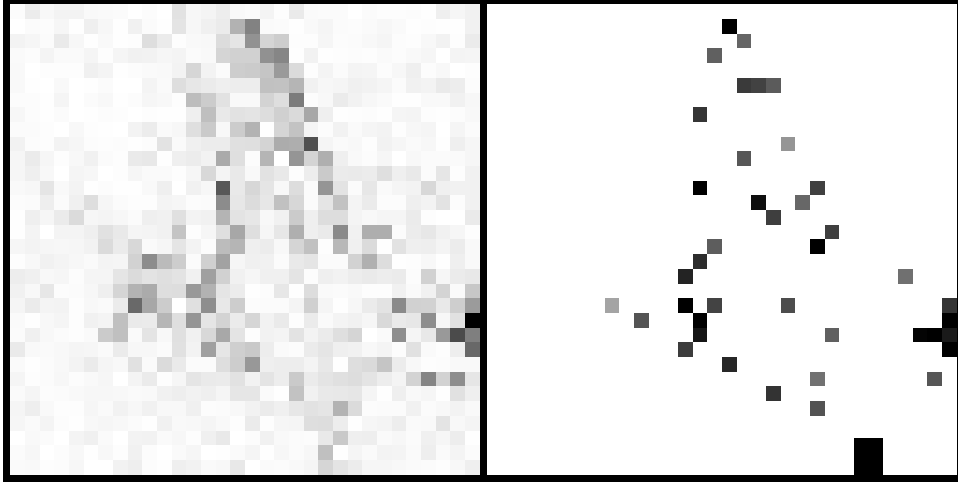


Figure 1.3: **Left**, sub-band of the representation of a natural image under a wavelet-like (DT-CWT) filter bank. Dark pixels represent high amplitude coefficients and light ones those with low amplitude. **Right**, same sub-band of a sparse approximation to the natural image.

this term, interpreting it as a continuous concept based on the concentration of most of the energy of a discrete signal in a relatively small proportion of coefficients. In order to measure the degree of energy concentration we use some norm of the representation vector. In the Figure 1.3 we can see two different two-dimensional distribution of coefficients. Whereas the right panel shows a *sparse* distribution, the left shares-out more the energy among the coefficients.

Apart from Fourier, there are other representations which further favour an efficient processing, making easier to describe the image statistics. For example, linear representations based on multi-scale pass-band filters, called generically *wavelets*, are especially well adapted to represent several properties of the natural images, as scale invariance and the existence of locally oriented structures. It has been experienced that the responses to this kind of filters of natural images typically produce sparse distributions [3, 6].

Redundant representations, those using more coefficients than pixels in the image, favor the extraction of relevant local features [7, 5], so they allow for a more powerful image analysis and processing compared to critically sampled ones. Also, being invariant to translation, rotation, phase, etc. [8, 9], they usually provide better results in image restoration (e.g., [7, 8]).

Nevertheless, direct transformation to a redundant domain does not necessarily increase the sparseness of the representation with respect to

non-redundant ones (see, for example, [10], or Chapter 2 in this Thesis). We need, in addition, some non-linear methods to further increase the sparseness. We define a dictionary as a set of vectors, also called atoms. The *sparse representation problem* is defined as finding the expression of an input signal as linear combination of as few elements as possible from a given representation. In this Thesis, we will tackle a common variant, called *sparse approximation problem*, that allows some tolerance in the fidelity to the input signal. Thus, we can obtain sparse representations with different approximation error, which is very useful for coding and restoration of images. The main issue that has traditionally prevented from solving these problems is their tremendous complexity [11], which has been considered, for a long time, as a too severe obstacle in practice. However, during the last years, and thanks to both a spectacular improvement in the computation speed and to increased consciousness of the importance of sparseness in signal processing, the methods for solving them have been object of big interest. Then, more or less efficient techniques to find sub-optimal solutions have been developed (e.g., [12, 11, 13, 14, 15, 16, 17]); mostly to use the sparseness as *a priori* knowledge for restoration purposes (e.g., [18, 14, 19, 20, 21]).

Current trends in these kinds of methods follow three main strategies. The oldest one tries to incrementally express the image by using greedy techniques, that is, sequentially using those vectors better approximating the part of the image still not represented (e.g., see [3, 22, 23]). The main problem of these techniques is that, being very inefficient, they often get trapped in non-favourable local minima in terms of energy compaction. There are other greedy techniques based on different selection strategies, as [24], where, in a previous work to this Thesis, we selected the significant coefficients by directly thresholding the observed representation vector.

The next strategy, by order of appearance, is based on reducing the complexity of the problem by changing the search for strict sparseness by the minimisation of the sum of the absolute values of the coefficients of the representation. This results in the minimisation of a convex function (see, e.g., [25, 11, 26]). In fact, this technique is called *convex relaxation*. Solving it provides, in general, a local solution to the sparse representation or approximation problems. The first approaches to this problem used numerical optimization algorithms, as conjugate gradient or interior-point methods (e.g., [25, 11, 27]). However, these are often inefficient, moreover when dealing with many dimensions, as typically happens in image processing.

Finally, during the last years, a number of efficient methods have been developed that are providing good practical results in both compaction and

restoration problems. They are based on iteratively combining linear with shrinkage operations (see, e.g., [14, 28, 29, 16, 17]). In Chapter 2 we will make a more extended review of the existing literature in this area.

In the literature, much effort has been devoted to find the constraints under which these two types of techniques manage to find the optimal solution to the sparse representation or approximation problems (see, among others, [30, 31, 32, 33]). But these conditions appear to be too restrictive for being accomplished in most practical image processing scenarios, where natural images, typical representations and useful sparseness levels are used. This has been already shown, for example, in [34, 16].

## 1.2 Contribution of this Thesis

In this Thesis we derive two iterative, though relatively efficient, methods to solve the sparse approximation problem. Besides trying to minimize the  $\ell_0$ -quasi-norm<sup>1</sup> for a given approximation error, we have also developed the resulting version of both methods when using the sum of absolute values of the coefficients ( $\ell_1$ -norm) as the criterion to be minimised, as proposed by the convex relaxation methods.

The first of the methods presented is based on reformulating the sparse approximation problem as finding, given  $p$  and  $R$ , the best approximation to the image inside the  $\ell_p$ -ball of radius  $R$ . Our solution uses alternated orthogonal projections [35, 36] between the  $\ell_p$ -ball and the set of vectors that represent perfectly the image. We have focused on the cases  $p = 0$  and  $p = 1$ . Similar methods can be found in the literature, using both  $p = 0$ , where they have been derived heuristically [15], as well as  $p = 1$  [37, 38]. We have called this method  $\ell_p$ -AP.

The second one is based on re-expressing the cost function of the sparse approximation problem, which is discontinuous and unconstrained, to obtain a continuous and constrained equivalent version. By means of gradient descent in this new cost function, we obtain a generalised version of the Iterative Hard Thresholding (IHT) method [39]. This derivation allows us to prove that the fixed point of the iterations of this method is a local minimum to the sparse approximation problem. Next, we will show the proposed method, which consists of performing gradient descent over gradually less smoothed versions of the new cost function. This method, that we call  $\ell_p$ -GM, has been used before [40, 15, 19]. Nevertheless, this is the first time that it is derived as the solution to an optimisation problem.

---

<sup>1</sup>Though the  $\ell_p$ -norm is not strictly a norm when  $0 \leq p < 1$ , in this Thesis we will usually use this term for every  $p$  value, for simplicity sake.

We have also studied the counterpart of this method,  $\ell_1$ -GM, which solves the convex relaxation problem.

We have followed a very rigorous methodology for obtaining our methods as solutions to optimisation of well-defined standard criteria. However, instead of trying to guarantee that these methods will reach the global optimum under some given constraints, as other authors do (e.g., [30, 31, 32]), in this Thesis we have sought to obtain useful results in practical image processing conditions. Opposite to the extended assumption that the methods based on maximising the strict sparseness are intractable in practice, we have experienced that our sub-optimal methods based on minimisation of the  $\ell_0$ -norm provide much better results than those based on minimising alternative convex criteria (as the  $\ell_1$ -norm). Especially,  $\ell_0$ -GM offers an excellent compaction performance, as other methods based on dynamically adjusting a threshold, and superior not only to our first method, but also to widely used greedy heuristics and convex relaxation techniques. In fact, we show that it has a nearly optimal asymptotic behaviour, when the number of active coefficients of the representation approaches the number of pixels of the image.

In addition, the interest of the proposed techniques is considerably increased when studying the application of the methods to different image restoration problems. We will show very high-quality results in a wide variety of applications, such as removing spatial quantisation artifacts (*de-quantising*), recovering missing pixels (*in-painting*), spatial-chromatic interpolation in digital camera mosaics (*de-mosaicing*), or static super-resolution. Up to our knowledge, this is the first time that this kind of methods are applied to de-quantising.

The content of this document is divided in the following chapters. In Chapter 2 the sparse approximation problem is stated, motivating it by the need for increasing the energy compaction achieved by linear transforms. The main traditional strategies for solving this problem are also analysed in detail. Next, Chapter 3 develops the first method proposed in this Thesis,  $\ell_p$ -AP. We focus on the cases  $p = 0$  and  $p = 1$ , and compare their performance one to each other and also with respect to other methods existing in the literature. In Chapter 4, we derive the IHT method and prove that its fixed point is a local minimum to the problem. Then, we obtain the second proposed method,  $\ell_0$ -GM. We also show an analogous derivation for  $p = 1$ , resulting in the  $\ell_1$ -GM method. In the results section we compare the behaviour of the proposed method versus the previous one and versus those existing in the literature. We also present the practical advantages of  $\ell_1$ -GM compared to other methods solving optimally the convex relaxation problem. In Chapter 5 we see how to adapt our methods to restoration

problems. Finally, in Chapter 6 we show the results of applying these methods to the restoration of several different degradations (see above). Chapter 7 concludes this Thesis.



## Chapter 2

# The Sparse Approximation Problem

The sparse approximation problem can be defined as minimising a measurement of the error when approximating an image as linear combination of a limited number of atoms taken from a redundant set (dictionary). In this work, we will measure this approximation error by means of the Mean Square Error (MSE). In this chapter we show that, in a redundant domain, there are infinite ways of representing an image. Traditionally, the minimum energy solution has been chosen, because it can be easily calculated (linearly). Nevertheless, non-linear methods have much higher potential to compact the energy in few coefficients. These kinds of methods have been extensively used and they are very useful for restoration purposes.

In Section 2.1 we motivate the use of non-linear methods to obtain sparse representations in redundant domains. In Section 2.2, we formulate the sparse approximation problem, and we describe the most important methods that has been used for solving it in Section 2.3. Finally, in Section 2.4 we analyse the conditions under which the sparse approximation problem can be solved optimally using convex optimisation, and same for greedy heuristics.

### 2.1 Analysis-based sparseness vs. Synthesis-based sparseness

Mathematically, to represent an image as a linear combination of vectors taken from some redundant set means to solve a system of linear equations with more equations than unknowns. There are, therefore, infinite solutions.

How to choose one of them? The minimum Euclidean norm solution has been traditionally chosen. Being a linear solution, it is easy to calculate it fast. In fact, usual "inverse" transforms used in image processing for representing images are minimising the Euclidean norm (through pseudo-inverse). However, as we have discussed in the Introduction, there are good reasons to look for solutions concentrating the energy in as few as possible coefficients. These solutions express the image, possibly with some error, as a linear combination of fewer vectors from the dictionary than the linear solution. Although some authors have observed that the linear response to wavelet filter banks of natural images already concentrates most of the energy in relatively few coefficients (e.g., [3, 41, 42, 6]), nevertheless the minimum Euclidean norm solution tends to spread as much as possible the energy among the coefficients, which makes it inadequate for sparse representation. Next we illustrate that the energy compaction in redundant dictionaries is much bigger for certain non-linear transformations of the image.

Figure 2.1 shows the fidelity (in dB) to the original image obtained by approximating it using different representations and a wide range of sparseness levels, that is, of number of vectors involved in the approximation. Data have been averaged (using the MSE) for the five images in our test set (see Appendix A). The fidelity to the original image is measured by the Peak Signal-to-Noise Ratio (PSNR), defined as  $10 \cdot \log_{10}(\frac{\rho^2}{MSE})$  and which is measured in decibels (dB), and where  $\rho$  is the maximum value of the involved signals. In our case  $\rho = 255$ , because we are dealing with 8-bits monochromatic images. In this Thesis, we usually represent the number of active coefficients normalized by the total number of pixels in the image). Each curve in Figure 2.1 has been obtained by reconstructing the image using the largest coefficients (in amplitude) of each linear transform, for different sparseness levels. Three of the representations used are critically sampled (pixels, Fourier, Haar Wavelets) and the fourth one is redundant (Dual Tree Complex Wavelets or DT-CWT [43]).

We can see how the quality of the approximation for a given number of elementary functions is increased when we transform the pixels to the Fourier domain, and even more when we use critically sampled wavelets. Unfortunately, the performance of the linear redundant representation suffers a brisk fall, due to many coefficients responding to the same feature of the image, which results in a sparseness decrease. Then, as we have pointed out before, if we want to outperform the critically sampled wavelets, we need to use a non-linear vector selection mechanism. In Figure 2.1 we also show the result obtained using DT-CWT with the best method proposed in this Thesis ( $\ell_0$ -GM, see Chapter 4). We can see that it provides a great



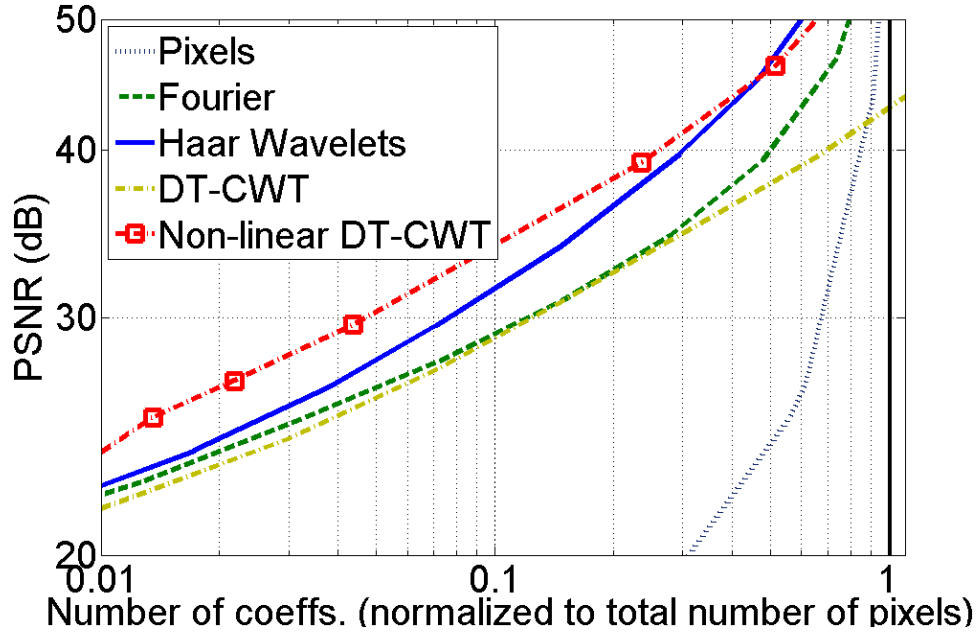


Figure 2.1: Sparse approximation results comparison for several representations. Data have been obtained using the largest in amplitude coefficients of the linear responses and then averaging in our test set. Performance is measured in terms of PSNR given a number of coefficients. This number is normalised by the total number of pixels in the image.

improvement over the approximations obtained using linear representations. The need for non-linear selection mechanisms to better concentrate the energy of the representation in few coefficients has been already treated by a number of authors, as, for example [27, 11, 26, 44, 45, 46].

Figure 2.2 illustrates the effect of using this type of selection mechanisms on the transform with DT-CWT of the *Peppers* image. Top left panel shows the coefficients of a sub-band of the linear representation of the image. In central panel we show the coefficients of the representation obtained non-linearly by maximising the sparseness (using  $\ell_p$ -AP, see Chapter 3). In the former we can see a less sparse distribution of coefficients than in the latter, which strongly decreases simultaneous responses to the same feature. This better compaction makes possible that, using only a small proportion of the total number of coefficients (around 7 times less than pixels in the image, in this case) a high reconstruction quality is preserved (35.70 dB in this example, see right panel).

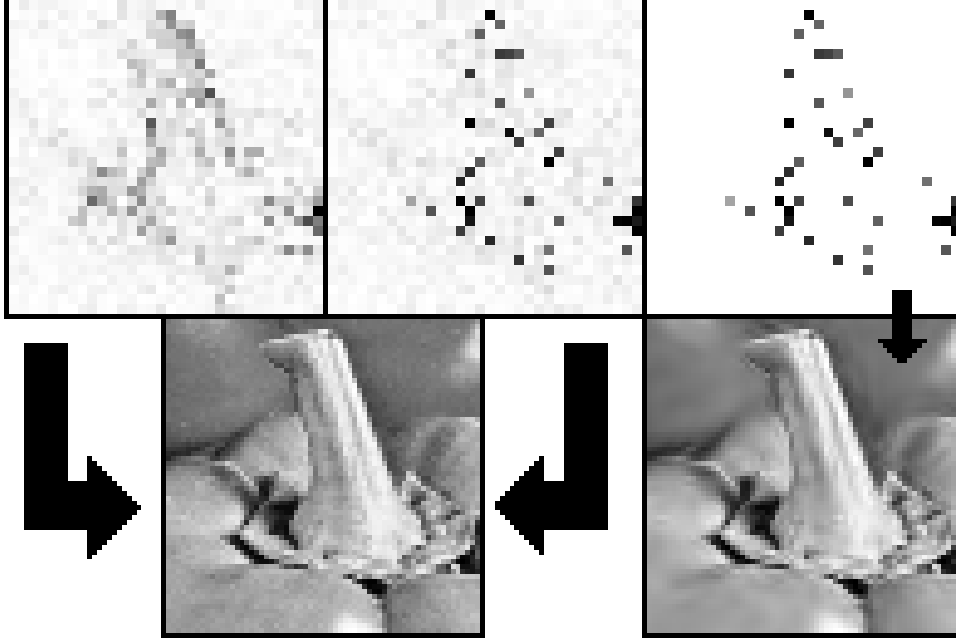


Figure 2.2: **Top-left**, highest-frequency sub-band of analysis vector of Peppers image using DT-CWT with 8 scales. Light and dark points correspond, respectively, with low and high amplitudes of the coefficients. The size of the sub-band has been doubled in both directions through replication of rows and columns in order to match the size of the image, and then it has been cropped to  $64 \times 64$  for visibility. **Top-centre**, same sub-band, but this time non-linearly obtained using the  $\ell_0$ -AP method (see Chapter 3). **Top-right**, result of applying a threshold in amplitude to the result in central panel (preserving 7 times less coefficients than pixels in the image). **Bottom-left**, original image, which is perfectly reconstructed by the representations corresponding to the left and central panels of the top row. **Bottom-right**, approximation obtained with the sparse coefficients corresponding to the top-right panel (35.67 dB).

## 2.2 Formulation of the sparse approximation problem

Next we formulate the sparse approximation problem. Let  $\Phi$  be a  $N \times M$  matrix with  $M > N$  and  $\text{range}(\Phi) = \mathbb{R}^N$ , representing the synthesis operator of a Parseval tight frame<sup>1</sup>.  $N$  is the number of pixels in the original domain and  $M$  the number of coefficients in the transformed domain. Then, for an observed image,  $\mathbf{x} \in \mathbb{R}^N$ , the linear system of equations:

$$\Phi \mathbf{a} = \mathbf{x}, \tag{2.1}$$

<sup>1</sup>A linear transform with a Parseval tight frame preserves the Euclidean norm of the original vector. We will use the term Parseval frame for simplicity.

has infinite solutions in  $\mathbf{a} \in \mathbb{R}^M$ . If we want to favour one of them, we should add some additional criteria. Thus, we can introduce a function  $f(\mathbf{a})$  to discriminate this solution, and the problem is set out as:

$$\hat{\mathbf{a}}^f = \arg \min_{\mathbf{a} \in \mathbb{R}^M} f(\mathbf{a}) \text{ s.t. } \Phi \mathbf{a} = \mathbf{x}. \quad (2.2)$$

Among the possible options for  $f(\mathbf{a})$ , the  $p$ -th power of the  $\ell_p$ -norm has often been used. For a given value of  $p$ , this is defined as  $\|\mathbf{a}\|_p^p = \sum_{i=1}^M |a_i|^p$ . In Figure 2.3 we show the shape of this function in its one-dimensional version and for several values of  $p$ . We have already mentioned that the most commonly used norm has been the Euclidean,  $p = 2$ , obtaining for it the minimum energy solution,  $\mathbf{a}^{LS}$ . This is especially easy to calculate for Parseval frames. In fact,  $\Phi^T = \Phi^T[\Phi\Phi^T]^{-1}$  is the analysis operator of the Parseval frame, calculated as the pseudo-inverse of  $\Phi$ , that is,  $\mathbf{a}^{LS} = \Phi^T \mathbf{x}$ . However, as we have seen in the previous section, this is not an appropriate solution in terms of maximising the sparseness, which is measured using the  $\ell_0$ -norm. This is expressed, by extension of the definition of norm, as the number of non-zero coefficients in the vector. Then, the sparse representation problem is expressed as:

$$\hat{\mathbf{a}}^0 = \arg \min_{\mathbf{a}} \|\mathbf{a}\|_0 \text{ s.t. } \Phi \mathbf{a} = \mathbf{x}. \quad (2.3)$$

However, redundant dictionaries typically used in image processing do not allow representations of natural images where most of the coefficients are exactly zero. It is more useful to search for representations concentrating most of the energy in as few as possible coefficients, so most of them have relatively small amplitudes. This kind of distributions, as we have exemplified in the previous section, have the property that a few high-amplitude coefficients can approximate the image with an error that can be acceptable for most applications. This is why many authors prefer to relax the constraint of Equation (2.3), and to formulate the sparse approximation problem as:

$$\hat{\mathbf{a}}^0(\lambda) = \arg \min_{\mathbf{a}} \{\|\mathbf{a}\|_0 + \lambda \|\Phi \mathbf{a} - \mathbf{x}\|_2^2\}, \quad (2.4)$$

where  $\lambda \in \mathbb{R}^*$  is a real positive number controlling the relative importance between the sparseness and fidelity terms in the cost function; so the higher its value, the smaller reconstruction error of the solution, whilst the sparseness is reduced. Note that Equation (2.4) is equivalent to either minimise  $\|\mathbf{a}\|_0$  for a given quadratic error (which depends on  $\lambda$ ) or to minimise the quadratic error for a given  $\ell_0$ -norm of the approximation (which also depends on  $\lambda$ ). Equation (2.3) is a particular case of

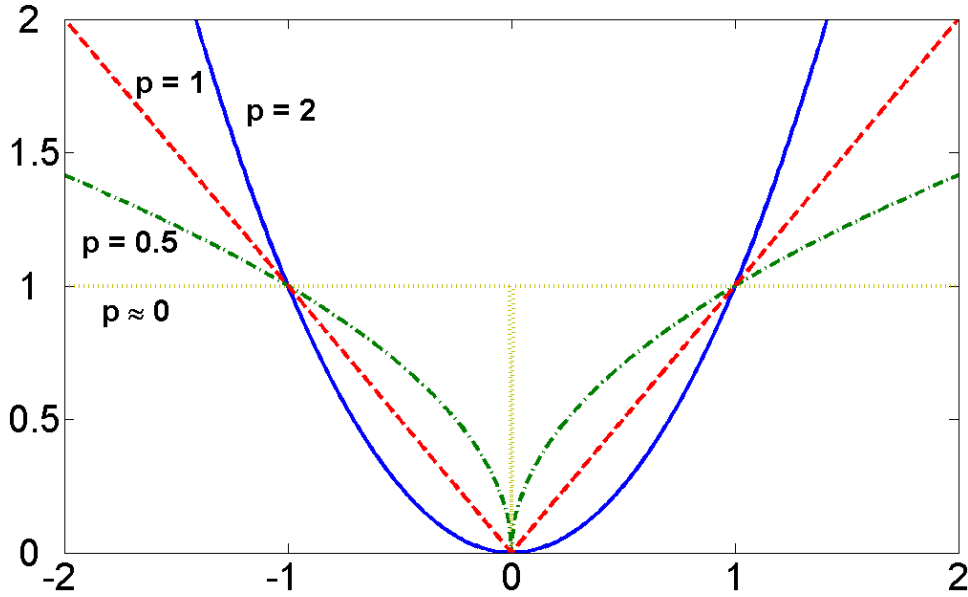


Figure 2.3: One-dimensional  $p$ -th power of the  $\ell_p$ -norm for different  $p$  values.

Equation (2.4), when  $\lambda$  tends to infinite. As an example, the sub-band shown in the central panel of Figure 2.2 can be understood as belonging to the solution of (2.4) when  $\lambda$  is infinite, where the image is perfectly reconstructed but there is no sparseness. On the other hand, the sub-band in the right panel would correspond to the solution with a lower value of  $\lambda$ , using less coefficients but reducing the quality of the reconstruction of the image.

## 2.3 The sparse approximation problem in the literature

Finding the global optimum of the sparse approximation problem is a combinatorial problem, and, thus, NP-complex. It requires to look for every possible combination of columns of  $\Phi$ , and solve for the least squares solution, choosing the one providing the lowest MSE. Some of the first approximations were restricted to some dictionaries (for example, wavelet packets or localised trigonometric functions) where it is possible to extract the orthonormal basis representing more efficiently the image [41]. But, during the last two decades, more effective and general methods to find local solutions to both problems (2.3) and (2.4) have become popular. We can classify these techniques in three main groups: greedy heuristics, methods

based on convex relaxation and methods based on iterative shrinkage. Next we will review the most important literature references in these three families.

### 2.3.1 Greedy heuristics

The first strategies to obtain local solutions to Equation (2.4) using general dictionaries were derived heuristically. Among them, the most frequently used comes from the observation that it is convenient, in order to approximate a given vector from a given a collection of vectors, start by selecting from the latter that vector having maximum correlation with the former (see, for example, [47]).

This family of algorithms is well known and widely used. In fact, these methods have been re-invented in several fields. In statistical modeling they are called *forward stepwise regression*, and they have been used since the 60s (see, for example, [48, 49] and references therein). When used in signal processing, they have been termed *Matching Pursuit* (MP) [12] and *Orthogonal Matching Pursuit* (OMP) [22], among others. In approximation theory they are referred as Greedy Algorithms [50, 51, 52, 53]. A wide review of these methods, applied to non-linear approximation, can be found in [54].

In our context, MP is the simplest greedy method. It is implemented through a set of indices,  $I$ , indicating the functions of the dictionary that have been already selected to form the approximation, and a residual,  $\mathbf{r}$ , which is the part of the image not yet represented. The set of indices is initialised empty,  $I^{(0)} = \emptyset$ , and the residual to the entire image,  $\mathbf{r}^{(0)} = \mathbf{x}$ . In each iteration  $k + 1$ , the selected basis is updated by adding that vector having maximum correlation with the residual:

$$I^{(k+1)} = I^{(k)} \cup \{i : \langle \phi_i, \mathbf{r}^{(k)} \rangle \geq \langle \phi_j, \mathbf{r}^{(k)} \rangle \forall j \neq i, j \notin I^{(k)}\},$$

where  $\langle \cdot, \cdot \rangle$  indicates inner product of two vectors. The estimation is updated then as:

$$\hat{\mathbf{x}}^{(k)} = \sum_{j=1}^k \langle \phi_{i(j)}, \mathbf{r}^{(j)} \rangle \phi_{i(j)},$$

where  $i(j)$  represents the chosen index at  $i$ -th iteration. The next step is to update the residual as:

$$\mathbf{r}^{(k+1)} = \mathbf{r}^{(k)} - \langle \phi_i, \mathbf{r}^{(k)} \rangle \phi_i.$$

The iterations end either when the desired error level is reached, or when the desired number of terms has been selected. The main problem of MP is that, given the set of elementary functions selected at each iteration, the amplitudes of the coefficients are not optimised to represent the image. OMP adds an intermediate least squares optimisation step. Given the subset of indices selected in iteration  $k$ ,  $I^{(k)}$ , we define the  $N \times k$  matrix  $\Phi_{I^{(k)}}$  formed by all columns  $\phi_i$  from  $\Phi$  such that  $i \in I^{(k)}$ . Then, in all OMP iterations, additionally to the MP steps, the following problem is solved:

$$\hat{\mathbf{a}}_I^{(k)} = \arg \min_{\mathbf{a}_I \in \mathbb{R}^k} \|\mathbf{x} - \Phi_{I^{(k)}} \mathbf{a}_I\|_2.$$

To update the residual we now use:  $\mathbf{r}^{(k)} = \mathbf{x} - \Phi_{I^{(k)}} \hat{\mathbf{a}}_I^{(k)}$ . Not only this method provides better compaction results than MP, but it also converges faster [22]. However, as often happens with greedy strategies, OMP gets stuck in local optima that frequently are not satisfactory for the problem at hand. A number of modifications have been proposed based mostly on recursive searches, sometimes also in hierarchical trees, trying to explore the maximum possible number of combinations at each step (see, for example, [55, 56, 57, 58]). Another disadvantage of OMP, even more serious in the variants mentioned, is that selecting only one coefficient at each step is unfeasible in terms of computation time for most commonly used dictionaries in image processing. There exist more efficient variants of OMP which selects more than one coefficient at each step, either with a fixed (e.g., [59]) or variable (e.g., [60, 23]) step size. In this Thesis, we refer to these methods with the name of maybe the more extended of them, Stage-wise OMP (StOMP) [23].

OMP has been applied with different degree of success to several applications, as for example noise removal [21, 61], video coding [62, 63, 64], colour image compression [65] or audio signals separation [66, 53]. However, all these applications either need few steps of the algorithm or they used orthogonal dictionaries. Only StOMP-like techniques can be applied, in most practical situations, with redundant dictionaries commonly used in image processing.

In a previous work to this Thesis [24], we presented a method for removing spatial quantization artifacts through finding a sparse approximation to the observation. The resulting method selects the significant coefficients via direct thresholding the amplitudes of vector  $\mathbf{a}^{LS}$ , which can be also seen as a greedy strategy.

### 2.3.2 Convex relaxation problem and *Basis Pursuit*

As we have already seen, using the Euclidean norm to obtain solutions (either approximate or exact ones) to the linear system of equations (2.1) is not efficient in terms of obtaining sparse solutions. On the other hand, the  $\ell_0$ -norm is not convex nor even continuous, which makes the optimisation problem difficult to solve. This disadvantage, in practice, is not well solved by greedy heuristics. Can we find an intermediate way allowing to profit from the advantages of both approaches? In some cases the answer is provided by the  $\ell_1$ -norm. In this case,  $p = 1 < 2$ , so it promotes sparse solutions; but, as it is convex, global optima are achievable in polynomial time. Note also that, as  $\ell_1$ -based techniques optimise all the coefficients at the same time, we can reach approximation levels in practice that OMP cannot. This variant is termed *convex relaxation problem*, and it is formulated, analogously to Equation (2.4), as:

$$\hat{\mathbf{a}}^1(\lambda) = \arg \min_{\mathbf{a}} \{\|\mathbf{a}\|_1 + \lambda \|\Phi \mathbf{a} - \mathbf{x}\|_2^2\}. \quad (2.5)$$

This problem has been frequently formulated, since the 50s, in terms of a linear program (LP). For example, in [25], a *Simplex* method is proposed to get minimum  $\ell_1$ -norm solutions formulating the problem as an LP. Nevertheless, this technique was not systematically exploited until the microprocessor speed and the memory capacity of computers were boosted in the last three decades or so.

*Basis Pursuit* (BP) and *Basis Pursuit De-noising* (BPDN) methods [11, 26] respectively solve the convex representation and approximation problems through an LP equivalent to the Equation (2.5), by using Interior-Point Methods. Both BP and BPDN have had so much diffusion that are nowadays taken as synonymous for the convex relaxation problem.

The appearance of BP and BPDN was preceded by two significant advances: 1) the surprising discovery that we can nearly optimally estimate a piece-wise smooth function from a noisy observation, by solving a convex relaxation problem using the appropriate wavelet basis and a value of the  $\lambda$  parameter related to the variance of the noise [30]; and 2) the development of LASSO, which proposed the convex relaxation to solve the problem of subset selection in linear regression [67]. The technique known as *Least Angle Regression* [60] was later adapted to solve the LASSO formulation. The great disadvantage of this and other techniques (for example, FOCUSS [27]) is that they need to explicitly deal with the dictionary matrix. This is not practicable in most image processing scenarios, where the matrix often have hundred of thousands rows and columns. However, we can use very efficient techniques to solve the product of the big matrix corresponding to

the representation used with a vector corresponding to the image. Both BP and iterative methods seen in the next section take advantage of these tools. A brief but complete history of the different approximations to the convex relaxation problem can be found in [68].

Apart from coding, the main areas where this kind of methods have been applied are statistical regression [67, 27] and noise removal [11, 26], although it is possible to find other applications are, for example, missing pixels recovery [69].

Recently, a great sense of expectancy has been generated with a new application for the convex relaxation problem, which is named *Compressed Sensing* (e.g., [70, 71]). This technique is based on the observation that a relatively small number of random projections of a sparse signal suffice to recover from them a good approximation of that signal [72]. It has been proposed as a powerful alternative to the traditional Nyquist Sampling Theorem, when the signal can be expressed as linear combination of few elements.

It is also possible to use intermediate quasi-norms ( $0 < p < 1$ ) when regularising the problem (2.2) (see, for example, [73, 14]). Although these norms do not lead to convex cost functions, and therefore the global optimum is difficult to calculate, several authors have pointed the fact that the marginal distribution of the coefficients of the linear transform of a natural image under a wavelet filter bank is appropriately modeled using these norms (see, for example, [74, 18]). This approach to the sparse approximation problem is very interesting, but its study is beyond the scope of this Thesis.

### 2.3.3 Iterative shrinkage

We have already said that application of greedy methods to sparse approximation of natural images is not completely satisfactory, because they are too expensive in computational terms and they get trapped in unfavourable local optima. On the other hand, methods based on a direct approach to the convex relaxation problem require too much computation for natural images of usual size.

During the last years new efficient techniques have been developed in order to solve the problem (2.4). They are based on iteratively apply shrinkage operations (coefficient-wise) combined with linear projections. By shrinkage we understand those operations that decrease the amplitude of the coefficients in the representation, possibly setting to zero the ones below and amplitude threshold. We call these techniques *Iterative Shrinkage Methods*<sup>2</sup>.

---

<sup>2</sup>Other possible name is *Thresholded Landweber Iteration*.



The iterative shrinkage methods have been widely used. For example, [75] already used similar techniques for image segmentation. In [8] it is proved that, in presence of observations degraded by additive white Gaussian noise, the simple method of applying a (hard or soft) thresholding to an orthogonal linear transform of the image achieves, for some signal models, optimal results in terms of MSE.

The optimality of these shrinkage operations depends on the orthogonality of the dictionary matrix<sup>3</sup>. But a single shrinkage, by itself, does not lead to optimal solutions when using redundant transforms. Even though, it has been widely used in all types of redundant representations (see [76] and references therein). Shrinkage, within an iterative scheme, can provide very efficient and reasonable good approximate solutions.

More recently, other authors [74, 77, 78] simplified the ideas in [79, 8] and re-stated the problem as in Equation (2.2), starting from a Bayesian formulation to find the *Maximum A Posteriori* (MAP). The Iterative Shrinkage Method, applied frequently to solve this problem, can be described with the following iterations:

$$\mathbf{a}^{(k+1)} = S_p(\mathbf{a}^{(k)} + \Phi^T(\mathbf{x} - \Phi\mathbf{a}^{(k)}), \theta), \quad (2.6)$$

where  $S_p(\mathbf{a}, \theta)$  indicates a certain shrinkage operation on a vector  $\mathbf{a}$  with threshold  $\theta$ , which in our case is function of the parameter  $\lambda$ . The linear operation that is the argument of the shrinkage comes from minimising the fidelity term in the Equation (2.2), for a given vector  $\mathbf{a}$ , and it is the orthogonal projection of the vector onto the affine subspace of perfect reconstruction (see Section 3.1).

Up to our knowledge, the first method proposing an iterative thresholding technique to solve problem (2.2) was [73, 14], which use  $\ell_p$ -norms with  $0 < p \leq 1$ . The method is derived from a formulation based on *Expectation-Maximisation* (EM). Later on, in [28], the same method was derived in the case  $p = 1$ , using extra terms added to the cost function without changing its minima. Also [76] derives a similar algorithm using  $p = 1$ , but from a different perspective, because it tries to maximise the sparseness of the representation. The method used in all these works is based on alternating a soft thresholding with a linear projection. It is commonly named *Iterative Soft-Thresholding* (IST). It corresponds to iterate with Equation (2.6) using soft-thresholding, defined as  $S_1(\mathbf{a}, \theta) = \mathbf{b}$ , where:

$$b_i = \begin{cases} \text{sign}(a_i) \cdot (|a_i| - \theta), & |a_i| > \theta \\ 0, & |a_i| \leq \theta. \end{cases} \quad (2.7)$$

---

<sup>3</sup>It also depends on using the Euclidean norm as error term.

Here,  $\text{sign}(\cdot)$  indicates the sign function. The value of the threshold results in  $\theta = \frac{1}{2\lambda}$ . Convergence of the method was proven in [28, 37]. These techniques have also been used to separate the different morphological components of a signal, using several dictionaries, each having the property of having sparse response to a different family of input signals (Morphological Component Analysis, MCA) [45, 19]. Other authors have derived similar methods from different points of view (e.g., [80]). See [81] for a review on iterative shrinkage-based methods.

Other iterative algorithms exist, as the one described in [82] using soft thresholding as a gradient descent operation, and using a linear search to find the optimal step size for each iteration. Another example is [77], that applies soft thresholding in redundant representations within a variational formulation to remove noise and for compression; or the application of the generic optimisation technique called *Iterative Re-weighted Least Squares* [83], which reformulates the Equation (2.5) as a quadratic programming problem (requiring direct manipulation of the transform as a large-scale matrix). We want to emphasise, among them, the gradient descent method recently proposed in [38]. This method is based on the same type of operations than IST, but the fixed threshold is changed by a threshold adapted in each iteration to keep constant the  $\ell_1$ -norm. This method is also derived in Chapter 3 of this Thesis, when using the  $\ell_1$ -norm. In addition to this, it applies an optimised step size of the gradient descent to accelerate convergence.

An alternative method to IST can be derived using hard-thresholding instead, which is known as Iterative Hard-Thresholding (IHT). In this case, the operation used in the iterations (2.6) is  $S_0(\mathbf{a}, \theta) = \mathbf{b}$ , with:

$$b_i = \begin{cases} a_i, & |a_i| > \theta \\ 0, & |a_i| \leq \theta. \end{cases} \quad (2.8)$$

And the value of the threshold results in  $\theta = \lambda^{-\frac{1}{2}}$ . The first paper (heuristically) proposing such a method was [13]. It is also heuristically proposed in other works, as [39]. In [84] it has been derived using surrogate functions, and, moreover, its convergence to a local minimum of the sparse approximation problem has been proved<sup>4</sup>.

In [15] a similar heuristic is proposed, but instead of applying a fixed threshold, the number of non-zero coefficients after each hard-thresholding operation is fixed. This is a similar method to what we derive in Chapter 3 of this Thesis, but here we derive it as solution to an optimisation

---

<sup>4</sup>This work is parallel to our derivation of the method and proof of convergence proposed in Chapter 4 and already published in [10].

method. This modification provides better compaction results. A further improvement on these methods can be found in [85], and uses an adaptive threshold depending on the energy of each sub-band of the representation.

Alternative solutions to soft and hard-thresholding have been proposed. For example, Firm-Shrinkage [86] tries to improve the results obtained, in [79], using both thresholding types. In [87, 88] a variational formulation for IST, IHT and Firm-Shrinkage is shown.

Several authors have compared the performance of both soft and hard-thresholding. Unless exceptional cases, such as [78], most of them have experienced that hard outperforms soft-thresholding [78, 34, 89, 90, 91].

Many authors have experienced a great improvement in the general performance of this type of algorithms when the threshold is decreased at each iteration (dynamic thresholds). This is one of the ideas involved in the proximal-points methods [40]. These methods solve iteratively a succession of problems formulated using the Equation (2.5) with increasing values of  $\lambda$ . The dynamic version of IST is found in the MCA variant [19, 20, 92] but no theoretical justification is provided. Also heuristically, [38] proposes to increase dynamically the radius of the  $\ell_1$ -ball where each soft-thresholding operation is projected. With respect to heuristic versions of dynamic IHT, they can be found in [93, 13, 15]. [94] presents a method based on substituting the  $\ell_0$ -norm by a equivalent continuous function. They use a Gaussian function leading to an algorithm different from  $\ell_0$ -GM.

Despite their recent introduction, methods based on iterative shrinkage have already proved to be very powerful for a number of applications. For example, [20] uses IST and IHT to recover missing pixels in the image. Also several papers use IST to approach the classic restoration problem (blurring plus noise, e.g., [29, 95, 96]). Moreover, other applications can be found, as medicals imaging [97] or video coding [98]. Finally, we want to emphasise the application of this type of techniques to Compressed Sensing [99, 72].

## 2.4 Equivalence conditions when minimising $\ell_1$ and $\ell_0$ -norms

In previous sections we have reviewed the most common techniques to solve the sparse approximation problem of Equation (2.4). We have seen that the global solution cannot be found in practice, and that three main approaches have been proposed: greedy heuristics, convex relaxation methods and iterative shrinkage methods. Next question is: Do these methods offer equally good solutions? In this section we review some surprising results proving that, under certain conditions, both greedy methods (OMP)

and convex relaxation methods reach the global optimum to the sparse *representation* problem, and provide a solution with an error proportional to the level of noise when dealing with the sparse *approximation* problem.

In the foundation of these results we find the concept of mutual coherence of a matrix. This is defined as  $M(\Phi) = \sup\{\langle\phi_i, \phi_j\rangle; \forall i \neq j\}$ . There is also a stronger constraint, associated to a different measure of the richness of a dictionary, called *Spark*( $\Phi$ ) or Kruskal range. It is defined as the minimum number of matrix columns forming a linear dependent set. It has been stated the following relation between *Spark*( $\Phi$ ) and mutual coherence:  $\text{Spark}(\Phi) \geq \frac{1}{M(\Phi)}$ .

The first step to establish equivalence conditions between problems (2.4) and (2.5) was to prove that, if a solution is sparse enough, it is the only global optimum to the problem (2.3) (see [100, 101] and also [102, 103, 104, 105, 106]). These results are interesting because they allowed for the first time to have a simple way to check if solutions obtained with different methods were optimal or not. The condition to check was established as  $\|\hat{\mathbf{a}}^0\|_0 < \frac{\text{Spark}(\Phi)}{2}$ . But more general results were still missing, because there was no known method to obtain effectively the global optimum.

It has been empirically shown in [11, 26] (using small 1-D standard discrete functions) that the solution to the convex relaxation problem is sparser than the minimum Euclidean norm solution. Defining the mutual coherence, for two matrices of equal size, as  $M(\Phi_A, \Phi_B) = \sup\{|\langle\phi_a, \phi_b\rangle| : \phi_a \in \Phi_A, \phi_b \in \Phi_B\}$ , in [101] it is proven that, if the solution to the sparse representation problem holds that  $\|\hat{\mathbf{a}}^0\|_0 < \frac{1}{2}(1 + \frac{1}{M(\Phi)})$ , when  $\Phi$  is formed by concatenating two mutually incoherent dictionaries (those leading to a small value of  $M(\Phi_A, \Phi_B)$ ), then  $\hat{\mathbf{a}}^0$  is the unique global solution and it can be obtained through the minimisation of the  $\ell_1$ -norm. Later on, [107] improved this upper bound setting it to  $\|\hat{\mathbf{a}}^0\|_0 < \frac{0.9142}{M(\Phi)}$ . This result can be also extended to redundant dictionaries [108, 109, 110]. In these works the uniqueness upper bound for general dictionaries were decreased to  $\|\hat{\mathbf{a}}^0\|_0 < \frac{\text{Spark}(\Phi)}{2}$ , what is twice tighter than the upper bound for the convex relaxation methods to reach the global optimum of the sparse representation problem. Finally, [30] relaxed the conditions to prove that, if a signal has a representation with fewer than  $\tau N$  non-zero coefficients, where  $\tau > 0$  is a real proportionality factor, then the solution to the convex relaxation problem is equal to the sparse representation problem. However, it is not made clear how to calculate  $\rho$  for each dictionary.

In most practical situations it is not reasonable to assume that the observed coefficients perfectly represent the signal. Then, it is more interesting the scenario where an ideal signal has a sparse approximation, but we only observe a version degraded with white additive noise. [32]

studies the algorithms based on convex relaxations in the same conditions that [108, 109, 110]. Under these sparseness bounds, and if the dictionary has the property of being mutually incoherent, then the convex relaxation algorithms are globally stable. That is, the error made is proportional to the noise level even under an arbitrary amount of noise. It is also shown that, under certain conditions, the support of the results of these methods is contained within the ideal selection existing for the original signal. Similar results were also derived in [110, 68]. We refer to [46] to find a more complete review of these works.

On the other hand, in [31, 32] it is shown that greedy techniques, such as OMP, find the global solution in the same conditions as BP for the sparse approximation problem, with the difference that OMP is locally stable. That is, under a small quantity of noise we can recover the ideal sparse representation with an error that increases, at worst, proportionally to the noise level. However, in [111, 112] it is shown that, in practice, OMP gets better results, and it is also faster. We refer to [112] to see these results in detail. Nowadays, there are no similar results stating the conditions under which the iterative shrinkage methods reach the global optimum to the sparse representation or approximation problems.



# Chapter 3

## Sparse approximation using alternating projections

In this chapter we present a simple and robust optimisation non-linear method providing sub-optimal solutions to the sparse approximation problem (Equation (2.4)). It is based on, given a Parseval frame transforming the pixels of the image to a redundant transformed domain, and given values for parameters  $p$  and  $R$ , look for the vector with  $\ell_p$ -norm equal to  $R$  which best approximates the image, in terms of the MSE in the reconstruction. The method consists of applying alternating orthogonal projections<sup>1</sup> onto the set of vectors of the transformed domain with  $\ell_p$ -norm equal or less than  $R$  and the set of vectors representing perfectly the image. We call this method  $\ell_p$ -AP (for *Alternated Projections*). We show that it converges to the global optimum of the cost function when  $p \geq 1$ , and to a local optimum if  $0 \leq p < 1$ . Here we will focus on the cases  $p = 0$  and  $p = 1$ . We will show that, even being sub-optimal,  $\ell_0$ -AP clearly outperforms  $\ell_1$ -AP (which is equivalent to other Basis Pursuit-like methods). We will also see how to de-bias the coefficients of the solution given by  $\ell_1$ -AP, through a LS-optimization of the active coefficients. We obtain slightly better compaction results than those of  $\ell_0$ -AP. Finally, we will see that  $\ell_0$ -AP also outperforms other existing strategies, as greedy heuristics and iterative shrinkage methods based on using fixed thresholds. Previously, [15] had already proposed a heuristical method equivalent to  $\ell_0$ -AP, whilst a very similar method to  $\ell_1$ -AP appears in [38], developed simultaneously and independently to our work.

In Section 3.1 we describe the method  $\ell_p$ -AP, explaining in detail the particular cases  $p = 0$  and  $p = 1$ . In Section 3.2 we explain a method

---

<sup>1</sup>Here we use the term *orthogonal projection* in a wide sense, involving any minimum Euclidean distance projection.

that, given a set of indices, find the vector whose support is that set and best approximates the image. After that, we describe the implementation details in Section 3.3, to discuss the results of the compaction experiments in Section 3.4. Section 3.5 concludes this chapter.

### 3.1 $\ell_p$ -AP method

For clarity sake, we start rewriting the problem of approximation minimising a general  $\ell_p$ -norm:

$$\hat{\mathbf{a}}^p(\lambda) = \arg \min_{\mathbf{a}} \{ \|\mathbf{a}\|_p^p + \lambda \|\Phi \mathbf{a} - \mathbf{x}\|_2^2 \}, \quad (3.1)$$

where  $\|\mathbf{a}\|_p = (\sum_{i=1}^M |a_i|^p)^{\frac{1}{p}}$  denotes the  $\ell_p$ -norm of  $\mathbf{a}$ . If we assume a determined value of  $\lambda$ , then  $\hat{\mathbf{a}}^p(\lambda)$  will have a determined  $\ell_p$ -norm, that we denote<sup>2</sup>  $R(\lambda)$ . Then, to solve the Equation (3.1) for a given  $\lambda$  value is equivalent to minimise the approximation error for a given norm of the solution,  $\|\hat{\mathbf{a}}^p(\lambda)\|_p = R$ :

$$\hat{\mathbf{a}}^p(\lambda) = \hat{\mathbf{a}}^p(R) = \arg \min_{\mathbf{a} \in \mathbb{R}^M} \|\Phi \mathbf{a} - \mathbf{x}\|_2^2 \text{ s.t. } \|\mathbf{a}\|_p^p = R. \quad (3.2)$$

A  $\ell_p$ -ball with radius  $R$ , centred at the origin, is formed by all those vectors with  $\ell_p$ -norm less or equal than  $R$ ,  $B_p(R) = \{\mathbf{a} \in \mathbb{R}^M : \|\mathbf{a}\|_p^p \leq R\}$ . Then, we solve here the problem:

$$\hat{\mathbf{a}}^p(R) = \arg \min_{\mathbf{a} \in B_p(R)} \|\Phi \mathbf{a} - \mathbf{x}\|_2. \quad (3.3)$$

Although, strictly speaking,  $\hat{\mathbf{a}}^p(R) \neq \hat{\mathbf{a}}^p(\lambda)$  (because those vectors with  $\ell_p$ -norm lower than  $R$  are feasible), in practice, as we will see below, we use a method that provides, under certain conditions, solutions holding the constraint of the optimization in Equation (3.2).

Equation (3.3) can be solved using several techniques. We have chosen to use the Alternating Projections Method [35, 36] due to its simplicity and convergence properties. This method consists of orthogonally alternately projecting onto two or more sets until reaching convergence. When the involved sets are convex and they have intersection, the method converges to the orthogonal projection of the input vector onto the intersection of the sets. When they are convex but they have empty intersection, the method converges to a limit cycle of minimum distance between both sets. When one or more of them are non convex, the limit cycle is reached in a local

---

<sup>2</sup>Through this Thesis, we remove the dependency of  $R$  on  $\lambda$  for clarity sake.



minimum of that distance<sup>3</sup>. See [113] for a more complete discussion of the convergence properties when non-convex sets are involved.

In order to apply the Alternating Projections Method, we have to define two sets. First of them is the set of solutions to the Equation (2.1), defined as  $S(\Phi, \mathbf{x}) = \{\mathbf{a} \in \mathbb{R}^M : \Phi \mathbf{a} = \mathbf{x}\}$ . It is an affine sub-space of  $\mathbb{R}^M$ , and, thus, it is convex. The second set is the  $\ell_p$ -ball of radius  $R$ , centred at the origin,  $B_p(R)$ , for a given  $p$  and  $R$  values. This set is convex only if  $p \geq 1$ . Here we assume that the starting vector for the iterations has a  $\ell_p$ -norm larger than desired (as it happens in practice), which implies that the solution will lie on the boundary of the  $\ell_p$ -ball and, as  $B_p(R)$  is a closed set, the optimization constraint of Equation (3.2) holds.

We denote  $P_C^\perp(\mathbf{v})$  the orthogonal projection of a vector  $\mathbf{v}$  onto a given set  $C$ . The orthogonal projection of  $\mathbf{a}$  onto the affine sub-space  $S(\Phi, \mathbf{x})$  can be obtained easily, being:

$$P_{S(\Phi, \mathbf{x})}^\perp(\mathbf{a}) = \mathbf{a} + \Phi^T(\mathbf{x} - \Phi \mathbf{a}). \quad (3.4)$$

This result can be interpreted in terms of adding to the vector  $\mathbf{a}$  the difference between the minimum Euclidean-norm solution,  $\mathbf{a}^{LS} = \Phi^T \mathbf{x}$ , and the analysis vector of the reconstruction using  $\mathbf{a}$  ( $\Phi^T \Phi \mathbf{a}$ ).

On the other hand, the orthogonal projection onto  $B_p(R)$ ,  $P_{B_p(R)}^\perp(\mathbf{a})$ , depends, obviously, on the value of  $p$ . Next, we explore in detail the cases  $p = 0$  and  $p = 1$ .

Finally, the  $\ell_p$ -AP method is implemented through the following iterations:

$$\begin{aligned} \hat{\mathbf{a}}^p(R)^{(0)} &= P_{B_p(R)}^\perp(\mathbf{a}^{LS}), \\ \hat{\mathbf{a}}^p(R)^{(k+1)} &= P_{B_p(R)}^\perp(P_{S(\Phi, \mathbf{x})}^\perp(\hat{\mathbf{a}}^p(R)^{(k)})). \end{aligned}$$

We have chosen to stop the iterations when  $\|\hat{\mathbf{a}}^p(R)^{(k+1)} - \hat{\mathbf{a}}^p(R)^{(k)}\|_2 < \delta$  for a  $\delta > 0$  (see implementation details in Section 3.3). Now we study the  $p = 0$  and  $p = 1$  cases in detail.

### 3.1.1 $\ell_0$ -AP

#### 3.1.1.1 Projection onto the $\ell_0$ -ball of given radius

When  $p = 0$ , it is straightforward to derive that  $P_{B_0(R)}^\perp(\mathbf{a})$  is a hard-thresholding preserving the  $R$  largest coefficients in amplitude:

$$P_{B_0(R)}^\perp(\mathbf{a}) = \mathbf{a}^h,$$

---

<sup>3</sup>In this case, it can eventually happen that the orthogonal projection onto the non-convex sets is not unique, but this is a theoretical problem without practical consequences.

where:

$$a_i^h = \begin{cases} a_i, & |a_i| > \theta_h(\mathbf{a}, R) \\ 0, & |a_i| \leq \theta_h(\mathbf{a}, R). \end{cases}$$

Here,  $\theta_h(\mathbf{a}, R)$  is the lowest threshold between those preserving the  $R - n_0$  largest coefficients in amplitude, being  $n_0$  the smallest non-negative integer guarantying that a solution exists. Thus,  $n_0 = 0$  if there are no repeated amplitudes in the interval of interest. Following the previous definition, in practice the threshold is set to the amplitude of the  $R + 1$ -th largest coefficient in amplitude in vector  $\mathbf{a}$ .

This method can be also seen as a particular case of the method described in [15], but with the difference that in that paper the method was not formally justified as an optimisation method.

### 3.1.1.2 $\ell_0$ -AP scheme and convergence

Top panel of Figure 3.1 shows an illustration of  $\ell_0$ -AP using few dimensions ( $N = 2$ ,  $M = 3$ ,  $R = 1$ ).

We will prove next that this method converges to a local minimum, in the image domain, of the MSE of the reconstruction for the vectors of the  $\ell_0$ -ball. Substituting following the Equation (3.4):

$$\|\mathbf{a} - P_{S(\Phi, \mathbf{x})}^\perp(\mathbf{a})\|_2 = \|\Phi^T(\mathbf{x} - \Phi\mathbf{a})\|_2 = \|\mathbf{x} - \Phi\mathbf{a}\|_2, \quad (3.5)$$

where the last step is true because  $\Phi^T$  is a Parseval frame. Given that  $\hat{\mathbf{a}}^0(R)$  is a local minimum in  $B_0(R)$  of the distance to  $S(\Phi, \mathbf{x})$ , then there exists a  $\delta > 0$  such that for all  $\mathbf{a} \in B_0(R)$ , if  $\|\mathbf{a} - \hat{\mathbf{a}}^0(R)\|_2 < \delta$ , then  $\|\mathbf{a} - P_{S(\Phi, \mathbf{x})}^\perp(\mathbf{a})\|_2 \geq \|\hat{\mathbf{a}}^0(R) - P_{S(\Phi, \mathbf{x})}^\perp(\hat{\mathbf{a}}^0(R))\|_2$ . Using (3.5) we have that  $\|\mathbf{x} - \Phi\mathbf{a}\|_2 \geq \|\mathbf{x} - \Phi\hat{\mathbf{a}}^0(R)\|_2$ . That is,  $\hat{\mathbf{a}}^0(R)$  is a local minimum in vector  $\mathbf{a}$  and within set  $B_0(R)$  of the Euclidean distance between  $\Phi\mathbf{a}$  and  $\mathbf{x}$ .

Regarding the convergence properties, we have observed that the method evolves fast towards the solution during the first iterations, and then the convergence speed decreases, as shown in Figure 3.2 for images *Barbara*, *Boat* and *House* from our test set (see Appendix A). In the figure we see that the convergence speed also depends on the degree of sparseness imposed (the sparser, the faster). In this Thesis, we are interested in exploring the performance of the methods at convergence, and this has required making thousands of iterations for each experiment. However, in a practical implementation, less iterations can be made for obtaining satisfactory results. In this work, we have established the stopping criterion based on the PSNR increase every 10 iterations. Dotted curve corresponds to the increase rate used as tolerance (it would be a straight line if the figure

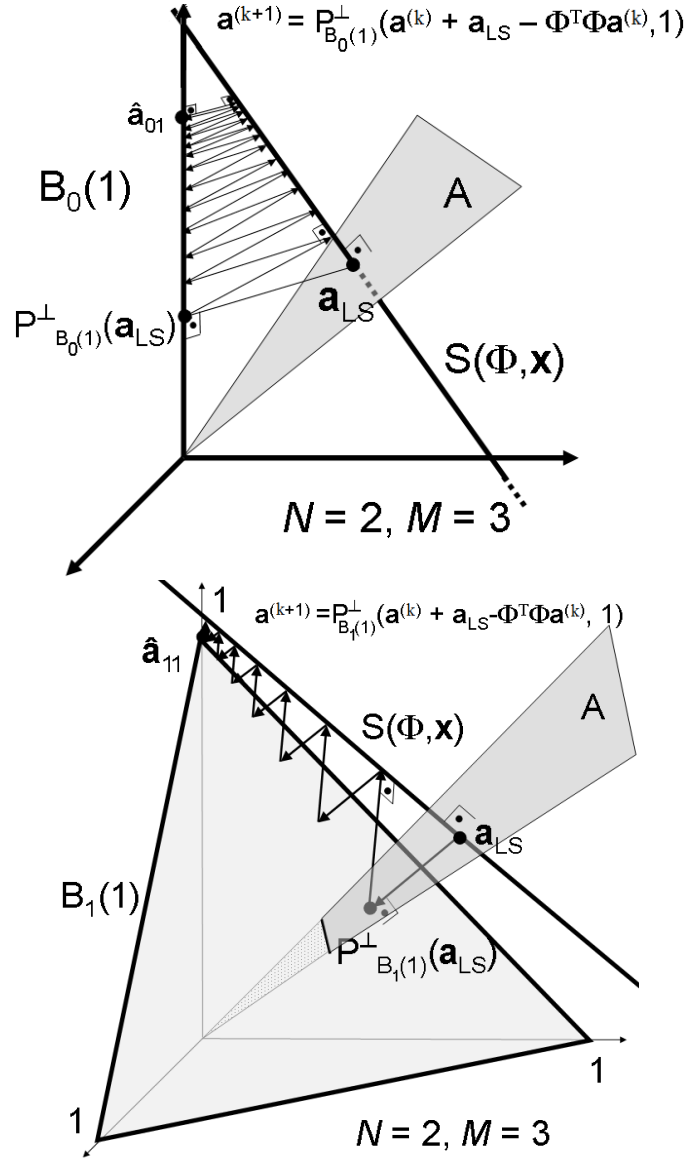


Figure 3.1: **Top**, graphical explanation of the  $\ell_0$ -AP method. **Bottom**, same for  $\ell_1$ -AP. Only a face of the ball is shown for clarity.

was not in logarithmic coordinates). In subsection 3.3.2 we will see more details about the stopping criterion.

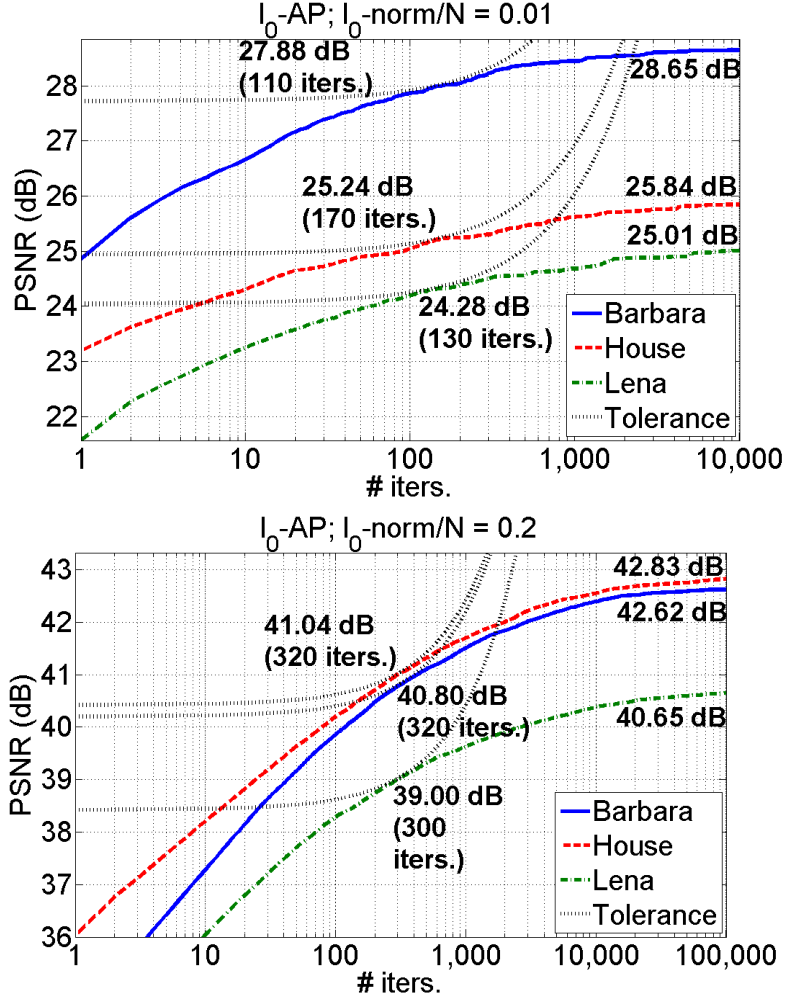


Figure 3.2: Logarithmic plot of the approximation quality (PSNR, in dB) vs. the number of iterations for  $l_0$ -AP for three images and two sparseness levels. The representation used here is DT-CWT. The number at the end of the curves is the PSNR at convergence. The numbers accompanying the tangency point (indicated by the intersection with the dotted curves) are the PSNR and the number of iterations obtained when the stopping criterion is reached.

### 3.1.2 $l_1$ -AP

#### 3.1.2.1 Projection onto the $l_1$ -ball of given radius

When  $p = 1$ , it can be proved that the orthogonal projection of a vector  $\mathbf{a}$  onto the  $l_1$ -ball of given radius  $R$ , that we denote as  $\mathbf{a}^s = P_{B_1(R)}^\perp(\mathbf{a})$ , is a soft-thresholding operation. This has been proved before, for example in [37, 38], but our alternative proof will also provide an iterative method to find the threshold associated to this operation.

We first assume that  $\|\mathbf{a}\|_1 > R$  (otherwise the projection onto  $B_1(R)$  would be the identity). In addition, we use the obvious fact that any projection onto a  $\ell_p$ -ball preserves the sign of the coefficients in the original vector ( $\text{sign}(\mathbf{a}^s) = \text{sign}(\mathbf{a})$ ). The problem is then reduced to project the vector formed by the components  $\{|a_1|, |a_2|, \dots, |a_M|\}$ , that we denote  $\mathbf{a}^{abs}$ , onto the positive hyper-quadrant of  $B_1(R)$ . Once we have that projection, we restore the sign of each element to obtain the projection of  $\mathbf{a}$  onto  $B_1(R)$ .

The positive hyper-quadrant of  $B_1(R)$  can be defined as the intersection of two convex sets. The first one is the set of all those vectors whose components sum, as much,  $R$ :

$$F(R) = \{\mathbf{b} \in \mathbb{R}^M : \sum_{i=1}^M b_i \leq R\}.$$

Given a vector  $\mathbf{c} \in \mathbb{R}^M$ , the expression of the orthogonal projection onto this set is:

$$P_{F(R)}^\perp(\mathbf{c}) = \mathbf{c} - \delta,$$

where  $\delta = \frac{\sum_{i=1}^M (c_i) - R}{M}$  if  $\sum_{i=1}^M c_i > R$  and 0 otherwise.

The second set is the positive hyper-quadrant of the  $M$ -dimensional vector space:

$$G^+ = \{\mathbf{b} \in \mathbb{R}^M : \forall i = \{1, \dots, M\}, b_i \geq 0\}.$$

The orthogonal projection onto this set is defined as:

$$P_{G^+}^\perp(\mathbf{c}) = \mathbf{D}\mathbf{c},$$

where  $\mathbf{D}$  is a diagonal  $M \times M$  matrix, such that  $d_{ii} = 1$  if  $c_i > 0$  and 0 otherwise.

Following the alternated projections theory, the orthogonal projection of a vector  $\mathbf{a}$  onto the intersection of  $F(R)$  and  $G^+$ , that we name  $\mathbf{a}^{pro} = P_{F(R) \cap G^+}^\perp(\mathbf{a})$ , is defined as:

$$\mathbf{a}^{pro} = \lim_{n \rightarrow \infty} [P_{G^+}^\perp(P_{F(R)}^\perp(\dots n \dots P_{G^+}^\perp(P_{F(R)}^\perp(\mathbf{a}^{abs})) \dots n \dots))]. \quad (3.6)$$

The orthogonal projection of  $\mathbf{a}$  onto  $B_1(R)$  is finally obtained as:

$$\mathbf{a}^s = \text{sign}(\mathbf{a}) \cdot \mathbf{a}^{pro}. \quad (3.7)$$

We prove next that the expression obtained in Equation (3.7) is a soft-thresholding. First, for  $k = \{1, \dots, n\}$ , we denote  $\delta^{(k)}$  the term substracted in

the  $k$ -th application of the orthogonal projection onto  $F(R)$  in the iterations of Equation (3.6). We also denote  $\mathbf{D}^{(k)}$  the mask applied in the  $k$ -th application of the orthogonal projection onto  $G^+$ . Then:

$$\mathbf{a}^{pro} = \lim_{n \rightarrow \infty} [\mathbf{D}^{(n)}(\dots \mathbf{D}^{(2)}(\mathbf{D}^{(1)}(\mathbf{a}^{abs} - \delta^{(1)}) - \delta^{(2)}) \dots - \delta^{(n)})].$$

This can be expressed as:

$$\mathbf{a}^{pro} = \mathbf{a}^{abs} - \mathbf{d}, \quad (3.8)$$

where each element  $d_i$  is defined as:

$$d_i = \begin{cases} \theta_s(\mathbf{a}, R), & |a_i| > \theta_s(\mathbf{a}, R) \\ |a_i|, & |a_i| \leq \theta_s(\mathbf{a}, R), \end{cases}$$

and where  $\theta_s(\mathbf{a}, R) = \sum_{k=1}^n \delta^k$ . Consequently, if we substitute the expression of Equation (3.8) into Equation (3.7), we obtain:

$$\mathbf{a}^s = \text{sign}(\mathbf{a}) \cdot (\mathbf{a}^{abs} - \mathbf{d}),$$

which is the definition of a soft-thresholding. That is,  $\mathbf{a}^s = S_1(\mathbf{a}, \theta_s(\mathbf{a}, R))$  (see Equation (2.7)).

Note that this proof provides a method based on alternating projections for, given  $\mathbf{a}$ , finding the value of the threshold that leads to the desired value of the  $\ell_1$ -norm after the shrinkage. This method starts by removing the sign from  $\mathbf{a}$ , projecting the result onto the intersection of  $F(R)$  and  $G^+$ , by using alternating projections, and finally restoring the original sign to every element of that projection. In practice, this method converges linearly in very few iterations. Next we develop a method where each iteration requires less calculation, so the final method is easier to implement.

Firstly, let's express the  $\ell_1$ -norm of the projected vector (that is,  $R$ ) as a function of the threshold<sup>4</sup>  $\theta_s(\mathbf{a}, R)$ . For that purpose, we define the set of indices corresponding to those coefficients in  $\mathbf{a}$  whose amplitudes are above a threshold  $\theta$ :  $\Upsilon(\mathbf{a}, \theta) = \{i \in \{1, \dots, M\} : |a_i| > \theta\}$ . Then, we can write:

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

---

<sup>4</sup>For notation clarity, in the following derivation we have removed the dependency of  $\theta_s$  upon  $\mathbf{a}$  and  $R$ .

where  $\text{card}(\cdot)$  indicates the cardinality of a set. We can express the previous iterations as:

$$\theta_s = \frac{\left( \sum_{\Upsilon(\mathbf{a}, \theta_s)} |a_i| \right) - R}{\text{card}(\Upsilon(\mathbf{a}, \theta_s))}. \quad (3.9)$$

Note that the right term depends on  $\theta_s$ , but nevertheless we can solve this equation iteratively using the following iterations:

$$\begin{aligned} \theta_s^{(0)} &= 0, \\ \theta_s^{(k+1)} &= \frac{\left( \sum_{\Upsilon(\mathbf{a}, \theta_s^{(k)})} |a_i| \right) - R}{\text{card}(\Upsilon(\mathbf{a}, \theta_s^{(k)}))}. \end{aligned} \quad (3.10)$$

The iterations end when  $\|\theta_s^{(k+1)} - \theta_s^{(k)}\|_2$  is below a tolerance threshold (see subsection 3.3.2 for more details on the stopping criterion).

Next, we prove that iterations (3.10) converge to  $\theta_s$ . We note first that  $R(\theta_s)$  is a strictly decreasing function and, then, so it is  $\theta_s(R)$ . This implies that Equation (3.9) has a unique solution in  $\theta_s$ . If we find  $\theta_s^{(k+1)} = \theta_s^{(k)}$  then Equation (3.9) is satisfied with that  $\theta$  value, so we know that if the iterations converge then they do to the unique solution,  $\theta_s$ . Thus, to prove the convergence to  $\theta_s$ , it is left to prove that the succession  $\theta_s^{(k)}$  converges. This can be made by proving that 1)  $\theta_s^{(k)}$  is monotonically increasing, and that 2) it is upper bounded by  $\theta_s$ . This is what we do next.

We start by observing that  $\theta_s^{(0)} = 0 \leq \theta_s$ . Assuming that  $\theta_s^{(k)} \leq \theta_s$ , then:

$$\sum_{\Gamma(\mathbf{a}, \theta_s^{(k)}, \theta_s)} |a_i| \leq \sum_{\Gamma(\mathbf{a}, \theta_s^{(k)}, \theta_s)} \theta_s,$$

where  $\Gamma(\mathbf{a}, \theta_1, \theta_2) = \{i \in \{1, \dots, M\} : \theta_1 < |a_i| \leq \theta_2\}$ . From here we obtain that:

$$\begin{aligned} \sum_{\Upsilon(\mathbf{a}, \theta_s^{(k)})} |a_i| - \sum_{\Upsilon(\mathbf{a}, \theta_s)} |a_i| &\leq \sum_{\Upsilon(\mathbf{a}, \theta_s^{(k)})} \theta_s - \sum_{\Upsilon(\mathbf{a}, \theta_s)} \theta_s, \\ \sum_{\Upsilon(\mathbf{a}, \theta_s^{(k)})} |a_i| - R &\leq \sum_{\Upsilon(\mathbf{a}, \theta_s^{(k)})} \theta_s, \\ \sum_{\Upsilon(\mathbf{a}, \theta_s^{(k)})} |a_i| - R &\leq \text{card}(\Upsilon(\mathbf{a}, \theta_s^{(k)})) \cdot \theta_s, \\ \frac{\sum_{\Upsilon(\mathbf{a}, \theta_s^{(k)})} |a_i| - R}{\text{card}(\Upsilon(\mathbf{a}, \theta_s^{(k)}))} &\leq \theta_s, \\ \theta_s^{(k+1)} &\leq \theta_s. \end{aligned}$$

Now we see that, as  $\sum_{\Gamma(\mathbf{a}, \theta_s^{(k)}, \theta_s)} |a_i| \geq \sum_{\Gamma(\mathbf{a}, \theta_s^{(k)}, \theta_s)} \theta_s^{(k)}$ , and as  $\sum_{\Upsilon(\mathbf{a}, \theta_s)} \theta_s \geq \sum_{\Upsilon(\mathbf{a}, \theta_s)} \theta_s^{(k)}$ , then:

$$\sum_{\Gamma(\mathbf{a}, \theta_s^{(k)}, \theta_s)} |a_i| + \sum_{\Upsilon(\mathbf{a}, \theta_s)} \theta_s \geq \sum_{\Upsilon(\mathbf{a}, \theta_s^{(k)})} \theta_s^{(k)},$$

and we have the following inequalities:

$$\begin{aligned} \sum_{\Upsilon(\mathbf{a}, \theta_s^{(k)})} |a_i| - \sum_{\Upsilon(\mathbf{a}, \theta_s)} |a_i| + \sum_{\Upsilon(\mathbf{a}, \theta_s)} \theta_s &\geq \sum_{\Upsilon(\mathbf{a}, \theta_s^{(k)})} \theta_s^{(k)}, \\ \sum_{\Upsilon(\mathbf{a}, \theta_s^{(k)})} |a_i| - R &\geq \sum_{\Upsilon(\mathbf{a}, \theta_s^{(k)})} \theta_s^{(k)}, \\ \sum_{\Upsilon(\mathbf{a}, \theta_s^{(k)})} |a_i| - R &\geq \text{card}(\Upsilon(\mathbf{a}, \theta_s^{(k)})) \cdot \theta_s^{(k)}, \\ \frac{\sum_{\Upsilon(\mathbf{a}, \theta_s^{(k)})} |a_i| - R}{\text{card}(\Upsilon(\mathbf{a}, \theta_s^{(k)}))} &\geq \theta_s^{(k)}, \\ \theta_s^{(k+1)} &\geq \theta_s^{(k)}. \end{aligned}$$

Consequently, the succession is monotonically increasing, so the proof is complete.

### 3.1.2.2 $\ell_1$ -AP scheme and convergence

Figure 3.1 (bottom panel) illustrates the behaviour of  $\ell_1$ -AP with  $N = 2$ ,  $M = 3$ , and  $R = 1$ . Only a face of  $B_1(1)$  is shown for visibility sake.

It is easy to prove that  $\ell_1$ -AP provides the global minimum for the distance, in the image domain, from the reconstruction from vectors in  $B_1(R)$  to image  $\mathbf{x}$ . We first note that  $\hat{\mathbf{a}}^1(R)$  is the global minimum in  $B_1(R)$  of the Euclidean distance to  $S(\Phi, \mathbf{x})$  (because both sets are convex). Then, for every  $\mathbf{a} \in B_1(R)$ , we have that  $\|\mathbf{a} - P_{S(\Phi, \mathbf{x})}^\perp(\mathbf{a})\|_2 \geq \|\hat{\mathbf{a}}^1(R) - P_{S(\Phi, \mathbf{x})}^\perp(\hat{\mathbf{a}}^1(R))\|_2$ . Applying the Equation (3.4) and being  $\Phi^T$  a Parseval frame, we obtain that  $\|\mathbf{x} - \Phi \mathbf{a}\|_2 \geq \|\mathbf{x} - \Phi \hat{\mathbf{a}}^1(R)\|_2$ . That is,  $\Phi \hat{\mathbf{a}}^1(R)$  is the global minimum, for all  $\mathbf{a} \in B_1(R)$ , of the Euclidean distance of  $\Phi \mathbf{a}$  to  $\mathbf{x}$ .

Figure 3.3 illustrates the convergence properties of  $\ell_1$ -AP. The interpretation is similar to that of Figure 3.2. As there are no local solutions to avoid, the convergence is more regular than using  $\ell_0$ -AP and fewer iterations are required to converge. We have included an example, in the bottom panel, where perfect reconstruction is achieved.



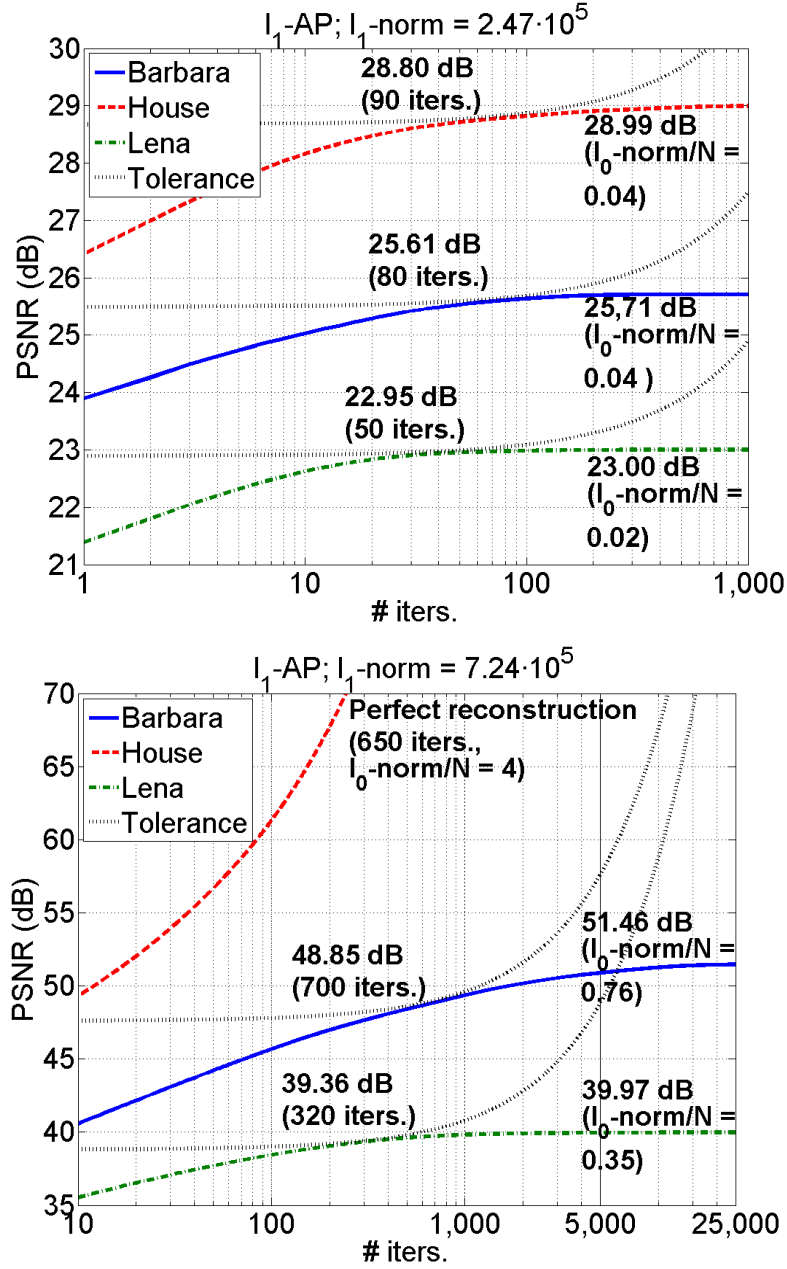


Figure 3.3: Convergence curves in semi-logarithmic scale for  $l_1$ -AP, using three images and two sparseness levels. Details are similar to Figure 3.2. It is also indicated the  $l_0$ -norm, normalised by  $N$ , of the solution at convergence.

### 3.2 Mean square error minimisation for a given selection of coefficients

As  $l_0$ -AP iterations are performed, the selection of coefficients becomes more and more stable, and the final solution becomes LS-optimal for that

selection<sup>5</sup>. As it was already pointed in [15], when, in the limit, the number of active coefficients becomes fixed, the two involved sets (vector subspace generated by the selected atoms and affine subspace of perfect reconstruction) are convex, and so the iterations converge to the global optimum, for that selection, linearly.

However, this is not true when we use a generic  $\ell_p$ -norm, because the projection onto the  $\ell_p$ -ball is not, in general, a hard-thresholding operation. As our final target is to solve the sparse approximation problem of Equation (2.4), we should use some method in order to improve the quality of the approximation for a given set of selected functions. We use here a method based on alternating projections which has been previously used by a number of authors, as [11, 13, 24, 32, 96].

Given a set  $I$  of  $R$  indices extracted from  $\{1, \dots, M\}$ , we define  $\Phi_I$  as a  $N \times R$  matrix formed by columns  $\phi_i$  of  $\Phi$  such that  $i \in I$ . Then, we want to find:

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

which translates into  $\hat{\mathbf{a}}_I = \Phi_I^\# \mathbf{x}$ , where  $\Phi_I^\#$  is the pseudo-inverse of  $\Phi_I$ . Note that  $\Phi_I^\# = \Phi_I^T [\Phi_I \Phi_I^T]^{-1}$  if  $R > N$  and  $\Phi_I^\# = [\Phi_I^T \Phi_I]^{-1} \Phi_I^T$  if  $R \leq N$ . When dealing with images, the size of  $\Phi_I$  makes the calculation of the pseudo-inverse a completely intractable task. Instead, we follow these iterations:

$$\begin{aligned} \mathbf{a}^{(0)} &= \mathbf{D}_I \Phi^T \mathbf{x}, \\ \mathbf{a}^{(k+1)} &= \mathbf{D}_I [\mathbf{a}^{(k)} + \Phi^T (\mathbf{x} - \Phi \mathbf{a}^{(k)})], \end{aligned} \quad (3.11)$$

where  $\mathbf{D}_I$  is a diagonal  $M \times M$  matrix such that  $d_{ii} = 1$  if  $i \in I$  and 0 otherwise. In Appendix B we show that this method effectively solves for the pseudo-inverse in  $\hat{\mathbf{a}}_I = \Phi_I^\# \mathbf{x}$ .

## 3.3 Implementation

### 3.3.1 Representations

To test the methods, we initially used four different Parseval frames (DT-CWT [43], Curvelets [114], Steerable Pyramid [7] and a redundant version of the Haar Wavelets [115]). From them, we chose those two giving the best

---

<sup>5</sup>Note that the method is still suboptimal because the selection of elementary functions is not optimal in general.

averaged compaction performance. These are DT-CWT and Curvelets<sup>6</sup>. Redundancy factor for DT-CWT is 4, and for Curvelets is  $\approx 7.2$ .

In order to make a homogeneous treatment of the two representations, we have divided DT-CWT coefficients in real and imaginary parts. On the other hand, to optimise the approximation in the extremely high sparseness range, we have inserted, in both representations, an extra scale composed only by the global mean of the image. This way we adapt to the fact that, usually, the best approximation to a natural image, using only one coefficient, is the global mean.

MATLAB® code for DT-CWT is available in [116]. We have also used MATLAB® code for the Curvelets implementation (*CurveLab 2.0* [117]).

### 3.3.2 Convergence and stopping criterion

The stopping criterion for  $\ell_p$ -AP, in our implementation, is based on using two tolerance constants. The first one controls the PSNR increase in order to decide if convergence has been reached. The method stops when the increase after 10 iterations is below 0.02 dB. This stopping criterion is represented as dotted curves in Figures 3.2 and 3.3. These curves would be straight and tangents to the convergence curves if the horizontal axis was plotted in a linear scale. We have experienced that this criterion, typically, provides differences with respect to the PSNR at convergence below 1 dB in the high sparseness range and below 2 dB in the low sparseness range. These differences are even lower for  $\ell_1$ -AP (favouring, thus, to this method in their comparison).

Note that, if the radius of the  $\ell_p$ -ball is big, the method achieves perfect reconstruction of the image. In this case, the increase of PSNR is, as stated by theory, linear. To detect this situation we have used a second tolerance criterion, controlling the PSNR increase after every 10 iterations and stopping them when the difference between the last two increases is below a constant ( $10^{-6}$  for  $\ell_0$ -AP and  $10^{-4}$  for  $\ell_1$ -AP).

We optimise the threshold in each  $\ell_0$ -AP iteration through a golden search. This requires an extra tolerance parameter controlling the size of the search interval, which we have set to 0.1. In  $\ell_1$ -AP, we have used the method described in subsection 3.1.2, taking for the stopping criterion the difference between the desired radius for  $B_1(R)$  and that obtained after each iteration. We have experienced that, eventually, this iterative method provides exactly the radius required. However, to reduce computation in practice, we have also chosen 0.1 for this tolerance.

---

<sup>6</sup>The experiments were actually made with all the representations, and all the results are qualitatively similar to those presented here.

## 3.4 Results and discussion

In the following experiments we compare the compaction performance of our methods with respect to some reference algorithms in the field. Results have been collected for a wide sparseness range, and for our set of standard test images. We have used PSNR to measure the approximation error. We have used a logarithmic scale for the vertical axis, though PSNR is already a logarithmic measure. This may be unusual, but we think that it is justified, in this case, because of the great improvement achieved in the visualisation. Regarding the sampling of the curves, each marker corresponds to averaging the results of the corresponding method for all test images, and the intermediate values have been linearly interpolated.

### 3.4.1 Some previous methods

Our first experiment compares some widely used sparse approximation strategies. We have the following two objectives: a) compare Iterative Shrinkage Methods, with fixed threshold, in their two variants: hard and soft-thresholding; and b) compare the direct and accumulative strategies for selecting coefficients in greedy methods.

Regarding the former, we have implemented the methods IHT and IST, as described in subsection 2.3.3. We remind that these methods iterate between a thresholding operation and the projection onto the affine space of perfect reconstruction (Equation (3.4)), using a fixed threshold. We have used the same stopping criterion as in our implementation of  $\ell_p$ -AP (see subsection 3.3.2). Thus, our implementation of these methods only differs from our implementation of  $\ell_p$ -AP in using a fixed threshold instead of a fixed number of non-zero coefficients after each thresholding.

To compare greedy heuristics, we have implemented StOMP [23] and the method we presented in [24], called here DT+OP (from Direct Thresholding plus Least Squares-Optimisation). To choose the threshold used by StOMP, we previously set how many coefficients will be selected at each iteration. On the other hand, DT+OP applies the threshold directly (and only once) upon the linear representation of the image for each sample. Both methods use Equations (3.11) to LS-optimize the quality of the reconstruction after the thresholding. Here, we have also used the same stopping criteria as for  $\ell_p$ -AP.

Figure 3.4 shows graphically some numerical results of this experiment. Top panel shows fidelity results using 8-scale DT-CWT, and bottom one using 6-scale Curvelets. This figure shows that the compaction performance of IHT is, on average, clearly better than IST for medium-high levels of sparseness. In low levels, the number of local minima is considerably

increased and therefore IHT has more probabilities of getting trapped in non-favorable local minima. Figure also shows that the results obtained with DT+OP outperform our implementation of StOMP, except in the very high sparseness range (of little practical relevance). This indicates that directly selecting the coefficients is better than accumulating them by means of the correlation with the residual. Among the compared methods, IHT provides the best results. Previously, some authors have also pointed that hard thresholding outperforms soft one when compacting the energy [78, 89, 90, 16], but no careful and systematic comparisons using natural images were presented.

### 3.4.2 Comparison of $\ell_0$ -AP, $\ell_1$ -AP and previous methods

This second experiment compares  $\ell_0$ -AP to  $\ell_1$ -AP. Because the result of  $\ell_1$ -AP is not LS-optimal for the selection of atoms from the dictionary, we also compare with the result of de-biasing these coefficients with Equations (3.11). We label this method as  $\ell_1$ -AP+OP. In addition, we have included IHT and StOMP as representatives of the iterative shrinkage and greedy methods, respectively.

Figure 3.5 shows the result of this experiment. Top panel shows fidelity results with 8-scale DT-CWT, and bottom one with 6-scale Curvelets. We can see that  $\ell_0$ -AP clearly outperforms  $\ell_1$ -AP, even though the latter is optimally minimising the  $\ell_1$ -norm for each sparseness level. We also see that  $\ell_1$ -AP+OP improves drastically the results of  $\ell_1$ -AP, providing slightly better results than  $\ell_0$ -AP. This shows that the selection of coefficients made by  $\ell_1$ -AP is slightly better, in general, than that of  $\ell_0$ -AP, especially in the low sparseness range. This, as before, seems to be a natural consequence of  $\ell_0$ -AP getting trapped in non-favourable local minima, whose number rapidly increases when the sparseness gets low.

We can also see that  $\ell_0$ -AP significantly improves the results of IHT and StOMP. It is interesting to note that fixing the radius of the  $\ell_p$ -ball provides much better results than fixing the threshold.

Tables 3.1 and 3.2 show the results plotted in Figure 3.5.

Figure 3.6 compares visually the methods using the *Einstein* image<sup>7</sup> using  $0.0765 \cdot N$  Curvelets coefficients. From top to bottom, left column shows the original image, the result of  $\ell_1$ -AP (30.85 dB) and that of  $\ell_1$ -AP+OP (33.52 dB). Note the great visual improvement obtained when

<sup>7</sup>For every experiment in this Thesis using *Einstein* image, we have removed its black border, by replication of adjacent rows and columns. This makes it a more representative natural image.

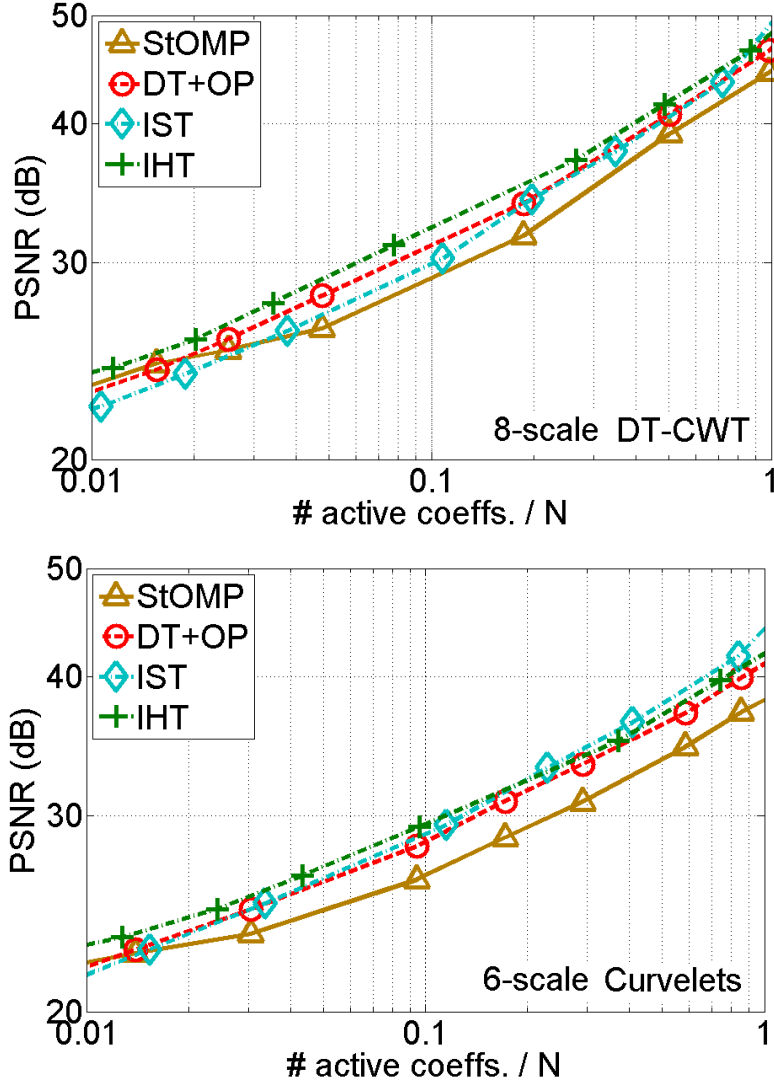


Figure 3.4: Averaged compaction results (fidelity of the approximation as PSNR, in dB) for our test set using StOMP, DT+OP, IHT and IST. **Top**, using 8-scale DT-CWT. **Bottom**, using 6-scale Curvelets.

post-optimising the selected coefficients. Right column shows StOMP (28.66 dB), IHT (29.10 dB) and  $\ell_0$ -AP (32.98 dB). Although more than half dB below in terms of PSNR, there is no significant visual difference between  $\ell_0$ -AP and  $\ell_1$ -AP+OP in this example. Not shown here, this difference becomes even smaller for lower PSNR values.

As we have already pointed,  $\ell_0$ -AP is equivalent to [15] when a fixed number of coefficients, and no extra heuristics, are used. They also use DT-CWT, but they apply the threshold to the magnitudes of complex

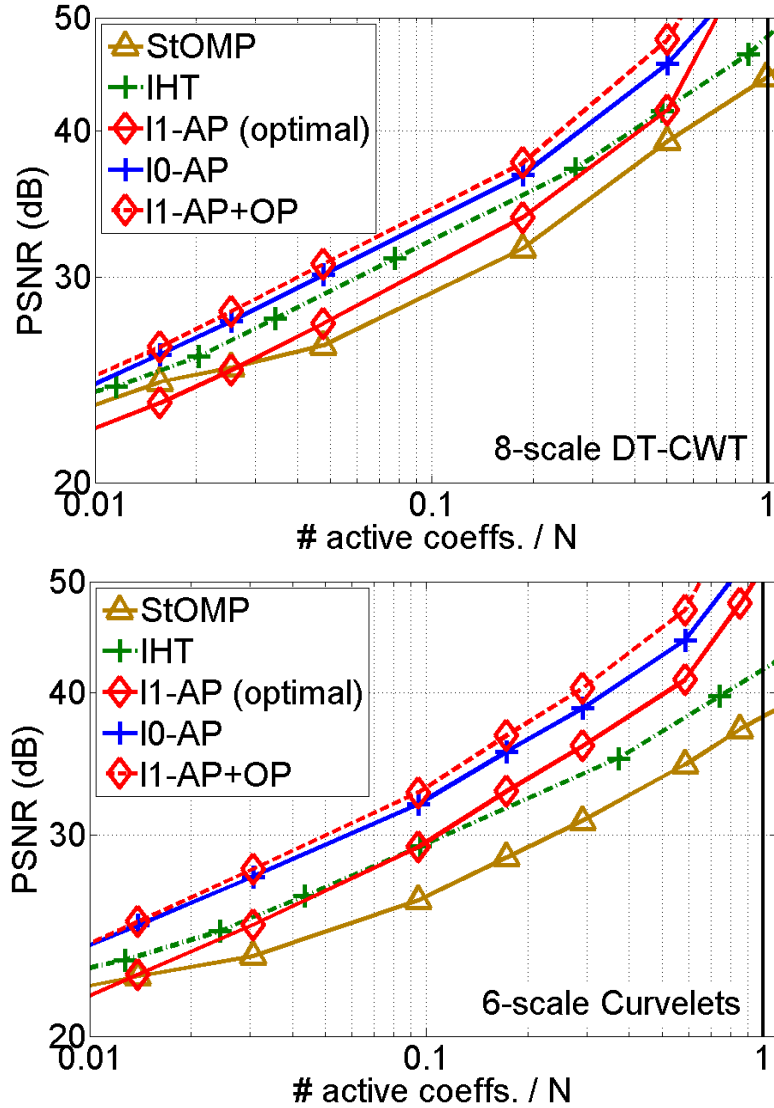


Figure 3.5: *Compaction results, averaged in our test set, of methods  $\ell_0$ -AP,  $\ell_1$ -AP,  $\ell_1$ -AP+OP, IHT and DT+OP. **Top**, using 8-scale DT-CWT. **Bottom**, using 6-scale Curvelets.*

coefficients. Then, to compare properly their results with ours, we have doubled the given number of selected coefficients in their results. We used 5-scale DT-CWT, as these authors do, and using 24000 fixed coefficients and the  $512 \times 512$  *Lena* image<sup>8</sup>, obtaining a result 2.02 dB above theirs (39.09 vs. 37.07 dB).

In [15] it is also presented a dynamic version increasing the number of

<sup>8</sup>We thank Prof. Kingsbury for helping us replicating their experiments.

DT-CWT/Curvelets		$R/N$		
Image	Method	0.00305	0.00944	0.02914
Barbara	StOMP	25.75/24.02	27.91/27.26	35.99/32.56
	IHT	28.18/25.98	33.25/30.51	40.58/34.30
	$\ell_1$ -AP	26.39/25.75	30.21/31.02	38.65/38.38
	$\ell_0$ -AP	<i>29.23/28.61</i>	<i>33.38/34.29</i>	<i>41.76/41.51</i>
	$\ell_1$ -AP+OP	<b>29.95/29.10</b>	<b>34.12/34.94</b>	<b>43.09/43.08</b>
House	StOMP	28.64/25.52	30.81/29.23	37.88/34.22
	IHT	31.19/28.64	34.76/32.69	40.25/37.77
	$\ell_1$ -AP	29.60/27.98	32.79/33.32	39.56/39.35
	$\ell_0$ -AP	<i>32.09/30.45</i>	<i>35.18/35.63</i>	<i>43.00/41.87</i>
	$\ell_1$ -AP+OP	<b>32.61/31.19</b>	<b>35.79/36.69</b>	<b>44.41/43.78</b>
Boat	StOMP	24.17/22.46	26.35/24.70	32.53/29.23
	IHT	25.46/24.07	30.07/27.62	34.69/31.17
	$\ell_1$ -AP	24.10/23.41	27.53/27.00	34.56/33.62
	$\ell_0$ -AP	<i>26.48/26.15</i>	<i>30.43/29.97</i>	<i>38.00/36.73</i>
	$\ell_1$ -AP+OP	<b>26.92/26.19</b>	<b>31.05/30.24</b>	<b>39.08/38.02</b>
Lena	StOMP	24.95/23.10	27.31/25.88	34.42/30.46
	IHT	27.09/24.86	<i>32.17/28.63</i>	39.38/32.54
	$\ell_1$ -AP	25.50/24.78	28.92/28.72	36.89/35.69
	$\ell_0$ -AP	<i>27.72/27.10</i>	<i>31.63/31.27</i>	<i>40.29/38.66</i>
	$\ell_1$ -AP+OP	<b>28.49/27.64</b>	<b>32.50/32.15</b>	<b>41.41/40.40</b>
Peppers	StOMP	24.48/22.97	26.85/25.66	33.14/29.69
	IHT	25.80/24.53	<i>31.57/28.41</i>	38.43/32.28
	$\ell_1$ -AP	24.36/24.47	28.46/28.80	35.99/34.89
	$\ell_0$ -AP	<i>27.43/26.85</i>	<i>31.47/30.88</i>	<i>38.81/37.35</i>
	$\ell_1$ -AP+OP	<b>27.82/27.43</b>	<b>32.26/32.13</b>	<b>40.12/39.17</b>

Table 3.1: Detailed comparison of the methods using 8-scale DT-CWT and 6-scale Curvelets in our test set. Bold numbers indicate the method providing the best fidelity results for each image and sparseness level. Cursive numbers indicate the second best. Each column corresponds to a number of selected coefficients, whose value is normalised by  $N$ . The precise sparseness value for each normalised row correspond respectively to 2001, 6189 and 19096 active coefficients. We have extracted directly the PSNR values from the experiments, except for IHT, where the values have been linearly interpolated.

active complex coefficients used at each iteration (from 12000 to 24000 in 30 iterations in the experiment they describe). In this case they obtain 38.68 dB in the approximation, still 0.41 dB below our result. However, it is easy to check that this difference is caused by the extra flexibility of our



DT-CWT/Curvelets		$R/N$		
Image	Method	0.05851	0.08552	1.4873
Barbara	StOMP	42.88/36.91	45.02/38.56	49.86/43.03
	IHT	44.73/39.69	47.69/42.78	53.44/47.16
	$\ell_1$ -AP	45.39/44.16	50.24/50.20	> <b>100</b> /> <b>100</b>
	$\ell_0$ -AP	<del>48.31/47.37</del>	<del>51.76/53.00</del>	<del>61.73/64.86</del>
	$\ell_1$ -AP+OP	<b>52.03/51.33</b>	<b>56.88/61.48</b>	> <b>100</b> /> <b>100</b>
House	StOMP	42.78/37.52	45.49/41.29	51.65/44.32
	IHT	45.85/42.02	50.01/43.95	55.98/47.94
	$\ell_1$ -AP	46.22/43.78	50.09/50.94	> <b>100</b> /> <b>100</b>
	$\ell_0$ -AP	<del>50.92/46.52</del>	<del>54.56/53.96</del>	<del>67.38/63.13</del>
	$\ell_1$ -AP+OP	<b>53.18/49.17</b>	<b>57.05/60.94</b>	> <b>100</b> /> <b>100</b>
Boat	StOMP	37.73/32.67	40.20/35.00	45.56/38.01
	IHT	40.61/36.07	44.29/38.27	52.50/42.61
	$\ell_1$ -AP	40.71/38.73	45.86/45.29	> <b>100</b> /57.46
	$\ell_0$ -AP	<del>45.50/42.54</del>	<del>50.22/49.66</del>	<del>63.29/58.70</del>
	$\ell_1$ -AP+OP	<b>47.76/45.14</b>	<b>52.90/55.64</b>	> <b>100</b> / <b>71.97</b>
Lena	StOMP	41.07/34.39	43.31/36.93	48.48/40.95
	IHT	43.65/38.07	46.15/41.61	51.57/45.47
	$\ell_1$ -AP	43.71/41.30	48.14/49.70	> <b>100</b> /> <b>100</b>
	$\ell_0$ -AP	<del>47.54/44.74</del>	<del>51.16/51.02</del>	<del>61.60/60.94</del>
	$\ell_1$ -AP+OP	<b>50.67/47.47</b>	<b>55.09/58.18</b>	> <b>100</b> /> <b>100</b>
Peppers	StOMP	39.10/33.41	41.91/36.31	47.61/39.61
	IHT	42.38/37.22	45.52/40.31	51.60/44.74
	$\ell_1$ -AP	41.92/39.75	49.84/46.17	> <b>100</b> /58.62
	$\ell_0$ -AP	<del>45.76/43.04</del>	<del>52.13/50.40</del>	<del>63.78/61.45</del>
	$\ell_1$ -AP+OP	<b>48.12/45.66</b>	<b>56.05/56.30</b>	> <b>100</b> / <b>72.91</b>

Table 3.2: Continuation of Table 3.1. Sparseness values in the normalised columns correspond, respectively, to 38342, 56048 and 97471 active coefficients.

implementation which activate independently the real and imaginary parts of the complex coefficients. Actually, if we use non-separated coefficients in our implementation of  $\ell_0$ -AP, then our result is 1.31 dB *below* their dynamic version (37.37 vs. 38.68 dB). Better results can be achieved by using dynamic threshold and separation of real and imaginary part (see Chapter 4). The method presented in [85] improves heuristically the results of  $\ell_0$ -AP.

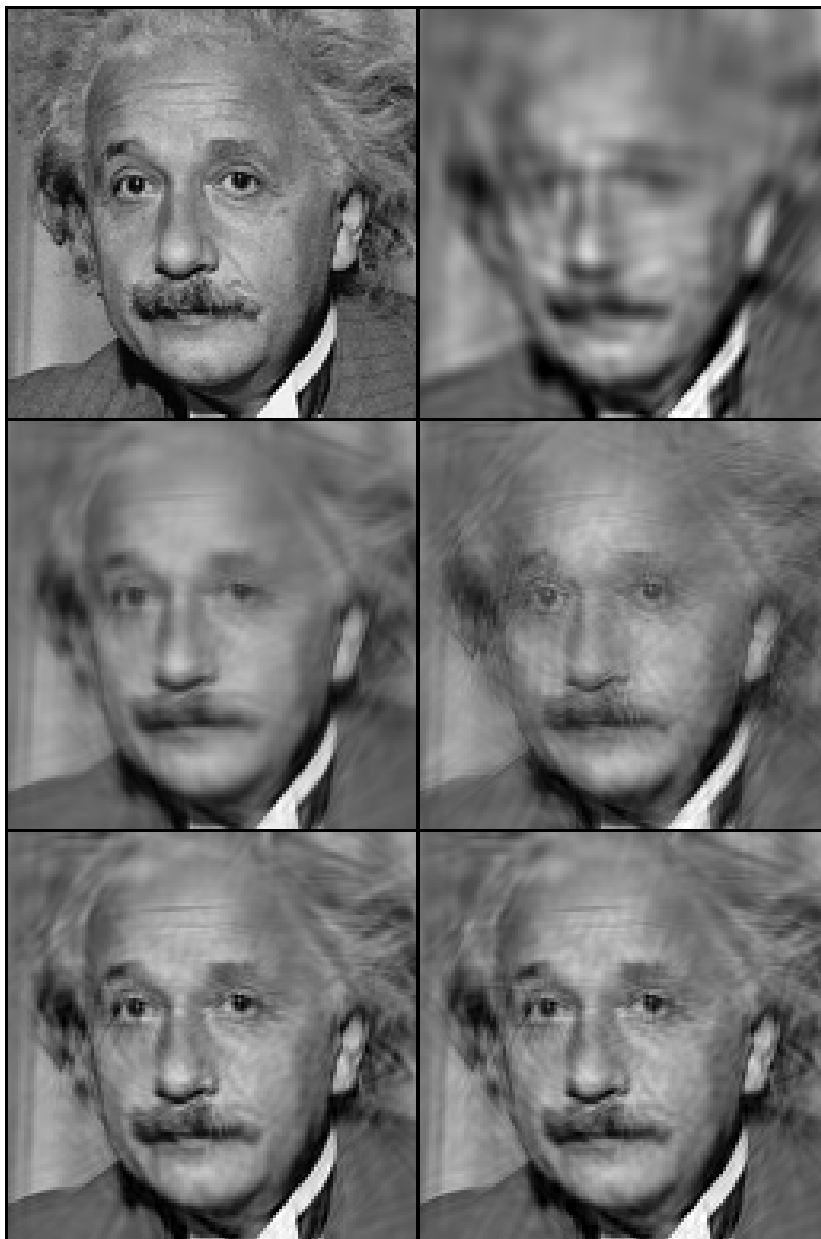


Figure 3.6: Visual comparison of the methods using  $0.0765 \cdot N$  Curvelets coefficients and the *Einstein* image, , where  $N$  is the number of pixels in the image. Results are cropped to  $128 \times 128$ , starting at pixel (71, 41), to improved the visibility. **Left column**, from top to bottom: original image and results of  $\ell_1$ -AP (30.85 dB) and  $\ell_1$ -AP+OP (33.52 dB). **Right column**, from top to bottom: results from StOMP (28.66 dB), IHT (29.10 dB) and  $\ell_0$ -AP (32.98 dB).

### 3.4.3 Computational load

The time per iteration is dominated in all the methods by one analysis and one synthesis operation. In addition to this, the search for the threshold in

$\ell_p$ -AP also takes a significant amount of time. Other methods like DT+OP and IHT do not require a threshold search, so they are relatively faster. Even so, the time consumed by the methods depends more critically on the number of iterations before reaching the stopping criterion. Table 3.3 shows that  $\ell_0$ -AP requires more iterations than  $\ell_1$ -AP. This difference is partly due to the tolerance used to detect the perfect reconstruction is tighter for  $\ell_0$ -AP than for  $\ell_1$ -AP (see subsection 3.3.2).

It is important to note, as we did in Section 3.1, that most of the real applications do not require so many iterations as shown in these experiments. In this chapter we did not aim to achieve a good compromise between performance and computation time, but we want to explore the quality ceiling of each method to appropriately compare them. However, as we have also experienced (see Chapter 4), methods based on dynamic thresholding (e.g., [93, 13, 15, 19, 38]) are intrinsically faster than those based on a fixed threshold or a fixed number of selected coefficients.

For our experiments, we have used an Intel® Core™2 Duo processor, with 1.66 GHz and with 2 GB RAM. As examples of execution time over  $256 \times 256$  images,  $\ell_0$ -AP takes around 7 minutes using DT-CWT and around 1 hour using Curvelets. On the other hand,  $\ell_1$ -AP takes around 3 minutes using DT-CWT and 30 minutes using Curvelets. Again, these running times are not representative of a real application, for which much fewer iterations would be applied.

Methods	# Iterations	
	DT-CWT	Curvelets
IHT	180	231
DT+OP	188	174
$\ell_1$ -AP	263	360
$\ell_1$ -AP+OP	333	440
$\ell_0$ -AP	495	920

Table 3.3: Averaged number of iterations in our test set using 8-scale DT-CWT and 6-scale Curvelets, for the different methods compared.

## 3.5 Conclusions

In this chapter, we have presented an optimisation method, which we call  $\ell_p$ -AP, based on minimising the MSE of the reconstruction of an image using a Parseval frame and given a maximum  $\ell_p$ -norm for that vector in

that representation. Given  $p$  and  $R$ , the method consists of alternatively orthogonally projecting between the  $\ell_p$ -ball of radius  $R$ , centred at the origin, and onto the set of vectors reconstructing perfectly the image. A global optimum is achieved when  $p \geq 1$ , and a local one when  $0 \leq p < 1$ . We have applied this method to the sparse approximation problem. We have focused on  $p = 0$  and  $p = 1$ . The case of  $\ell_0$ -AP translates into a heuristical algorithm previously proposed in [15]. On the other hand,  $\ell_1$ -AP is similar to the method, developed in a parallel and independent work, in [38].

Through systematic experiments, we have shown that  $\ell_0$ -AP clearly outperforms  $\ell_1$ -AP in terms of energy compaction of natural images using widely used pyramidal representations, despite  $\ell_1$ -AP being optimal for the convex relaxation problem. Moreover, this behaviour is consistent throughout the representations studied. This result shows that the conditions for achieving a global optimum to the sparse approximation problem by using convex relaxation are not held when using natural image and typical representations. Nevertheless, we can improve  $\ell_1$ -AP results by LS-optimising *a posteriori* the amplitudes of the selected coefficients. Applying this, we have shown that the selection of coefficients made by  $\ell_1$ -AP is slightly superior to that of  $\ell_0$ -AP. In the next chapter, however, we will show that this selection is still far from optimal.

We have also compared iterative to shrinkage methods based on fixed thresholds and greedy strategies, showing that  $\ell_0$ -AP also outperforms IHT, IST, our implemented version of StOMP, and DT+OP. We would need more intensive test to establish the superiority of  $\ell_0$ -AP over greedy methods in general, but the huge computational effort required by more strict greedy algorithms prevented us from doing this comparison. Among the methods mentioned before (excluding ours), we have seen that IHT achieves the best compaction results.

Although not compared in detail here, but in the next chapter, methods based on dynamically adjusting the threshold through iterations provide, until this date, the best compaction performance. But, up to now, these methods have not been mathematically formulated, as we have done here with  $\ell_p$ -AP. It is easy to adapt our method to iteratively increase the number of selected coefficients (as in [15]). An additional fact is that, for some restoration tasks (as, for example, spatial quantisation artifacts, see Chapter 6) we have experienced that not always sparser solutions are used.

Another additional advantage of our method is that we use less parameters than other similar ones [13, 15, 19]. However, it still requires to establish a radius for the  $\ell_p$ -ball. This disadvantage is overcome in the method proposed in next chapter.

## Chapter 4

# Sparse approximation using gradient descent

In this chapter we mathematically derive another method to solve the sparse approximation problem. It is more accurate and efficient than the one described in the previous chapter, but it maintains the advantage of being a solution to an explicit optimisation problem. It is designed from the next question: Is it possible to include gradient descent in the criterion to be minimised in the Equation (2.4)? The answer, due to the discontinuous nature of the  $\ell_0$ -norm, is "not directly". However, we will write an equivalent continuous criterion which allows to calculate the gradient direction. Then we will obtain a generalised version of IHT and proof that the fixed point of its iterations is a local minimum of the cost function at hand.

Moreover, to avoid getting trapped by unfavourable local minima, we will apply a deterministic annealing technique similar to other non-convex global optimisation algorithms [118, 119, 19]. We name the resulting method  $\ell_0$ -GM. We show through experiments that  $\ell_0$ -GM is competitive with current state-of-art in terms of energy compaction, outperforming both  $\ell_0$ -AP and our LS-optimized version of  $\ell_1$ -AP ( $\ell_1$ -AP+OP).

We derive analogously the IST method throughout the gradient descent in a function equivalent to the criterion to be minimised in the convex relaxation problem (Equation (2.5)). We also derive a convex variant of  $\ell_0$ -GM, which we name  $\ell_1$ -GM. We will show that it achieves comparable results to other convex relaxation methods, and we will describe the practical cases where it should be used.

We have noted that  $\ell_0$ -GM is a dynamic thresholding method. The idea of decreasing the threshold as iterations are executed is not new [13, 15, 19, 89, 90, 17]. Nevertheless, up to our knowledge, this is the first time that it

has been formally derived as a direct solution to the sparse approximation problem. In addition, nobody has analysed, in a certain depth, the reasons why this solution behaves so well.

We start by reformulating the sparse approximation cost function (discontinuous and unconstrained) in a continuous and constrained form (Section 4.1). Then, in Section 4.2 we derive the generalised IHT as local solution to the sparse approximation problem. We justify the use of a decreasing threshold in Section 4.3. In Section 4.4 implementation details of  $\ell_0$ -GM are given, and in Section 4.5 we compare the energy compaction capacity of  $\ell_0$ -GM to the methods studied in previous chapter. Finally, we derive the IST and  $\ell_1$ -GM methods in Section 4.6, and compare them to  $\ell_1$ -AP. Section 4.7 concludes this chapter.

## 4.1 An alternative formulation with a continuous cost function

We repeat here, for convenience, the sparse approximation problem formulation of Equation (2.4):

$$\hat{\mathbf{a}}^0(\lambda) = \arg \min_{\mathbf{a}} \{\|\mathbf{a}\|_0 + \lambda \|\Phi \mathbf{a} - \mathbf{x}\|_2^2\}. \quad (4.1)$$

The associated cost function is not only non-convex, but it is also discontinuous. This prevents a direct calculation of its gradient. Next we derive a new equivalent continuous and constrained function, whose gradient can be calculated. We start from the following formulation:

$$(\hat{\mathbf{a}}, \hat{\mathbf{b}}) = \arg \min_{\mathbf{a}, \mathbf{b}} \{\|\mathbf{a}\|_0 + \lambda \|\mathbf{b} - \mathbf{a}\|_2^2 \text{ s.t. } \Phi \mathbf{b} = \mathbf{x}\}, \quad (4.2)$$

and prove the equality  $\hat{\mathbf{a}} = \hat{\mathbf{a}}^0(\lambda)$ . Firstly we express Equation (4.2) as:

$$\hat{\mathbf{a}} = \arg \min_{\mathbf{a}} \{\|\mathbf{a}\|_0 + \lambda \min_{\mathbf{b}} \{\|\mathbf{b} - \mathbf{a}\|_2^2 \text{ s.t. } \Phi \mathbf{b} = \mathbf{x}\}\}. \quad (4.3)$$

Note that the inner minimisation is the orthogonal projection of  $\mathbf{a}$  onto the affine subspace  $S(\Phi, \mathbf{x})$  of perfect reconstruction of  $\mathbf{x}$ . This projection was already defined in Equation (3.4). We repeat its expression here for convenience:

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

Substituting it in Equation (4.2) we obtain:

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

Given that  $\Phi$  is a Parseval frame, it finally yields:

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

as we wanted to prove. In the next step, in order to obtain a continuous and constrained cost function only depending on  $\mathbf{b}$ , we start from Equation (4.2) and swap the minimization variables with respect to Equation (4.3):

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

It is easy to see that, in this case, minimising this cost function for vector  $\mathbf{a}$  is equivalent to minimising independently for each index. We express the cost as  $c(\mathbf{a}, \mathbf{b}) = \sum_{i=1}^M c'(a_i, b_i)$ , where:

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

Given  $\mathbf{b}$ , it is easy to see that if the value  $\tilde{a}_i(b_i)$  minimising  $c'(a_i, b_i)$  is not zero, then  $\tilde{a}_i(b_i) = b_i$ , and  $c'(\tilde{a}_i(b_i), b_i) = 1$ . Then, we have:

$$c(\tilde{\mathbf{a}}(\mathbf{b}), \mathbf{b}) = \sum_{i=1}^M \min(1, \lambda b_i^2).$$

Figure 4.1 shows a one-dimensional illustration of this minimum (with  $\lambda = 1$ ). Given some  $\lambda$  value, we note  $\theta$  the value holding  $\lambda\theta^2 = 1$ . Therefore:

$$\theta = \lambda^{-\frac{1}{2}},$$

and we have that:

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

This is a hard-thresholding operation with threshold  $\theta$ , which we note  $\tilde{\mathbf{a}}(\mathbf{b}) = S_0(\mathbf{b}, \theta)$ . Substituting  $S_0(\mathbf{b}, \theta)$  for  $\mathbf{a}$ , in Equation (4.2):

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

When evaluating this criterion for each coefficient in  $\mathbf{b}$ , one of the two terms (fidelity or sparseness) is zero. Thus, we can express the same as:

$$\begin{aligned} \hat{\mathbf{b}} &= \arg \min_{\mathbf{b}} \{C_0(\mathbf{b}, \theta) \text{ s.t. } \Phi \mathbf{b} = \mathbf{x}\}, \\ \hat{\mathbf{a}} &= S_0(\hat{\mathbf{b}}, \theta), \end{aligned} \tag{4.4}$$

where:

$$C_0(\mathbf{b}, \theta) = \sum_{i=1}^M \min \left( 1, \left( \frac{b_i}{\theta} \right)^2 \right). \tag{4.5}$$

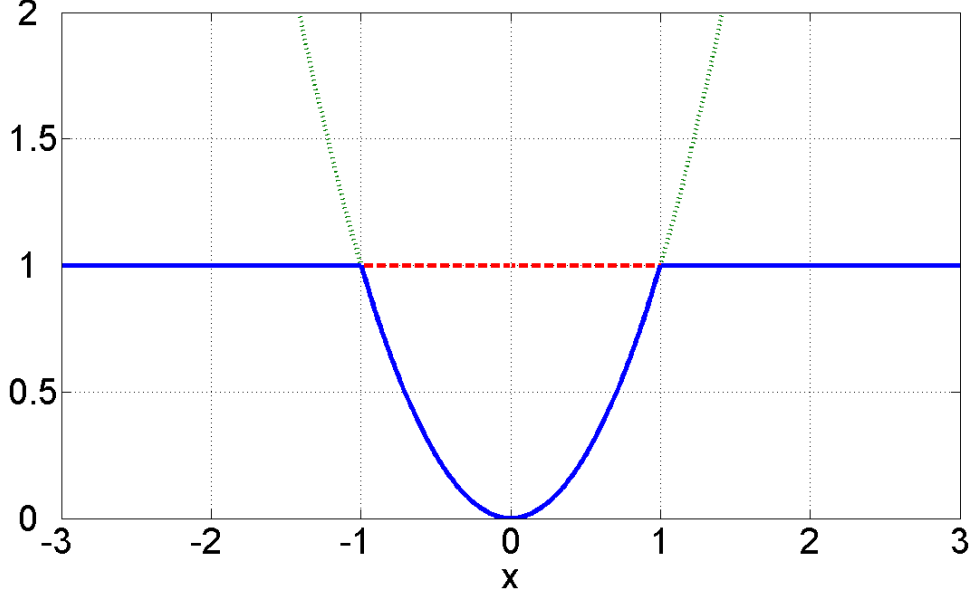


Figure 4.1: Bold line shows the minimum between  $y(x) = 1$  (dashed) and  $y(x) = x^2$  (dotted).

## 4.2 Local minimisation with $\ell_0$ -norm: IHT

The gradient of the new (unconstrained) cost function is  $\nabla C_0(\mathbf{b}, \theta) = \mathbf{c}$ , where:

$$c_i = \begin{cases} 0, & |b_i| > \theta \\ \frac{2}{\theta^2} b_i, & |b_i| \leq \theta. \end{cases}$$

This can be expressed more compactly as:

$$\nabla C_0(\mathbf{b}, \theta) = \frac{2}{\theta^2} (\mathbf{b} - S_0(\mathbf{b}, \theta)).$$

The projection of this gradient onto the affine subspace of perfect reconstruction,  $S(\Phi, \mathbf{x})$ , is:

$$\nabla^{S(\Phi, \mathbf{x})} C_0(\mathbf{b}, \theta) = (\mathbf{I} - \Phi^T \Phi) \nabla C_0(\mathbf{b}, \theta).$$

Every iteration of the gradient descent method is:

$$\mathbf{b}^{(k+1)} = \mathbf{b}^{(k)} - \alpha \nabla^{S(\Phi, \mathbf{x})} C_0(\mathbf{b}^{(k)}, \theta).$$

As this projection is the component of the gradient in the null space of  $\Phi$ ,  $\mathbf{b}^{(k)}$  always provides perfect reconstruction, no matter which value of  $\alpha$  we use. Substituting the gradient expression we obtain:

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



A necessary and, in our case, sufficient condition to reach a local minimum of the cost function is that:

$$\nabla^{S(\Phi, \mathbf{x})} C(\mathbf{b}^*, \theta) = \mathbf{0}.$$

This is the convergence condition of the previous iterations. This means that, if those iterations converge, they do it to a local minimum of the cost function in Equation (4.5).

Note that the choice of  $\alpha = \alpha_0 = \frac{1}{2\lambda} = \frac{\theta^2}{2}$  minimises the unconstrained cost function of Equation (4.5) for a single descent step, resulting in:

$$\mathbf{b}^{(k+1)} = S_0(\mathbf{b}^{(k)}, \theta) + \Phi^T (\mathbf{x} - \Phi S_0(\mathbf{b}^{(k)}, \theta)),$$

that is, the same Iterative Hard Thresholding (IHT) method described in subsection 2.3.3.

We have shown that this procedure provides, when converging, a local minimum in the classical sparse approximation criterion (Equation (4.1)). However, in general, choosing the  $\alpha$  value which minimises in one step the unconstrained cost function ( $\alpha_0$ ) is not optimal in terms of convergence speed. We have hand-optimised the convergence speed by using  $\alpha \sim 1.85\alpha_0$ .

Recently, we have known that [84] also proved, in a parallel and independent way to our work, that the convergence point of the IHT iterations is a local minimum of the cost function<sup>1</sup>. However, they prove, in addition, that the iterations converge indeed, provided that the eigenvalues of  $(\mathbf{I} - \Phi^T \Phi)$  are between 0 and 1, where  $\mathbf{I}$  is the  $M \times M$  identity matrix.

Figure 4.2 shows some convergence curves using fixed thresholds and different  $\alpha$  values, and using 8-scale DT-CWT as representation<sup>2</sup>, whose redundancy factor is 4. We can see that, although the striking simplicity of this method, doing gradient descent for a given  $\lambda$  value until convergence is too expensive in computational terms. In addition, we know that the local minimum obtained is clearly worse than that of  $\ell_0$ -AP (see Section 3.4).

### 4.3 Global minimisation with $\ell_0$ -norm: $\ell_0$ -GM

We propose next an efficient alternative to IHT and  $\ell_0$ -AP, inspired by deterministic global optimisation techniques, which drastically reduces

<sup>1</sup>This work was published in April 2007, while ours [10] appeared in August same year.

<sup>2</sup>Except when indicated, this is the representation used throughout the chapter. We have experienced that other representations, as Curvelets, provide a qualitatively similar behaviour.

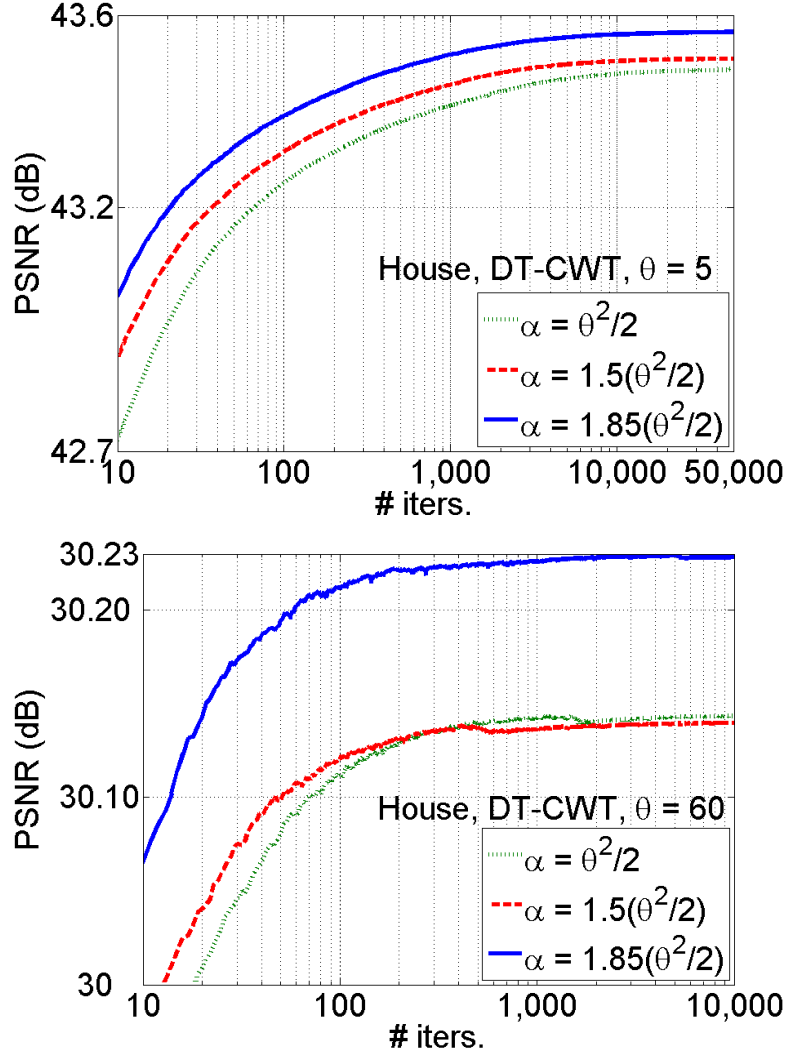


Figure 4.2: **Top**, IHT convergence curves using a low threshold ( $\theta = 5$ ) and three different  $\alpha$  values. We have used House image and 8-scale DT-CWT. **Bottom**, same result for a higher threshold ( $\theta = 60$ ).

computational cost compared to IHT, while increasing the energy compaction capacity.

The cost function in Equation (4.5) can be re-written as:

$$C_0(\mathbf{b}, \theta) = \sum_{i=1}^M (1 - h(b_i/\theta)), \quad (4.6)$$

where  $h(x) = \max(1 - x^2, 0)$  is the inverted parabolic arc centred at zero, going from  $-1$  to  $1$ , and reaching a maximum value of one at zero, with an

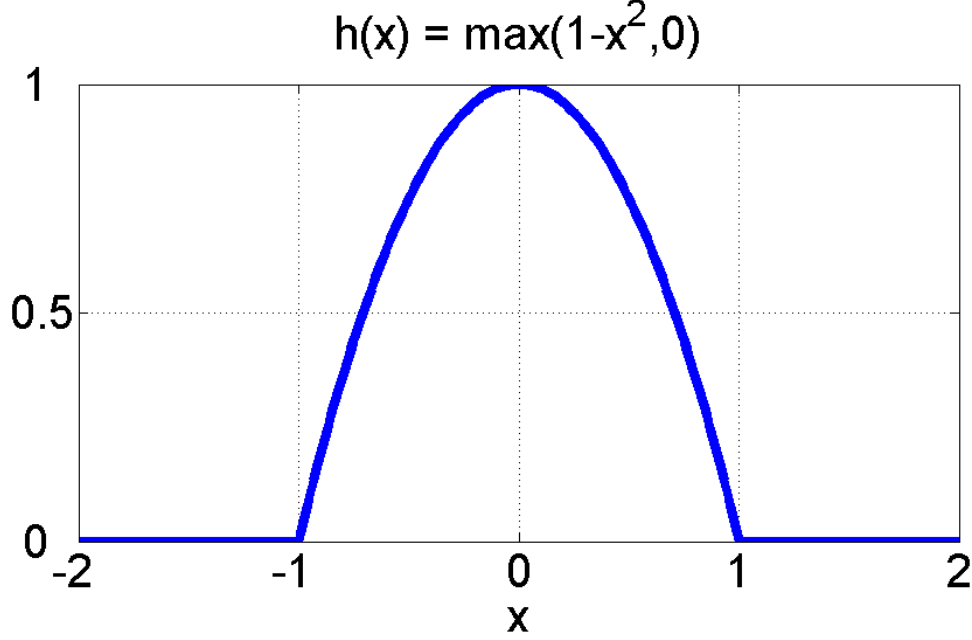


Figure 4.3: 1-D smoothing function: an inverted parabola in the interval  $[-1, 1]$ , centred at 0 and with maximum 1. Outside that interval is 0.

amplitude of one. A plot of this function is shown in Figure 4.3. We can re-write the optimisation problem in Equation (4.4) as follows:

$$\begin{aligned}
 \hat{\mathbf{b}} &= \arg \max_{\mathbf{b}} C'(\mathbf{b}, \theta), \\
 \hat{\mathbf{a}} &= S_0(\hat{\mathbf{b}}, \theta), \\
 C'(\mathbf{b}, \theta) &= \sum_{i=1}^M h(b_i/\theta) = M - C_0(\mathbf{b}, \theta).
 \end{aligned} \tag{4.7}$$

It is easy to express this in terms of an infinitely sharp cost function,  $C_\delta(\mathbf{b})$ , convolved with a smoothing kernel,  $H(\mathbf{b}/\theta)$ , which has a variable width depending on  $\theta$  (the higher is  $\theta$  the stronger the smoothing effect):

$$C'(\mathbf{b}, \theta) \propto C_\delta(\mathbf{b}) * H(\mathbf{b}/\theta), \tag{4.8}$$

where  $H(\mathbf{b}) = \prod_{i=1}^M h(b_i)$  and  $C_\delta(\mathbf{b}) = \sum_{i=1}^M \delta(b_i)$ . In order to prove this proportionality, we start by writing:

$$C_\delta(\mathbf{b}) * H(\mathbf{b}/\theta) = \left( \sum_{i=1}^M \delta(b_i) \right) * \left( \prod_{k=1}^M h(b_k/\theta) \right).$$

We can express the delta function as an outer product of constants expressed as functions of each dimension separately:

$$C_\delta(\mathbf{b}) * H(\mathbf{b}/\theta) = \left( \sum_{i=1}^M \delta(b_i) \left[ \prod_{j \neq i} c_1(b_j) \right] \right) * \left( \prod_{k=1}^M h(b_k/\theta) \right),$$

where  $c_1(b_j) = 1$  is a constant function of  $b_j$ . Now, we can write the convolution of outer products as outer product of convolutions:

$$C_\delta(\mathbf{b}) * H(\mathbf{b}/\theta) = \sum_{i=1}^M \left( \prod_{j \neq i} h(b_j/\theta) * c_1(b_j) \right) (\delta(b_i) * h(b_i/\theta)).$$

The convolution with the unit constant is an integral, and the convolution with the unity delta at zero does not change the function being convolved. So, we write:

$$C_\delta(\mathbf{b}) * H(\mathbf{b}/\theta) = \sum_{i=1}^M h(b_i/\theta) \left( \prod_{j \neq i} A(\theta) \right),$$

where  $A(\theta) = \int_{-\infty}^{\infty} h(x/\theta) dx = \int_{-\theta}^{\theta} h(x/\theta) dx = 4\theta/3$ . Then:

$$C_\delta(\mathbf{b}) * H(\mathbf{b}/\theta) = A(\theta)^{M-1} \sum_{i=1}^M h(b_i/\theta).$$

Finally, by substituting the definition of  $C'(\mathbf{b}, \theta)$ :

$$C_\delta(\mathbf{b}) * H(\mathbf{b}/\theta) = A(\theta)^{M-1} C'(\mathbf{b}, \theta).$$

So we can express:

$$C'(\mathbf{b}, \theta) = A(\theta)^{1-M} C_\delta(\mathbf{b}) * H(\mathbf{b}/\theta),$$

and the relation of proportionality in Equation (4.8) is proved. The scale factor,  $A(\theta)^{-M+1}$ , is irrelevant in terms of the minimisation of vector  $\mathbf{b}$  in Equation (4.7).

As Figure 4.2 illustrates, it is faster to find a local optimum when  $\theta$  is high, or, equivalently,  $\lambda$  is low, which corresponds to a smooth cost function. Moreover, having a good candidate for the global optimum for a given  $\lambda$ , we can expect a good result by searching from it the nearest optimum corresponding to a similar, slightly higher,  $\lambda$ . From here we derive the following method. Starting from a small  $\lambda$ , we do gradient descent until

reaching convergence, then set a slightly higher  $\lambda$ , do again gradient descent from the previous convergence point, and so on until reaching the desired  $\lambda$  value. We call this method  $\ell_0$ -GM (from Gradual Minimisation). A faster and simpler approximated version is to increase slowly  $\lambda$  at each iteration, so drastically reduce the number of iterations. In fact, both versions become equivalent in the limit when the increase of  $\lambda$  at each iteration becomes infinitesimal. In terms of the threshold  $\theta$ , we start from the highest possible threshold (highest amplitude in  $\mathbf{a}^{LS}$ ) and slowly decrease it at each *de-smoothing* iteration, until reaching the desired value. In Figure 4.4 we illustrate the concept guiding this method with an example of a function of multiple minima smoothed until getting a function with one single minimum. The path joining all the minima throughout the different scales is drawn. In this example there is continuity of the global minima as a function of the scale, which is a necessary condition for  $\ell_0$ -GM to reach the global optimum. This, in general, does not happen in real cases. We have seen that this method finds the global optimum for extremely high sparseness levels (around few tens of DT-CWT coefficients). Nevertheless, global optimality conditions are beyond the scope of this Thesis, where we are more interested in the methods behaviour under practical conditions.

The idea of smoothing a cost function to avoid getting trapped by unfavourable local minima is closely related to other deterministic annealing schemes, such as [118, 119]. Some authors had already proposed this idea as an heuristic to obtain algorithms promoting the energy compaction, by using either soft-thresholding [19] or hard-thresholding [13, 89]. But the referred authors did not propose their algorithms as means to solve well-founded optimisation problems.

Figure 4.5 shows, on the one hand, some convergence trajectories (dashed lines) of IHT for different fixed thresholds (which corresponds to search for a local optimum by doing gradient descent with a fixed  $\lambda$ ). On the other hand, it shows two trajectories (circles and solid lines) corresponding to exponentially decrease the threshold with the rule  $\theta^{(k)} = \theta^{(0)}\beta^k$  for two different  $\beta$  values. The closer to 1 is  $\beta$ , the better is the compaction, but also the slower is the convergence. We, as other authors [15, 19], have experienced that, in practice, exponential decreasing of the threshold provides a better compromise between computational cost and quality of the result than other decreasing functions, as linear. By decreasing dynamically the threshold we are not only dramatically reducing the required number of iterations for reaching convergence, but we are also significantly improving the achievable fidelity for any given sparseness level.

Top panel of Figure 4.6 shows a family of fidelity-sparseness curves for different  $\beta$  values. An ideal curve would have an asymptote to perfect

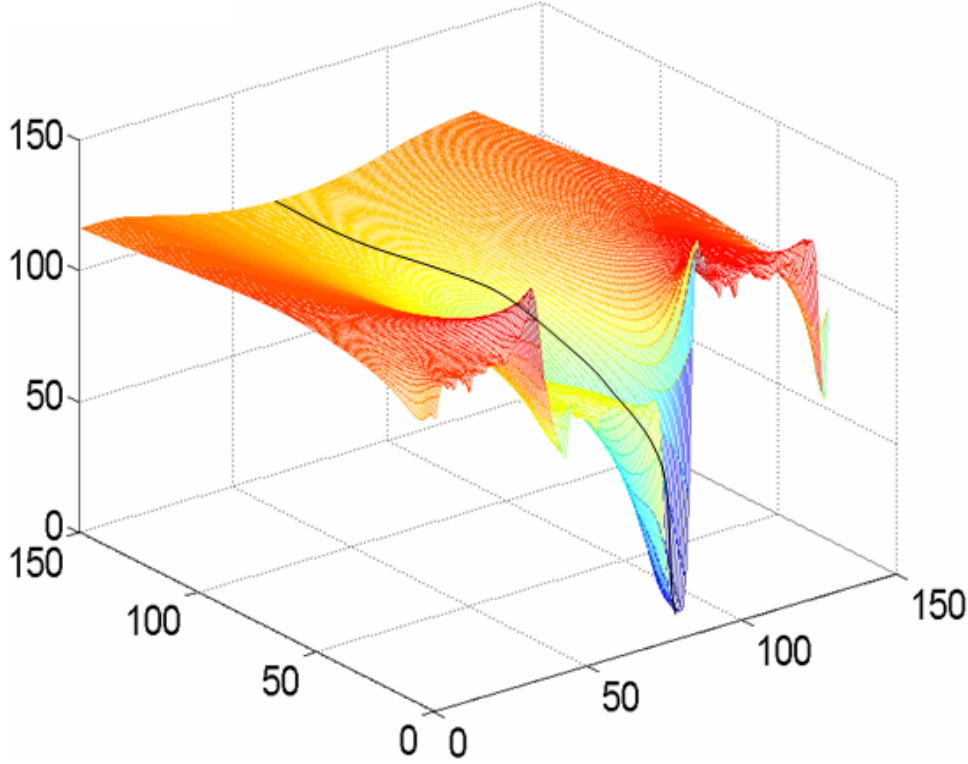


Figure 4.4: 1-D Function with multiple minima progressively smoothed until obtaining only one. The black continuous line indicates the path joining the global optima through the scale of the smoothing kernel. We have used here as smoothing kernel a normalised (in area) version of  $h(x)$  (See Figure 4.3).

reconstruction in  $N$ . Note that  $\ell_0$ -GM approximates this asymptote as  $\beta$  gets closer to 1. This is even more significant if we consider that achieving a global optimum for low sparseness levels is much more difficult than for high sparseness, because the number of local optima increases very rapidly with  $\lambda$ .

#### 4.3.1 Using a single solution for all the sparseness levels

If we optimise using  $\ell_0$ -GM for a set of  $\lambda$  values, we end up having multiple solutions, one for every value taken by the threshold in its descending path. Which criterion should we apply to choose a particular solution? Is it possible to find a  $\theta_0$  value whose associated minimum of the cost function,  $C_0(\hat{\mathbf{b}}(\theta_0), \theta_0)$ , can be extended as  $C_0(\hat{\mathbf{b}}(\theta_0), \theta)$  to approximate the minimum of the cost function for other thresholds  $\theta$  (that is, such that

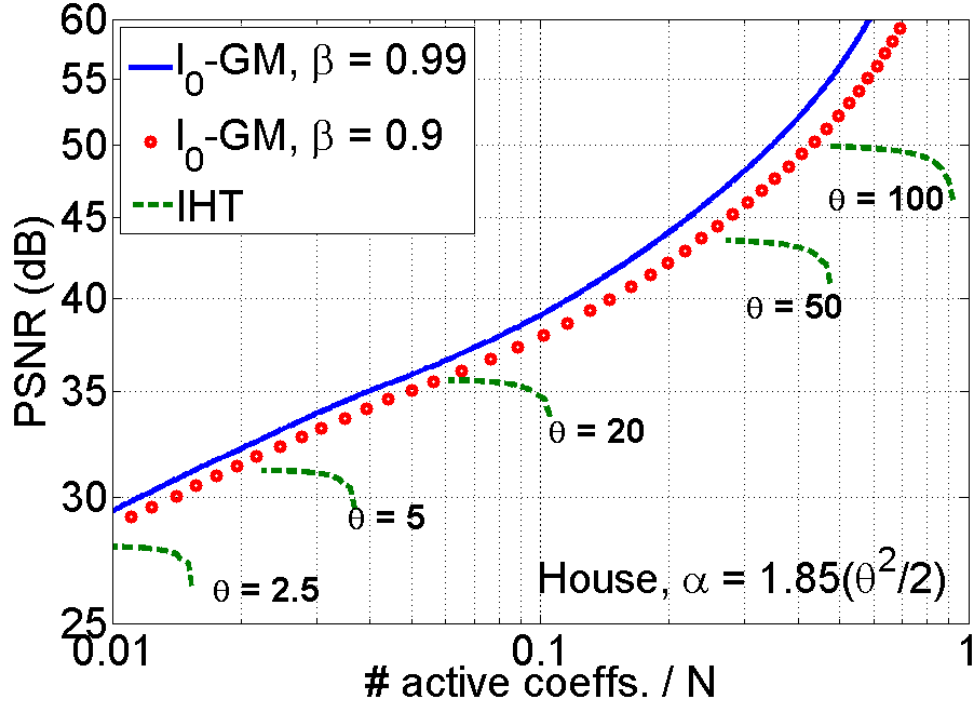


Figure 4.5: Fidelity-sparseness results of  $\ell_0$ -GM, using  $\beta = 0.9$  (circles,  $1.5 \cdot 10^2$  iterations) and  $\beta = 0.99$  (solid,  $1.5 \cdot 10^3$  iterations), compared to IHT, using several thresholds (dashed,  $10^5$  iterations). We use House image and DT-CWT with 8-scales.

$C_0(\hat{\mathbf{b}}(\theta_0), \theta) \approx C_0(\hat{\mathbf{b}}(\theta), \theta)$ , for all  $\theta > \theta_0$ ? The answer, surprisingly enough, is "yes". This problem has an important practical impact, because in that case we could use  $\hat{\mathbf{a}}'(\theta) = S_0(\hat{\mathbf{b}}(\theta_0), \theta)$  as an almost equally good substitute of  $\hat{\mathbf{a}} = S_0(\hat{\mathbf{b}}(\theta), \theta)$ , that does not require to use, and store<sup>3</sup>,  $\hat{\mathbf{b}}(\theta)$ .

We have explained before how, to get a good solution for some  $\lambda_i > \lambda_j$ , it is good to start from the solution associated to  $\lambda_j$  and, from there, refine it until reaching the solution associated to  $\lambda_i$ . This seems to imply that good solutions for high  $\lambda$  values should be reasonable good for lower values. In bottom panel of Figure 4.6 we show the fidelity-sparseness curve obtained by *a posteriori* thresholding the solution obtained using a very high  $\lambda$  value. As we can see, for  $\beta < 0.99$ , the results are even better than the curve obtained throughout the execution of the direct  $\ell_0$ -GM method. This means that only an optimisation solution, for a determined sparseness level (the lowest one) is enough to have a good approximate solution to all  $\lambda$  values considered. This is a practical advantage, because it means that we do

<sup>3</sup>However, this does not prevent us from computing  $\hat{\mathbf{b}}(\theta)$ , if  $\theta > \theta_0$ , because the computation of  $\hat{\mathbf{b}}(\theta_0)$  in  $\ell_0$ -GM requires the computation of  $\hat{\mathbf{b}}(\theta)$ , for all  $\theta > \theta_0$  (in practice, a dense enough sampling of the interval  $[\theta_{max}, \theta_0]$ )

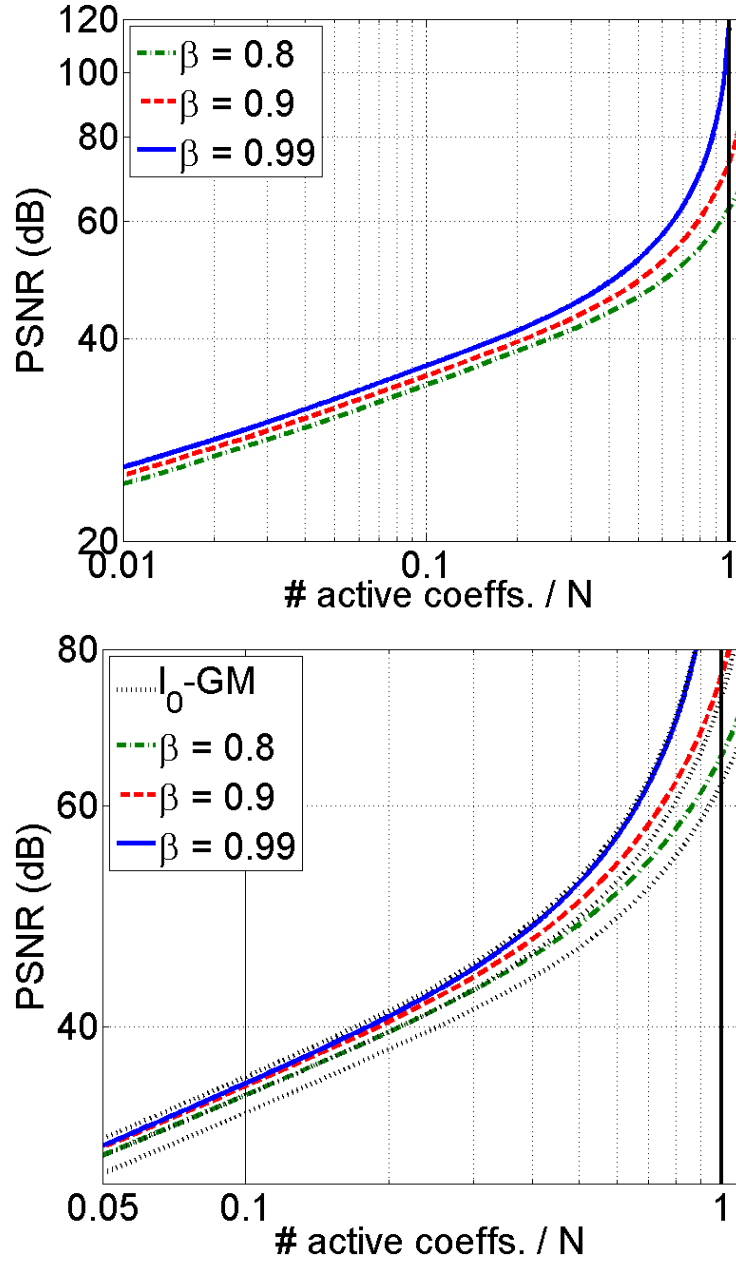


Figure 4.6: **Top**, sparse approximation fidelity averaged in our test set using  $\ell_0$ -GM with  $\alpha = 1.85(\theta^2/2)$ , three different  $\beta$  values and 8-scale DT-CWT. **Bottom**, quality of the reconstruction from the highest amplitude coefficients of the vector obtained using  $\ell_0$ -GM for a very high  $\lambda$  value (very low sparseness), and the same  $\beta$  values. Dotted curves correspond to that of top panel. The vertical axis has been re-scaled to improve visibility.



not need to store all  $\hat{\mathbf{b}}(\theta)$  to choose a threshold level  $\theta$ , corresponding to a certain  $\lambda$ , in real time. This allows, for example, for adapting to a variable channel bandwidth in communications, and it provides, in general, a flexible approach to quickly shift the trade-off between fidelity and sparseness.

## 4.4 Implementation

We have experimented with several Parseval frames, as we did with  $\ell_p$ -AP method. Although the qualitative conclusions of the experiments are similar using any of them, we have chosen DT-CWT to show the experiments in this chapter. Together with Curvelets, it offers the best compaction results among the compared representations. In addition to this, the MATLAB® implementation available [116] is much faster than the one for Curvelets [117].

Similarly to the previous chapter, the complex coefficients of DT-CWT have been separated in real and imaginary parts, in order to make a homogeneous treatment of them. Moreover, it has also been added an extra scaled composed by one only coefficient storing the global mean of the image.

We have checked that, in our  $\ell_0$ -GM implementation, the best results are obtained when the decrease interval of  $\theta$  is as large as possible. Then, the threshold is initialised to the second largest amplitude of the linear response to the image (to choose at least one coefficient in the first iteration). Then it is decreased until reaching the desired value. This value depends on the application. Following the previous section, we choose a small final value for  $\theta$ , then obtaining a solution with good performance at every sparseness level.

## 4.5 Results and discussion for $\ell_0$ -GM

Figure 4.7 compares the compaction performance of the following methods:  $\ell_0$ -GM with  $\beta = 0.99$ ,  $\ell_0$ -AP, post LS-optimised  $\ell_1$ -AP ( $\ell_1$ -AP+OP), IHT and StOMP. See more details about these methods in Chapter 3. The improvement on the behaviour of  $\ell_0$ -GM with respect to the rest is very remarkable. These plot show clearly that we can obtain, in the conditions of this experiment, a much better local minimum to the sparse approximation problem by directly minimising the  $\ell_0$ -norm than solving the convex relaxation problem, even LS-optimising the coefficients for the selected support. One important difference of  $\ell_0$ -GM with the methods based on alternating projections is that we can sweep all the sparseness

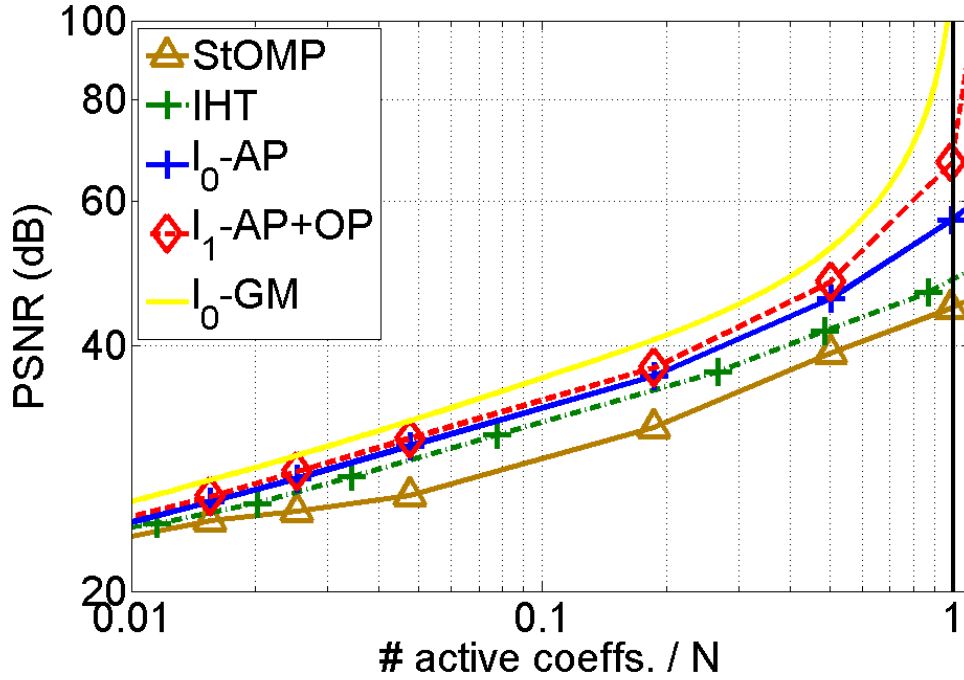


Figure 4.7:  $\ell_0$ -GM sparse approximation results averaged in our test set compared to other methods previously seen (StOMP, IHT,  $\ell_0$ -AP and  $\ell_1$ -AP+OP).

levels in the same execution of the method, instead of making a lot of iterations for each level, each time. Tables 4.1 and 4.2 show the numerical data of Figure 4.7.

Other strategies exist in the literature for the dynamic thresholding, and depending on the precise case, they can give slightly better results than those of  $\ell_0$ -GM. For example, in [15], the number of preserved coefficients at each iteration is linearly increased.

Figure 4.8 shows a visual comparison of sparse approximation with the different methods using  $0.04 \cdot N$  DT-CWT coefficients and *Einstein* image. We see that  $\ell_0$ -GM preserves significantly better the perceptually relevant information of the original.

Image	Method	# active coeffs./ $N$		
		0.00868	0.02536	0.04761
Barbara	StOMP	24.48	25.55	26.45
	IHT	23.16	27.25	29.89
	$\ell_0$ -AP	24.89	28.67	31.89
	$\ell_1$ -AP+OP	<i>25.24</i>	<i>29.38</i>	<i>32.60</i>
	$\ell_0$ -GM	<b>26.18</b>	<b>30.47</b>	<b>33.93</b>
House	StOMP	26.40	28.27	29.82
	IHT	24.08	30.35	32.91
	$\ell_0$ -AP	27.33	31.50	34.16
	$\ell_1$ -AP+OP	<i>27.75</i>	<i>32.01</i>	<i>34.73</i>
	$\ell_0$ -GM	<b>28.85</b>	<b>33.18</b>	<b>35.65</b>
Boat	StOMP	20.75	23.89	24.89
	IHT	16.82	24.74	27.09
	$\ell_0$ -AP	21.57	25.76	27.86
	$\ell_1$ -AP+OP	<i>21.96</i>	<i>26.12</i>	<i>28.37</i>
	$\ell_0$ -GM	<b>23.77</b>	<b>27.32</b>	<b>30.13</b>
Lena	StOMP	23.21	24.67	26.02
	IHT	17.27	26.22	28.93
	$\ell_0$ -AP	23.62	27.20	30.09
	$\ell_1$ -AP+OP	<i>24.17</i>	<i>27.96</i>	<i>30.95</i>
	$\ell_0$ -GM	<b>25.19</b>	<b>29.29</b>	<b>32.34</b>
Peppers	StOMP	21.43	24.19	25.28
	IHT	16.76	25.15	27.93
	$\ell_0$ -AP	22.61	26.39	29.17
	$\ell_1$ -AP+OP	<i>22.81</i>	<i>26.85</i>	<i>29.74</i>
	$\ell_0$ -GM	<b>24.03</b>	<b>28.43</b>	<b>31.57</b>

Table 4.1: Fidelity (PSNR, in dB) for several sparseness levels, using the images in our test set and five different methods, and using 8-scale DT-CWT. Bold numbers indicate the method providing the best approximation for each image and sparseness level, and italic indicate the second best. Columns correspond to 569, 1662 and 3120 active coefficients. We have directly extracted the PSNR values from the experiments, except for IHT, where they have been linearly interpolated.

## 4.6 Gradient descent for minimisation of $\ell_1$ -norm: IST & $\ell_1$ -GM

### 4.6.1 Alternative formulation of the convex cost function

We described the convex relaxation problem in Equation (2.5), which we repeat here for convenience:

$$\hat{\mathbf{a}}^1(\lambda) = \arg \min_{\mathbf{a}} \{\|\mathbf{a}\|_1 + \lambda \|\Phi \mathbf{a} - \mathbf{x}\|_2^2\}, \quad (4.9)$$

		# active coeffs./N		
Image	Method	0.1865	0.5021	0.9869
Barbara	StOMP	34.58	42.39	46.63
	IHT	35.68	44.12	48.56
	$\ell_0$ -AP	40.39	47.63	55.64
	$\ell_1$ -AP+OP	<i>41.57</i>	<i>51.22</i>	> <b>100</b>
	$\ell_0$ -GM	<b>43.05</b>	<b>54.93</b>	> <b>100</b>
House	StOMP	33.86	41.83	47.89
	IHT	38.14	44.61	50.75
	$\ell_0$ -AP	38.40	49.25	58.98
	$\ell_1$ -AP+OP	<i>39.20</i>	<i>51.25</i>	<i>62.31</i>
	$\ell_0$ -GM	<b>43.47</b>	<b>56.12</b>	> <b>100</b>
Boat	StOMP	29.06	36.42	42.05
	IHT	32.07	39.30	45.65
	$\ell_0$ -AP	33.52	43.29	58.91
	$\ell_1$ -AP+OP	<i>34.25</i>	<i>45.02</i>	> <b>100</b>
	$\ell_0$ -GM	<b>38.09</b>	<b>50.10</b>	> <b>100</b>
Lena	StOMP	31.70	39.48	45.06
	IHT	34.33	42.30	48.21
	$\ell_0$ -AP	36.85	45.66	55.51
	$\ell_1$ -AP+OP	<i>37.79</i>	<i>47.95</i>	<i>63.72</i>
	$\ell_0$ -GM	<b>40.97</b>	<b>52.97</b>	> <b>100</b>
Peppers	StOMP	31.60	38.50	43.17
	IHT	33.98	41.19	48.66
	$\ell_0$ -AP	37.20	44.84	57.09
	$\ell_1$ -AP+OP	<i>38.41</i>	<i>47.13</i>	> <b>100</b>
	$\ell_0$ -GM	<b>39.93</b>	<b>51.63</b>	> <b>100</b>

Table 4.2: Continuation of Table 4.1. Columns correspond, respectively, to 12221, 32905 and 64682 active coefficients.

The associated cost function, in contrast to the  $\ell_0$ -norm case, is convex and, thus, continuous. Nevertheless, we are interested in doing a similar transformation as for that case. The proof of that solution  $\hat{\mathbf{a}}$  of the problem:

$$(\hat{\mathbf{a}}, \hat{\mathbf{b}}) = \arg \min_{\mathbf{a}, \mathbf{b}} \{\|\mathbf{a}\|_1 + \lambda \|\mathbf{b} - \mathbf{a}\|_2^2 \text{ s.t. } \Phi \mathbf{b} = \mathbf{x}\}, \quad (4.10)$$

is equivalent to  $\hat{\mathbf{a}}^1(\lambda)$  is analogous to the one  $\ell_0$ -norm case. We can express  $\hat{\mathbf{b}}$  as:

$$\hat{\mathbf{b}} = \arg \min_{\mathbf{b}} \{\min_{\mathbf{a}} \{\|\mathbf{a}\|_1 + \lambda \|\mathbf{b} - \mathbf{a}\|_2^2\} \text{ s.t. } \Phi \mathbf{b} = \mathbf{x}\}. \quad (4.11)$$

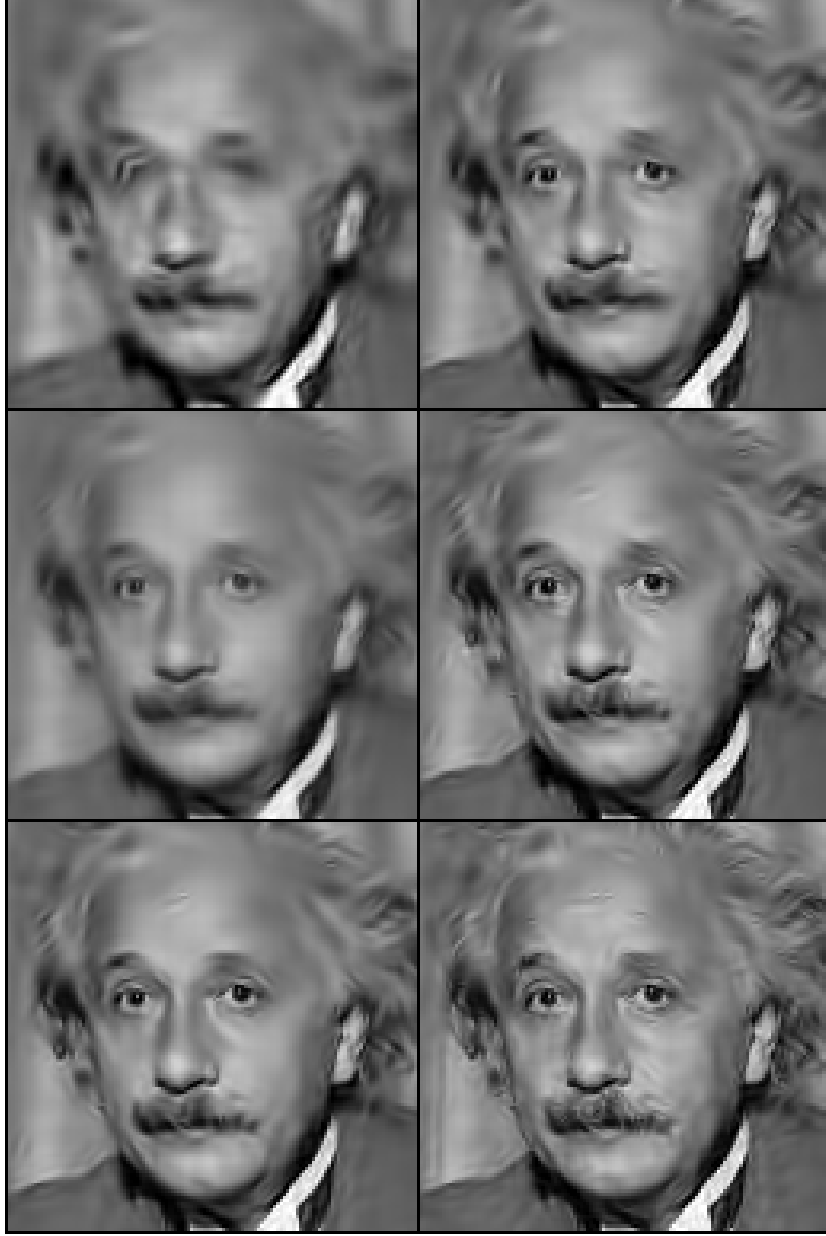


Figure 4.8:  $64 \times 64$  crop of the reconstruction of Einstein image using  $0.04 \cdot N$  (2605) active DT-CWT coefficients, for several sparse approximation methods. **Top-left**, result of StOMP, implemented as described in Section 3.4.1 (28.98 dB). **Top-right**, IHT (31.20 dB). **Centre-left**,  $\ell_1$ -AP (29.70 dB). **Centre-right**,  $\ell_0$ -AP (31.97 dB). **Bottom-left**,  $\ell_1$ -AP+OP (32.38 dB). **Bottom-right**,  $\ell_0$ -GM (33.28 dB).

First, we find the generic expression minimising the inner cost function given **b**. This cost can be decomposed as a summation of a coefficient for each

element of the involved vectors, so the vector  $\tilde{\mathbf{a}}^s(\mathbf{b})$  minimising the inner criterion in Equation (4.11) is:

$$\tilde{\mathbf{a}}^s(\mathbf{b}) = \min_{\mathbf{a}} \left\{ \sum_{i=1}^M c(a_i, b_i) \right\},$$

where  $c(a, b) = |a| + \lambda(b - a)^2$ . The derivative in  $a$  of this function is  $\frac{\partial c(a, b)}{\partial a} = d + 2\lambda(a - b)$ , where:

$$d = \begin{cases} 1, & a > 0 \\ -1, & a < 0 \\ 0, & a = 0. \end{cases} \quad (4.12)$$

For the case  $a > 0$ , we have that:

$$\frac{\partial c(a, b)}{\partial a} = 1 + 2\lambda(a - b).$$

Equaling to zero we obtain:

$$a = b - \frac{1}{2\lambda},$$

from where it follows that  $b > \frac{1}{2\lambda}$ , given that  $\lambda > 0$  by definition. For the case  $a < 0$ , analogously, we obtain:

$$a = b + \frac{1}{2\lambda},$$

and then  $b < -\frac{1}{2\lambda}$ . Joining these two cases we have that:

$$a = \text{sign}(b) \cdot (|b| - \frac{1}{2\lambda}),$$

when  $|b| > \frac{1}{2\lambda}$ .

On the other hand, when  $|b| \leq \frac{1}{2\lambda}$ , the value of  $a$  minimising the associated cost function changes the sign with respect to  $b$ . Given that for every quadrant we only consider values in the same quadrant, this implies that the minimum is at zero.

Thus, by applying these results in our problem, we obtain that the vector  $\tilde{\mathbf{a}}^s(\mathbf{b})$  is the result of a soft-thresholding operation of  $\mathbf{b}$  with threshold  $\theta = \frac{1}{2\lambda}$ . We denote this operation as  $\tilde{\mathbf{a}}^s(\mathbf{b}) = S_1(\mathbf{b}, \theta)$ . Substituting in Equation (4.10):

$$\hat{\mathbf{b}} = \arg \min_{\mathbf{b}} \{ \|S_1(\mathbf{b}, \theta)\|_1 + \lambda \|\mathbf{b} - S_1(\mathbf{b}, \theta)\|_2^2 \text{ s.t. } \Phi \mathbf{b} = \mathbf{x} \}.$$

And finally, given that the cost function of the previous expression is separable as a sum of independent terms for each coefficient index, we can write:

$$\begin{aligned}\hat{\mathbf{b}} &= \arg \min_{\mathbf{b}} \{C_1(\mathbf{b}, \theta) \text{ s.t. } \Phi \mathbf{b} = \mathbf{x}\}, \\ \hat{\mathbf{a}} &= S_1(\hat{\mathbf{b}}, \theta),\end{aligned}$$

where:

$$\begin{aligned}C_1(\mathbf{b}, \theta) &= \sum_{i=1}^M c_i, \\ c_i &= \begin{cases} |b_i| - \frac{\theta}{2}, & |b_i| > \theta \\ \frac{b_i^2}{2\theta}, & |b_i| \leq \theta. \end{cases}\end{aligned}\tag{4.13}$$

#### 4.6.2 Cost function minimisation with a fixed threshold: IST

The derivation of the gradient descent-based method with the cost function  $C_1(\mathbf{b}, \theta)$  is analogous to the one shown for function  $C_0(\mathbf{b}, \theta)$ . From Equation 4.13 we obtain:

$$\nabla C_1(\mathbf{b}, \theta) = \frac{1}{\theta}(\mathbf{b} - S_1(\mathbf{b}, \theta)),$$

and, after projecting onto the affine space of perfect reconstruction,  $\nabla^{S(\Phi, \mathbf{x})} C_1(\mathbf{b}, \theta) = (\mathbf{I} - \Phi^T \Phi) \nabla C_1(\mathbf{b}, \theta)$ , we end up with the following iterations:

$$\mathbf{b}^{(k+1)} = \mathbf{b}^{(k)} - \frac{\alpha}{\theta} (\mathbf{I} - \Phi^T \Phi) (\mathbf{b}^{(k)} - S_1(\mathbf{b}^{(k)}, \theta)).$$

A necessary (and also sufficient, in this case) condition to reach the global minimum of the function  $C_1(\mathbf{b}, \theta)$  is that  $\nabla^{S(\Phi, \mathbf{x})} C_1(\mathbf{b}^*, \theta) = \mathbf{0}$ . Note that this is the convergence condition for the previous iterations. Choosing  $\alpha = \alpha_0 = \frac{1}{2\lambda} = \theta$  leads us to IST method:

$$\mathbf{b}^{(k+1)} = S_1(\mathbf{b}^{(k)}, \theta) + \Phi^T (\mathbf{x} - \Phi S_1(\mathbf{b}^{(k)}, \theta)).$$

It has been proved [28, 37] that this procedure provides the global minimum of the convex relaxation problem shown in Equation (4.9).

However, as in the case  $p = 0$ , in general the  $\alpha = \theta$  choice, though it minimises in  $\alpha$  in one step the unconstrained cost function, is not optimal in terms of convergence speed. We have also experienced that we can obtain

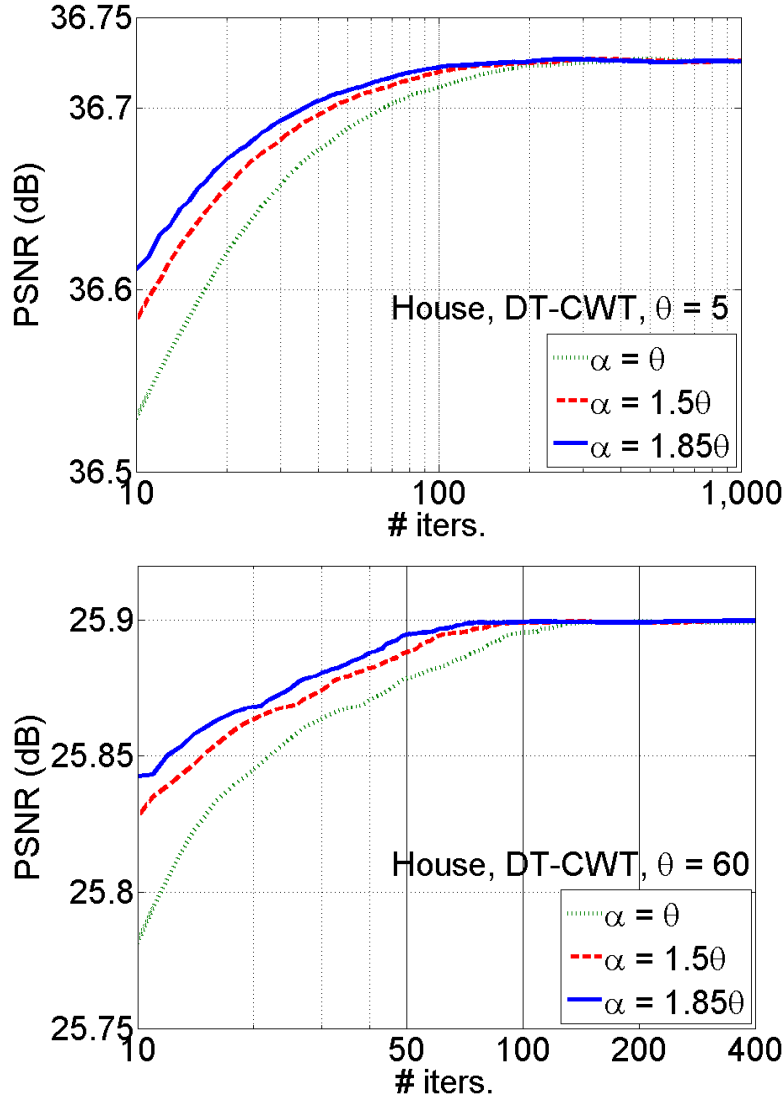


Figure 4.9: **Top**, convergence curves for IST with a low threshold ( $\theta = 5$ ) and three different  $\alpha$  values. We have used House image and 8-scale DT-CWT. **Bottom**, same result for a higher threshold ( $\theta = 60$ ).

a faster convergence using  $\alpha \sim 1.85\alpha_0$ , though now the difference of using  $\alpha = \alpha_0$  is very small. Figure 4.9 shows the corresponding convergence curves in the same conditions stated for IHT in Figure 4.2. The convergence here is much much faster and uniform than for IHT, because of the convexity of the cost function.



### 4.6.3 A more efficient convex minimisation: $\ell_1$ -GM

We can derive an equivalent, but often more efficient, alternative method to IST and  $\ell_1$ -AP. For any  $\theta$  value, the cost function  $C_1(\mathbf{b}, \theta)$  is convex and therefore one can find its global minimum using IST or  $\ell_1$ -AP. As the value of  $\theta$  is increased ( $\lambda$  decreased), the quadratic term of the function dominates, which provokes a faster convergence (as seen in Figure (4.9)). Moreover, we know that, in this case, it does exist continuity along the global minima of the function for different  $\lambda$  values, and this condition holds because there is only one minimum for each  $\lambda$  and because that minimum must be a continuous function of  $\lambda$ . This property ensures that, starting from the global optimum of a given  $\lambda$  we rapidly converge to the optimum for a slightly higher  $\lambda$ . From here we derive a method similar to  $\ell_0$ -GM but minimising the  $\ell_1$ -norm. That is, we fix a small  $\lambda$ , do gradient descent with IST until reaching convergence, then fix a slightly higher  $\lambda$ , we apply again IST from the previous convergence point, and so on until reaching to the desired value of  $\lambda$ . We call this method  $\ell_1$ -GM. Similarly to the  $\ell_0$ -GM case, a faster and simpler approximation consists of increasing very slowly  $\lambda$  at each iteration, so the number of iterations is significantly reduced.

Figure 4.10 compares the convergence trajectories of IST, using for different thresholds, with respect to two  $\ell_1$ -GM trajectories corresponding to exponentially decreasing the threshold using the rule  $\theta^{(k)} = \theta^{(0)}\beta^k$  for two different  $\beta$  values. We can see that, in practice, the result of the exponential decrease with  $\beta = 0.99$  needs less iterations ( $1.5 \cdot 10^3$  vs.  $10^3$  for IST) to provide a quasi-optimal result for many values of  $\lambda$ , executing in total as many iterations as IST takes for only one sparseness value. Furthermore, using  $\beta = 0.9$ , we achieve a good approximation to the optimal result with a much more reduced number of iterations. Also in this case, as other authors (e.g., [40]), we have experienced that the exponential decay of the threshold provides a better compromise between computational cost and quality of the result than other decreasing functions, as linear, for example.

Figure 4.11 shows a family of fidelity-sparseness for different  $\beta$  values. We have also indicated the results of  $\ell_1$ -AP as a reference. We can appreciate that the fidelity obtained with  $\ell_1$ -GM approximates better the result of  $\ell_1$ -AP as  $\beta$  gets closer to 1.

### 4.6.4 Practical advantages of $\ell_1$ -GM

We have observed that  $\ell_1$ -AP is faster if the solution has a medium-high sparseness level (equivalently, a medium-low  $\lambda$  value), whereas  $\ell_1$ -GM is better in case of having high  $\lambda$  values. Figure 4.12 shows a comparison, using *Barbara* image and 8-scale DT-CWT, of the iterations needed to reach

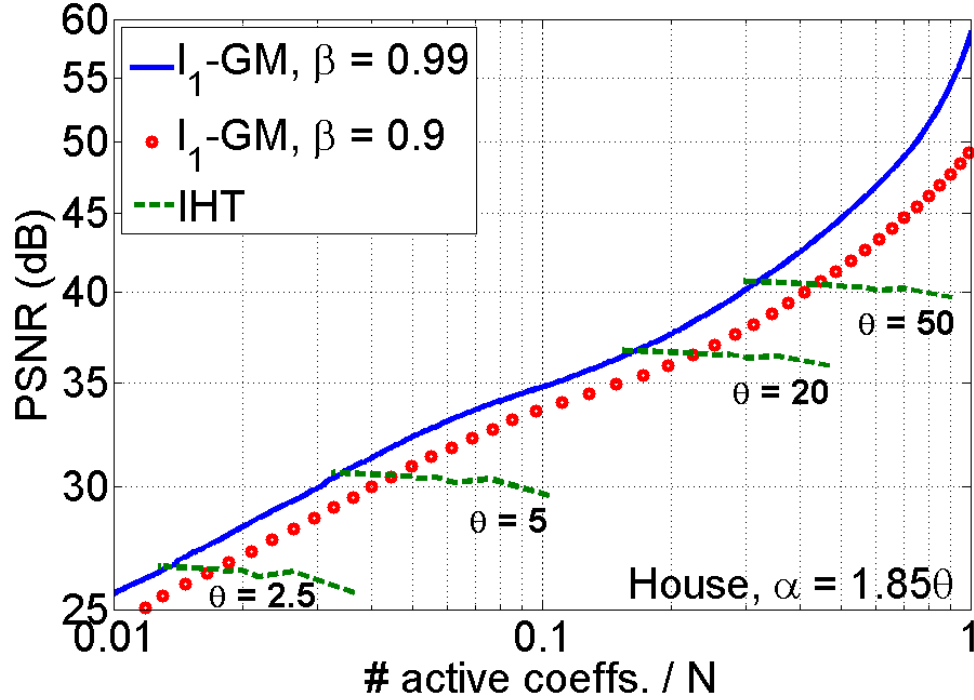


Figure 4.10: Fidelity-sparseness results of  $\ell_0$ -GM, using  $\beta = 0.9$  (circles,  $1.5 \cdot 10^2$  iterations) and  $\beta = 0.99$  (solid,  $1.5 \cdot 10^3$  iterations), compared to IHT, using several thresholds (dashed,  $10^3$  iterations). We use House image and DT-CWT with 8-scales.

a close to optimal result for different sparseness level by methods  $\ell_1$ -AP (as described in Chapter 3) and  $\ell_1$ -GM (with  $\alpha = \theta$  and  $\beta = 0.99$ ). Note that, for low sparseness level,  $\ell_1$ -GM is faster. This case appears often in practice, when trying to look for exact sparse representations, or when solving image restoration problems, if some localised information has been lost (see Chapter 6). That is, when the goal is to match, total or partially, an observation.

## 4.7 Conclusions

In this chapter we have derived an optimisation method, based on iterative shrinkage and with dynamic adjusting of the threshold, to solve the sparse approximation problem. In contrast to existing heuristics (for example [15, 19, 17]) our approximation is fully justified in theory and it is formulated as a classical optimisation problem solution.

Our first step has been to reformulate the sparse approximation problem to obtain an equivalent optimisation problem, but using a constrained continuous function, instead of the discontinuous and unconstrained original

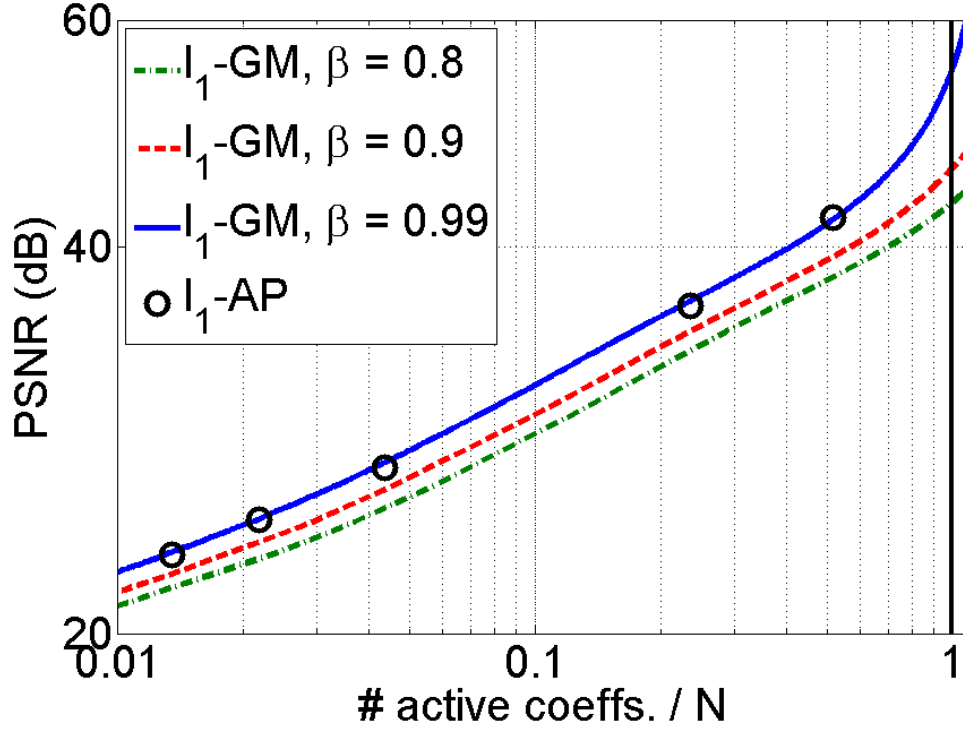


Figure 4.11: Averaged sparse approximation results in the test set using  $\ell_1$ -GM with  $\alpha = 1.85\theta$ , different  $\beta$  values and using DT-CWT with 8 scales. We also show the result of  $\ell_1$ -AP.

cost function, which prevented us to apply classical optimization tools. Then, we have derived a solution for the problem by applying gradient descent on this function, and projecting each iteration onto the set of vectors holding the constraint,  $S(\Phi, \mathbf{x})$ . The resulting method is a generalisation of IHT for the case  $p = 0$ . Finally, we have proposed the  $\ell_0$ -GM method, which is based on a dynamic update of the threshold, while doing gradient descent of the cost function. This method has been justified as a type of deterministic annealing equivalent based on expressing the cost function as the result of convolving tan infinitely sharp reference cost function with a decreasingly smooth kernel.

Our experiments show that  $\ell_0$ -GM is not only more efficient and requires less iterations than other methods ( $\ell_0$ -AP,  $\ell_1$ -AP+OP, IHT) but that it also provides much better compaction results. In fact, its performance, when the number of selected coefficients is close to the number of pixels in the image is close to optimal (asymptotical). This method is comparable to the state-of-the-art in sparse approximation performance. These results show that, under the practical conditions presented in this Thesis, trying to solve

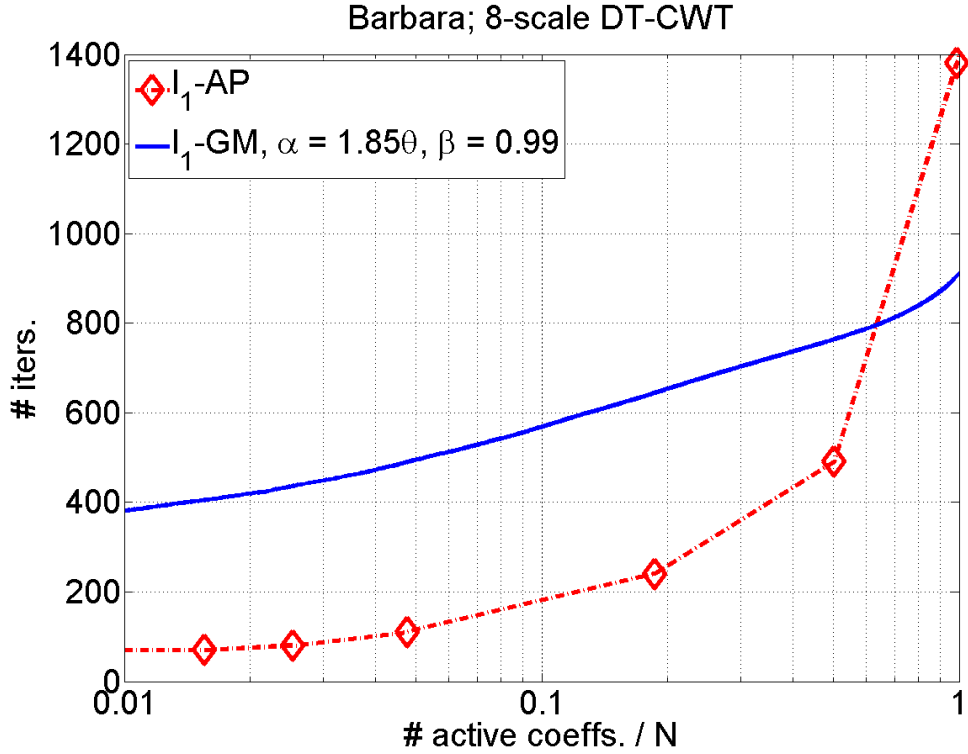


Figure 4.12: Iterations needed to provide nearly optimal result for different sparseness level using  $\ell_1$ -AP and  $\ell_1$ -GM ( $\alpha = 1.85\theta$ ,  $\beta = 0.99$ ). We use Barbara image and 8-scale DT-CWT.

directly for the sparse approximation problem leads to better local minima than solving for the convex relaxation problem.

Analogously, we have derived generalised IST and  $\ell_1$ -GM from the gradient descent in a constrained equivalent version of the cost function associated to the convex relaxation problem. Both methods provide the optimal solution to the problem ( $\ell_1$ -GM when  $\beta \rightarrow 1$ ). We have seen that using  $\ell_1$ -GM is recommended for those applications where we are constrained to preserve some part (of all) of the observation.

We could use the same ideas presented here with other norms. Although the mathematics behind using intermediate quasi-norms ( $0 < p < 1$ ) can be more complicated, but it may be beneficial in order to improve the compaction result.

## Chapter 5

# Application to image restoration

We now consider that we have an incomplete observation. For example, it has lost some pixels, chromatic components, bits, resolution, etc. Our goal is estimating that missing information. We approach the problem by maximising the fidelity to the observation regularised by an *a priori* model based on statistical properties of natural images.

Our fidelity model is based on the concept of consistency. We say that an image is consistent with a degraded observation when, applying the same degradation to the image; we obtain again the given observation. Consequently, to apply this in practice, the degradation should be perfectly reproducible from the observed image. In some cases, it is not possible to identify the precise degradation suffered by some given observation (e.g., white Gaussian noise). But, in some others, it is possible (e.g., missing pixels, bits, chromatic components, resolution, etc.). We call the latter *a posteriori* deterministic degradations.

Our *a priori* model is based on favouring the sparseness of the estimation. This is justified by the observation that most of the degradations decrease the sparseness of the representation (e.g., wavelets) with respect to the original image [120, 121, 24]. Figure 5.1 shows an example. Left column corresponds to a crop of *Peppers* image (top) and a high-frequency sub-band of the linear response using DT-CWT with this image (bottom). Right column corresponds to randomly missing 40% of the image pixels (top) and the corresponding sub-band (bottom). We note that the energy is less concentrated in the degraded sub-band.

The two observations made about natural images in Chapter 2 (energy compaction of the linear response and sparseness increase using non-linear methods) lead us to two different variants to describe the *a priori* knowledge

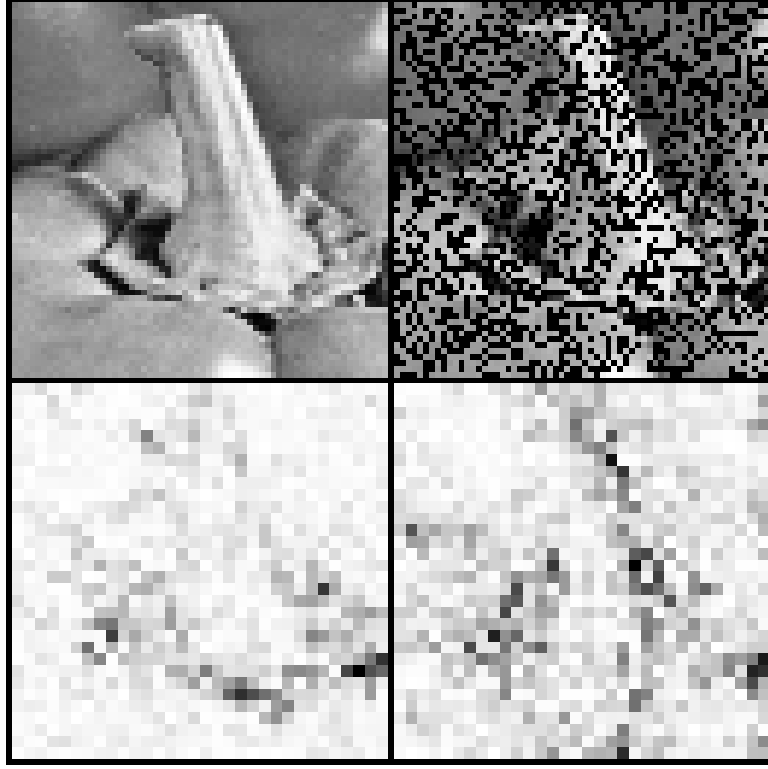


Figure 5.1: **Top-left**, Peppers crop, starting at row 111, column 91. **Bottom-left**, same crop of the high frequency sub-band of the linear response to Peppers using 8-scales DT-CWT, corresponding to orientation  $-45^\circ$ . We have previously doubled the size of this sub-band, through pixel replication, in order to match the image size. **Top-right**, degraded image by setting to zero, randomly, 40% of the pixels. **Bottom-right**, corresponding sub-band.

we have about them. On the one hand, we can assume that the original image can be expressed as a linear combination of few representation vectors. This implies a sparse vector of synthesis coefficients. Many authors have previously used this concept to approach image restoration [26, 29, 61]. We call this Synthesis-sense Sparseness (SS).

Although the use of the SS approach is perfectly legitimate and reasonable successful in practice, one could object to the lack of a direct empirical basis. The traditional Bayesian approximation to image restoration is based on building *a priori* models of the image reflecting the typical behaviour of the signals in many previous observations. However, the synthesis coefficients of an optimally sparse representation cannot be observed directly, and, often, they cannot be exactly calculated in practice (because, in general, the global optimum solution is not available).

Following this reasoning, it seems conceptually more consistent to use an

*a priori* statistical model based on direct observations, describing the typical distribution of the coefficients of the linear transformation of natural images. We call this Analysis-sense Sparseness (AS). This is a natural extension of many previous works that, under different points of view, have used sparse density models for the linearly transformed image (e.g., [74, 18]). Moreover, some authors have implemented practical methods based on AS for image processing, with very positive results (e.g., [19, 20]). As it will be seen in the next chapter, we have experienced a generally better restoration performance using AS than SS, in agreement with [122].

In this chapter we show how to apply the methods presented in previous chapters to restoration of *a posteriori* deterministic degradations. We have observed, for each method, which sparseness type is better in practice. Consequently, we have used  $\ell_p$ -AP method for SS-restoration and  $\ell_p$ -GM for AS.

We start by explaining and formulating the consistency set with a generic degraded observation (Section 5.1). Then we formulate the SS restoration problem (Section 5.2). Next we show how to adapt  $\ell_p$ -AP to solve it (Section 5.3). Then we formulate the AS problem (Section 5.4), and show how to adapt  $\ell_p$ -GM to solve it (Section 5.5).

## 5.1 Consistency with an observation

We have a degraded image  $\mathbf{y} \in \mathbb{R}^N$ . Consider that the degradation consists of missing some identifiable pieces of information. This could be bits, pixels, chromatic components, etc. We assume that we can exactly know, given  $\mathbf{y}$ , which elements of the original image are preserved (*a posteriori* deterministic degradations). Then, we can replicate the associated degradation, noted as:

$$\mathbf{y} = f_{\mathbf{y}}(\mathbf{x}). \quad (5.1)$$

We define the consistency set for observation  $\mathbf{y}$ ,  $R(\mathbf{y})$ , as all those images which, after being degraded with  $f_{\mathbf{y}}(\mathbf{x})$ , result in the same observation. Mathematically:

$$R(\mathbf{y}) = \{\mathbf{x} \in \mathbb{R}^N : f_{\mathbf{y}}(\mathbf{x}) = \mathbf{y}\}.$$

## 5.2 Formulation using synthesis-sense sparseness

If we are dealing with a redundant representation domain, we have that:

$$\mathbf{y} = f_{\mathbf{y}}(\Phi \mathbf{a}),$$

where  $\mathbf{a}$  is a synthesis vector whose reconstruction provides the original image. The *Maximum A Posteriori* estimate of  $\mathbf{a}$  is given by:

$$\hat{\mathbf{a}} = \arg \min_{\mathbf{a}} \{\log p(\mathbf{a}) + \lambda \|\mathbf{y} - f_{\mathbf{y}}(\Phi \mathbf{a})\|_2^2\},$$

where  $p(\mathbf{a})$  is a prior for the representation coefficients. As is common in the literature when using, as in our case, wavelets (e.g., [3, 74, 78, 29]), we assume independent coefficients and heavy-tailed priors, such as the Generalized Gaussian density :

$$p(\mathbf{a}) \propto \exp\{-k \|\mathbf{a}\|_p^p\}.$$

When  $0 \leq p \leq 1$ , this distribution is *sparse*, in the sense of having a probability density function concentrating most of the coefficients around zero, and having a small proportion of them with relatively high amplitudes. The logarithm of this prior is proportional to the  $p$ -th power of the  $\ell_p$ -norm of the vector plus some irrelevant constant ( $\log p(\mathbf{a}) \propto \|\mathbf{a}\|_p^p + A$ ). Then, our optimisation problem is set up as follows:

$$\hat{\mathbf{a}} = \arg \min_{\mathbf{a}} \{\|\mathbf{a}\|_p^p + \lambda \|\mathbf{y} - f_{\mathbf{y}}(\Phi \mathbf{a})\|_2^2\}.$$

Note that this is analogous to the  $\ell_p$ -norm minimisation problem in Equation (3.1), but the fidelity is measured in terms of the residual between the observation and the degradation of the estimation. Therefore, the cost function to be minimised is formed by the addition of two terms, one corresponding to the sparseness of the approximation, and the other to the quadratic distance to the consistency set  $R(\mathbf{y})$ . The parameter  $\lambda$  controls the relative importance of each term in the final solution. In practice, we require our estimation to belong to  $R(\mathbf{y})$ , same as the original image. Equivalently, we require  $\hat{\mathbf{a}}$  to be inside the set  $S(\mathbf{y})$  of synthesis vectors representing images in  $R(\mathbf{y})$ :

$$S(\mathbf{y}) = \{\mathbf{a} \in \mathbb{R}^M : \Phi \mathbf{a} \in R(\mathbf{y})\}.$$

Then, we set  $\lambda$  to infinite, which yields the following problem:

$$\begin{aligned} \hat{\mathbf{a}} &= \arg \min_{\mathbf{a}} \|\mathbf{a}\|_p^p \text{ s.t. } \mathbf{a} \in S(\mathbf{y}), \\ \hat{\mathbf{x}} &= \Phi \hat{\mathbf{a}}. \end{aligned} \tag{5.2}$$



### 5.3 Estimation using $\ell_p$ -AP and synthesis-sense sparseness

The solution to Equation (5.2) has a certain  $\ell_p$ -norm,  $\|\hat{\mathbf{a}}\|_p^p = R^*$ . This value can be found by solving the problem:

$$R^* = \min\{R \in \mathbb{R}^* : B_p(R) \cap S(\mathbf{y}) \neq \emptyset\}.$$

The intersection between the corresponding sets,  $B_p(R^*)$  and  $S(\mathbf{y})$ , will have more than one element in general. Among them, we choose the closest one to the observation:

$$\hat{\mathbf{a}} = P_{S(\mathbf{y}) \cap B_p(\hat{R}^*)}^\perp(\Phi^T \mathbf{y}).$$

It is easy to see that we can use  $\ell_p$ -AP (see Chapter 3) to solve this problem, only substituting the set  $S(\Phi, \mathbf{x})$  with the set  $S(\mathbf{y})$ . Then, we obtain the following iterations:

$$\begin{aligned} \hat{\mathbf{a}}^{(0)} &= P_{B_p(\hat{R}^*)}^\perp(\mathbf{a}^{LS}), \\ \hat{\mathbf{a}}^{(k+1)} &= P_{B_p(\hat{R}^*)}^\perp(P_{S(\mathbf{y})}^\perp(\hat{\mathbf{a}}^{(k)})). \end{aligned} \quad (5.3)$$

Iterations end when  $\|\hat{\mathbf{a}}^{(k+1)} - \hat{\mathbf{a}}^{(k)}\|_2 < \delta$ , for  $\delta > 0$ . The proof that the fixed point of these iterations is a local minimum to the distance to  $S(\mathbf{y})$  is completely analogous to that shown in Section 3.1 for the case of sparse approximation.

We derive now the expression of the orthogonal projection of a vector  $\mathbf{b}^o$  onto  $S(\mathbf{y})$ :

$$\hat{\mathbf{b}}_{S(\mathbf{y})}^p = P_{S(\mathbf{y})}^\perp(\mathbf{b}^o) = \arg \min_{\mathbf{b}} \{\|\mathbf{b} - \mathbf{b}^o\|_2^2 \text{ s.t. } \mathbf{b} \in S(\mathbf{y})\}. \quad (5.4)$$

$S(\mathbf{y})$  is orthogonal to the set  $A$  of linear responses to images, defined as:

$$A = \{\mathbf{b} : \exists \mathbf{x} \in \mathbb{R}^N, \Phi^T \mathbf{x} = \mathbf{b}\}.$$

Then, analogously to the case of the affine space of perfect reconstruction of an image (see Equation (3.4)), we have that:

$$\hat{\mathbf{b}}_{S(\mathbf{y})}^p = \mathbf{b}^o + \Phi^T \Phi [\hat{\mathbf{b}}_{S(\mathbf{y})}^p - \mathbf{b}^o].$$

We now define  $S_A(\mathbf{y})$  as the set of linear responses whose reconstruction is consistent with the observation:

$$S_A(\mathbf{y}) = \{\mathbf{b} \in R^M : \exists \mathbf{x} \in R(\mathbf{y}), \Phi^T \mathbf{x} = \mathbf{b}\}.$$

We have that  $\hat{\mathbf{b}}_{S_A(\mathbf{y})}^p = P_{S_A(\mathbf{y})}^\perp(\hat{\mathbf{b}}_{S(\mathbf{y})}^p) = \Phi^T \Phi \hat{\mathbf{b}}_{S(\mathbf{y})}^p$ , and thus:

$$\hat{\mathbf{b}}_{S(\mathbf{y})}^p = \mathbf{b}^o + \hat{\mathbf{b}}_{S_A(\mathbf{y})}^p - \Phi^T \Phi \mathbf{b}^o. \quad (5.5)$$

To solve  $\hat{\mathbf{b}}_{S_A(\mathbf{y})}^p$ , we have to derive an expression of the orthogonal projection onto  $S_A(\mathbf{y})$  in terms of our observation  $\mathbf{b}^o$ . We have that:

$$\hat{\mathbf{b}}_{S_A(\mathbf{y})}^p = \arg \min_{\mathbf{b}} \{ \|\mathbf{b}^o - \mathbf{b}\|_2^2 \text{ s.t. } \mathbf{b} \in S_A(\mathbf{y}) \}.$$

We can express:

$$\mathbf{b}^o - \mathbf{b} = (\mathbf{b}^o - \Phi^T \Phi \mathbf{b}^o) + (\Phi^T \Phi \mathbf{b}^o - \mathbf{b}).$$

These two bracketed differences are orthogonal vectors, as the first one belongs to the null space of  $\Phi$ , whereas the second one has no null component in  $\Phi$  (that is,  $\Phi^T \Phi \mathbf{b} = \mathbf{b}$ , because  $\mathbf{b} \in S_A(\mathbf{y})$ ). Then, we can write:

$$\hat{\mathbf{b}}_{S_A(\mathbf{y})}^p = \arg \min_{\mathbf{b}} \{ \|\Phi^T \Phi \mathbf{b}^o - \mathbf{b}^o\|_2^2 + \|\mathbf{b} - \Phi^T \Phi \mathbf{b}^o\|_2^2 \text{ s.t. } \mathbf{b} \in S_A(\mathbf{y}) \}.$$

As the first summation term is independent from  $\mathbf{b}$ , it can be ignored in the minimisation, resulting in:

$$\hat{\mathbf{b}}_{S_A(\mathbf{y})}^p = \arg \min_{\mathbf{b}} \{ \|\mathbf{b} - \Phi^T \Phi \mathbf{b}^o\|_2^2 \text{ s.t. } \mathbf{b} \in S_A(\mathbf{y}) \}.$$

We know that, for every vector  $\mathbf{b} \in S_A(\mathbf{y})$ , another vector  $\mathbf{x} \in R(\mathbf{y})$  exists such that  $\mathbf{b} = \Phi^T \mathbf{x}$ . Thus, substituting in previous expression, we get:

$$\hat{\mathbf{b}}_{S_A(\mathbf{y})}^p = \Phi^T [\arg \min_{\mathbf{x}} \{ \|\Phi^T \mathbf{x} - \Phi^T \Phi \mathbf{b}^o\|_2^2 \text{ s.t. } \mathbf{x} \in R(\mathbf{y}) \}].$$

And because  $\Phi^T$  is a Parseval frame:

$$\hat{\mathbf{b}}_{S_A(\mathbf{y})}^p = \Phi^T [\arg \min_{\mathbf{x}} \{ \|\mathbf{x} - \Phi \mathbf{b}^o\|_2^2 \text{ s.t. } \mathbf{x} \in R(\mathbf{y}) \}].$$

The minimisation in  $\mathbf{x}$  corresponds to the orthogonal projection of  $\Phi \mathbf{b}^o$  onto the set of images consistent with the observation,  $P_{R(\mathbf{y})}^\perp(\Phi \mathbf{b}^o)$ , so we obtain:

$$P_{S_A(\mathbf{y})}^\perp(\mathbf{b}^o) = \Phi^T P_{R(\mathbf{y})}^\perp(\Phi \mathbf{b}^o). \quad (5.6)$$

And, substituting in Equation (5.5) we finally have that:

$$P_{S(\mathbf{y})}^\perp(\mathbf{b}^o) = \mathbf{b}^o + \Phi^T (P_{R(\mathbf{y})}^\perp(\Phi \mathbf{b}^o) - \Phi \mathbf{b}^o).$$

Finding the orthogonal projection onto the consistency set  $R(\mathbf{y})$  is trivial for a wide number of strictly reproducible *a posteriori* degradations, by simply forcing the reconstruction to preserve the desired values. Of course, the precise form of this projection depends on each degradation. We explain some cases in detail in Chapter 6.

## 5.4 Formulation using analysis-sense sparseness

We cannot reach, in general, strict sparseness when dealing with linear responses of natural images, because we cannot avoid the simultaneous response of several coefficients to the same feature. Instead, we can consider that most of the energy is concentrated in a small proportion of coefficients. Then, we can model the linear representation as a strictly sparse vector whose support corresponds to the highest responses in amplitude, called  $\mathbf{a}$ , plus a Gaussian correction term, noted  $\mathbf{r}$ . Then, if we define  $S_A(\mathbf{y})$  as the set of linear responses whose reconstruction is consistent with the observation:

$$S_A(\mathbf{y}) = \{\mathbf{b} \in R^M : \exists \mathbf{x} \in R(\mathbf{y}), \Phi^T \mathbf{x} = \mathbf{b}\},$$

then, we can write our optimization problem as:

$$\begin{aligned} (\hat{\mathbf{a}}, \hat{\mathbf{r}}) &= \arg \min_{\mathbf{a}, \mathbf{r}} \{\|\mathbf{a}\|_p^p + \lambda \|\mathbf{r}\|_2^2 \text{ s.t. } (\mathbf{a} + \mathbf{r}) \in S_A(\mathbf{y})\}, \\ \hat{\mathbf{x}} &= \Phi(\hat{\mathbf{a}} + \hat{\mathbf{r}}). \end{aligned} \quad (5.7)$$

## 5.5 Estimation using $\ell_p$ -GM and analysis-sense sparseness

Following a completely parallel way to that of the sparse approximation problem, when solving Equation (5.7) we derive an expression which only depends on a vector  $\mathbf{b} = \mathbf{a} + \mathbf{r}$ , and where the constraint set,  $S(\Phi, \mathbf{x})$ , is substituted by the new constraint,  $S_A(\mathbf{y})$ . Note that this new set is no longer affine and that, therefore, we have to consider its curvature by projecting the cost function gradient onto its tangent hyperplane on every border point  $\mathbf{b}$ . This projection can be calculated as the limit:

$$\nabla^{S_A(\mathbf{y})} C_p(\mathbf{b}, \theta) = \lim_{\alpha \rightarrow 0} \frac{P_{S_A(\mathbf{y})}^\perp(\alpha \nabla C_p(\mathbf{b}, \theta))}{\alpha},$$

where  $P_{S_A(\mathbf{y})}^\perp$  is the orthogonal projection onto  $S_A(\mathbf{y})$  (Equation (5.6)), and where  $C_p(\mathbf{b}, \theta)$  indicates, with  $p = 0$  or  $p = 1$ , the cost functions defined in Equations (4.5) and (4.13), respectively. The gradient descent method is then formulated as:

$$\mathbf{b}^{(k+1)} = \mathbf{b}^{(k)} - \alpha \nabla^{S_A(\mathbf{y})} C_p(\mathbf{b}^{(k)}, \theta). \quad (5.8)$$

However, it is more convenient in practice to use the following simpler calculations in the estimation loop:

$$\mathbf{b}^{(k+1)} = P_{S_A(\mathbf{y})}^\perp \left( \mathbf{b}^{(k)} - \alpha \nabla C_p(\mathbf{b}^{(k)}, \theta) \right),$$

which ensures that the updated vector belongs to the consistency set  $S_A(\mathbf{y})$  for any value of  $\alpha$ . This update rule is equivalent to that of Equation (5.8) if the projection is linear.

Because of the similar structure of the minimisation problem described in Equation (4.2), and that described in Equation (5.7), we can apply the same strategy to look for a global minimisation of the cost function when  $p = 0$  or  $p = 1$ . This implies, in practice, that we get a significantly better restoration performance if we use an exponentially decaying threshold until reaching the desired value, and then we use that fixed threshold until convergence. We have empirically tested in different applications that the optimal final threshold in our optimisation is usually close to zero, as it was also indicated by [19, 20]. Thus, in the absence of any additional information source, an arbitrarily small threshold<sup>1</sup> is a suitable stopping criterion for the iterations.

---

<sup>1</sup>Too low thresholds demand more computation, so here there is again a trade-off between time and quality.

# Chapter 6

## Some applications

In this chapter we present several applications to restoration problems of the methods derived in this Thesis. We focus on *a posteriori* deterministic degradations (see Chapter 5). We motivate each problem and formulate the associated consistency set and the orthogonal projection onto it. Finally, we present some experiments showing that our methods are highly competitive for the applications studied.

In Section 6.1 our sparse approximation methods are applied to remove spatial quantisation artifacts. In Section 6.2 they are applied to estimation of missing pixels of the image. In Section 6.3 we study the interpolation of Bayer Color Filter Array mosaics. Finally, in Section 6.4 we approach the problem of increasing the details of images.

### 6.1 Removing quantisation artifacts

#### 6.1.1 Introduction

Spatial quantisation is an indispensable part of the capture of images with digital devices. Usually, the artifacts derived from it, as false contours and destruction of low-contrast texture, are close or even below the visibility threshold. However, in a number of situations they can become evident. For example, when the local luminance range of an image is stretched to inspect low-contrast details, or when blurry and quantised images are deconvolved, especially if there is little noise from other sources. It can also be useful as a previous step to extract sensitive local features, like the gradient of the luminance. Other possible applications are interpolating level curves in topographic or barometric maps, or using a reduced number of bits per pixel when there are not enough resources to perform a more advanced compression of the image.

Surprisingly enough, until very recently the removal of quantisation artifacts in the image domain (from now on, de-quantising) has received little attention in the scientific literature. In contrast, the quantisation in the transformed domain has been widely used, especially in the context of post-processing compressed images (e.g., [123, 124, 125, 126, 127]).

Nevertheless, during the last years there has been a growing interest in approaching the problem in the image domain. Up to our knowledge, the first work was [128], which used de-quantising as a previous step to edge detection. Recently, other methods have been published, based on iterating between some filtering operation and the correction of the difference with the original [129, 130]. But this type of strategies, though they result in efficient algorithms, are too simple to provide satisfactory results. In parallel to these works, we presented a method based on promoting the sparseness on a representation with redundant wavelets [24]. The selection of coefficients was made, in this method, by directly thresholding the analysis wavelet coefficients, so we can classify this technique within the greedy heuristics. Last three referred methods are described in this section in detail.

In this section we compare the performance of  $\ell_p$ -AP adapted to de-quantising in the two versions presented,  $p = 1$  and  $p = 0$ . We see, through intensive experiments, that  $\ell_0$ -AP significantly outperforms methods in [24, 129, 130], and also  $\ell_1$ -AP.

### 6.1.2 Consistency set

In this case the consistency set is made of those images that, when quantised using the same observed quantization levels, result in the same observation. Therefore, being  $\mathbf{y}$  a quantised observed image, the consistency set associated to it, noted as  $R_Q(\mathbf{y})$ , is defined as:

$$R_Q(\mathbf{y}) = \{\mathbf{x} \in \mathbb{R}^N : y_i - \frac{\delta_i}{2} < x_i \leq y_i + \frac{\delta_i}{2}, \forall i \in \{1, \dots, N\}\},$$

where  $\delta_i$  indicates the size of the quantisation interval associated to each pixel<sup>1</sup>.

Given an image  $\mathbf{x} \in \mathbb{R}^N$ , the orthogonal projection onto  $R_Q(\mathbf{y})$  is easily computed as  $\mathbf{z} = P_{R_Q(\mathbf{y})}^\perp(\mathbf{x})$ , where:

$$z_i = \begin{cases} x_i, & y_i - \frac{\delta_i}{2} < x_i \leq y_i + \frac{\delta_i}{2} \\ y_i - \frac{\delta_i}{2} + \epsilon, & x_i \leq y_i - \frac{\delta_i}{2} \\ y_i + \frac{\delta_i}{2}, & y_i + \frac{\delta_i}{2} < x_i, \end{cases}$$

<sup>1</sup>Here we have assumed uniform quantisation, for simplicity, but all the methods described here are easily applicable to other types of quantisation.

where  $\epsilon \in \mathbb{R}^*$  (ideally infinitesimal) is an artifice added to achieve empty intersection between closed adjacent quantisation intervals.

### 6.1.3 Implementation

In the experiments shown in next subsection we have compared our methods with three recent algorithms applied to quantisation artifacts removal. We briefly describe next these three methods and our implementation of them. The values given to the different parameters have been hand-optimised for every method to stop the iterations at a similar approximation level to the final convergence.

**Direct thresholding and optimisation.** In [24] we describe a method for de-quantizing in the image domain, by enforcing a high degree of sparseness in its representation with a redundant wavelet-based dictionary. For this purpose we devise a linear operator that returns the minimum  $\ell_2$ -norm image preserving a set of significant coefficients, and estimate the original by minimising the cardinality of that subset, always ensuring that the result is compatible with the quantized observation. We implement this solution by alternated projections onto convex sets. To select the set of significant coefficients, we threshold directly the amplitudes of the linear representation of the image, using a threshold proportional to the estimated energy of each original sub-band.

This application is based on the method that we call Direct Thresholding and Optimisation (DT+OP, see Chapter 3), and by extension this will be the name given to it here. Its details can be seen in [24]. In the following experiments we have used 7-scale DT-CWT and we have assumed that there is intersection between the sets when, in less than 30 iterations, the mean square difference of the subsequent projected vectors onto both sets is less or equal to 0.5.

**Constrained Diffusion,** In [130] the method of Constrained Diffusion (CD) is presented, based on combining linear filtering with non-linear correction of the difference with the observation.

Our implementation of the method follows the steps explained in [130], that is, an iterative method initialising the estimate with the observation and having two steps: 1) Convolution of the estimated image with the following matrix:

$$\begin{pmatrix} 0 & \frac{1}{5} & 0 \\ \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \\ 0 & \frac{1}{5} & 0 \end{pmatrix}, \quad (6.1)$$

and 2) projection onto the consistency set with the observation. The

iterations ended when the mean square difference of the estimation of one iteration with respect to the previous one is less than 0.5.

**Regularised Constrained Iterative Restoration.** In [129] it is described a method to remove quantisation artifacts, called Regularised Constrained Iterative Restoration (RCIR). The method is described applied to vector quantisation. However, when dealing with uniform quantisation, as in our case, it is reduced to a simple strategy based on minimising the  $\ell_2$ -norm of the output of a high-pass filter, enforcing, at each step, again that the estimation belongs to the consistency set.

Our implementation results in an iterative method, initialised with the observation at the first estimation. It consists of two steps: 1) Subtract from the estimation its own convolution with a Laplacian spatial filter. 2) Project onto the consistency set. The stopping criterion for the iterations is similar to that of CD.

**$\ell_p$ -AP.** We use a binary search to find the radius of the smaller  $\ell_p$ -ball having non-empty intersection with the set of vectors consistent with the observation (see Chapter 5), whose interval is initialised between 0 and  $M$  in  $\ell_0$ -AP and between 0 and  $\|\Phi^T \mathbf{y}\|_1$  in  $\ell_1$ -AP. We consider that we have found the required radius when the search interval has a length less or equal to 5000, and we assume that there exists intersection when in less than 30 iterations the mean square difference between the projected vector onto one of the sets and the other is less or equal to<sup>2</sup> 0.3.

## 6.1.4 Results and discussion

### 6.1.4.1 $\ell_0$ -AP vs. $\ell_1$ -AP

Figure 6.1 compares the de-quantising performance of  $\ell_0$ -AP vs.  $\ell_1$ -AP, using 3-bits quantisation of *Einstein* image. Results are shown using both 8-scale DT-CWT and 6-scale Curvelets. We observe that the performance of  $\ell_0$ -AP is much better, both in PSNR increase and visually: it manages to remove the quantisation discontinuities but preserves a high definition in the original edges of the image. We also see that results using Curvelets are slightly better in this case than those obtained using DT-CWT.

Figure 6.2 shows a similar example with a little texture image (*Peppers*). Note that the relative behaviour between the methods is qualitatively similar, although the difference in visual quality between  $\ell_1$ -AP and  $\ell_0$ -AP is not so big in this case as in the *Einstein* example. This is due to this image has less high-frequency texture, and then over-smooth estimations

---

<sup>2</sup>The higher tolerance given to methods based on linear filtering favour them, because making more iterations will smooth too much the estimation, thus decreasing the PSNR.



are having little visual impact. However, the difference in PSNR is still favourable to  $\ell_0$ -AP. Note also that now the performance is better using DT-CWT than Curvelets.

Regarding the behaviour of  $\ell_p$ -GM, we have experienced that it is not very satisfactory for both  $p = 0$  and  $p = 1$ . Using AS it is not able to remove the quantisation artifacts; and using SS too smooth estimations are obtained. In the latter case, the compaction capacity of  $\ell_0$ -GM is much better than that of  $\ell_0$ -AP, but this does not reflect in a parallel way on the method's performance.

#### 6.1.4.2 $\ell_0$ -AP vs. existing methods

Now we compare the  $\ell_0$ -AP performance vs. that of the methods previously described: RCIR [129], CD [130], and DT+OP [24]. Table 6.1 shows the performance average (obtaining the MSE for the average) of each method in our test set (see Appendix A) for all the range of possible quantisation bits in 8-bits images. We can see that methods based on enforcing sparseness, DT+OP and  $\ell_0$ -AP, outperform clearly those based on simpler linear operations, except when the image is quantised with only one bit. Note that  $\ell_0$ -AP is the best for those levels which, in practice, result in visible artifacts (low and medium range).

# Bits	PSNR (dB)						
	1	2	3	4	5	6	7
Observed	16.40	22.73	29.22	34.71	40.59	46.45	51.11
RCIR	<i>17.62</i>	24.25	29.75	34.37	39.20	44.56	51.87
CD	<b>17.88</b>	<i>24.33</i>	29.60	34.38	39.30	44.65	<i>51.89</i>
DT+OP	16.46	24.05	<i>30.83</i>	<i>35.67</i>	<b>40.61</b>	<b>45.28</b>	49.27
$\ell_0$ -AP	17.29	<b>24.74</b>	<b>31.49</b>	<b>35.91</b>	<i>40.21</i>	<i>45.11</i>	<b>52.03</b>

Table 6.1: PSNR (MSE averaged) using the images in our test set, quantised with all the possible range of bits and restored using methods RCIR, CD, DT+OP and  $\ell_0$ -AP. First row corresponds to the averaged PSNR of the observed images. Bold numbers indicate the best result for each number of bits, and italic the second best.

In addition, the visual appearance of the results of methods based on promoting sparseness is significantly better than those of their competitors, even for low number of quantisation bits. In Figures 6.3 and 6.4 we can see a visual comparison of the application of the methods to *Einstein* and *Peppers* images quantised with 3-bits. Among them, the best one, both visually and in PSNR, is  $\ell_0$ -AP. Both RCIR and CD destroy too many high-frequency

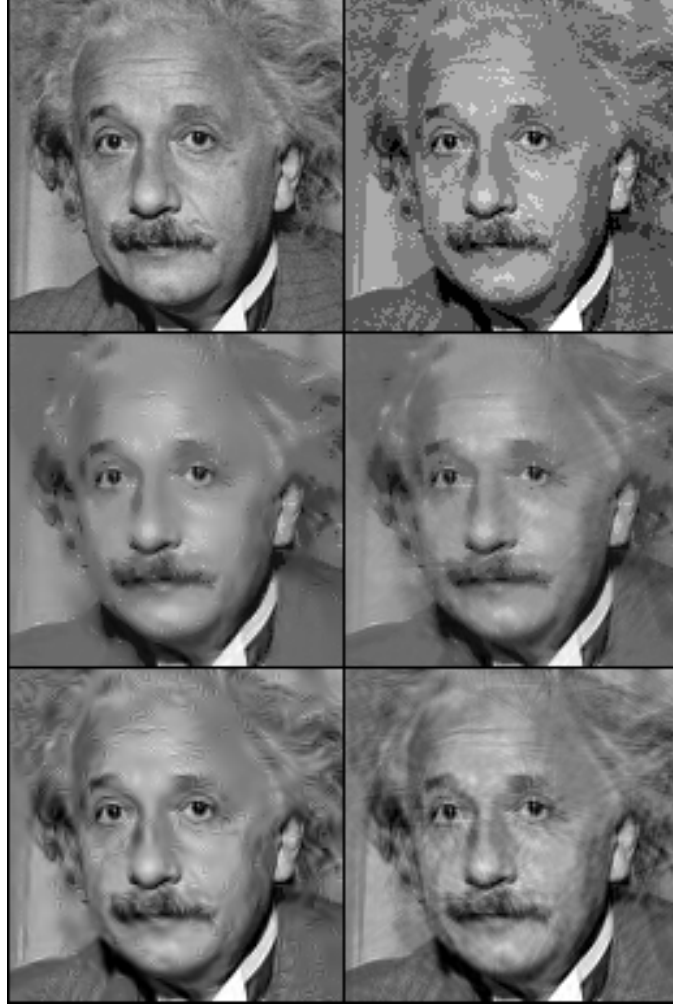


Figure 6.1: Example of application of  $\ell_1$ -AP and  $\ell_0$ -AP to de-quantizing. **Top-left**, original Einstein image, cropped to  $128 \times 128$  pixels. **Top-right**, 3-bits observed quantisation (PSNR: 27.98 dB). **Centre-left**,  $\ell_1$ -AP result using 8-scale DT-CWT (30.17 dB). **Centre-right**,  $\ell_1$ -AP result using 6-scale Curvelets (30.61 dB). **Bottom-left**,  $\ell_0$ -AP result using 8-scale DT-CWT (31.21 dB). **Bottom-right**,  $\ell_0$ -AP result using 6-scale Curvelets (31.38 dB).

components without removing completely the artifacts. Using an adaptive threshold allows DT+OP to remove isolated elementary functions (present in  $\ell_0$ -AP). However, presence of *ringing* effect and poor performance in areas of high-frequency texture contribute to the fact that the visual effect (and also the PSNR) is better in  $\ell_0$ -AP. Finally, we show in the last panel the result of  $\ell_0$ -AP using a joint DT-CWT - Curvelets representation (see Appendix C). Note that increasing the richness of the dictionary, not

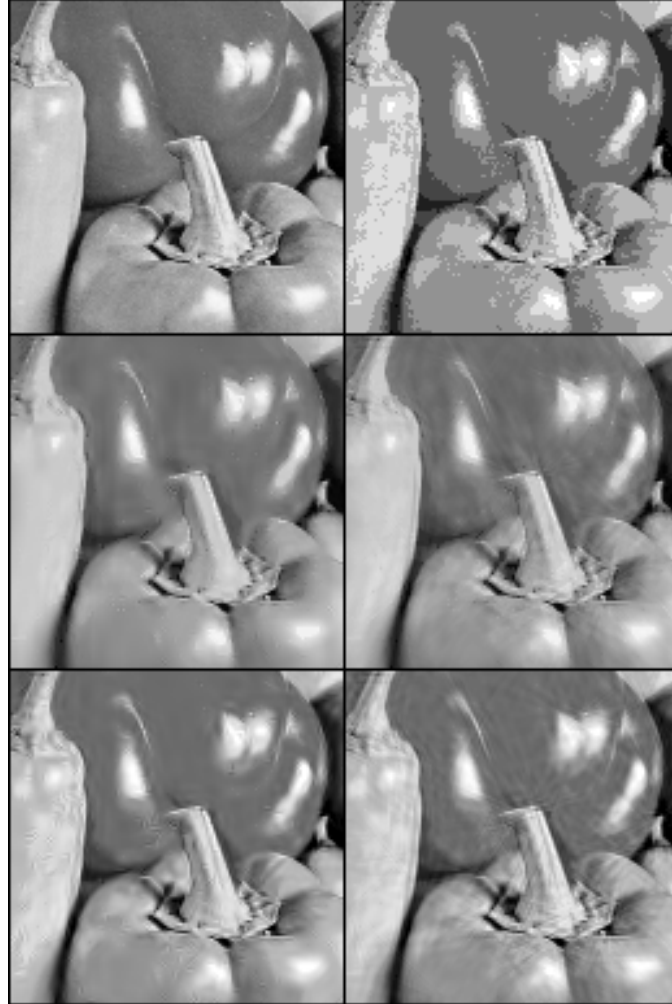


Figure 6.2: Example of application of  $\ell_1$ -AP and  $\ell_0$ -AP to de-quantizing. **Top-left**, original Peppers image, cropped to  $128 \times 128$ . **Top-right**, 3-bits observed quantisation (PSNR: 28.81 dB). **Centre-left**,  $\ell_1$ -AP result using 8-scale DT-CWT (29.08 dB). **Centre-right**,  $\ell_1$ -AP result using 6-scale Curvelets (29.50 dB). **Bottom-left**,  $\ell_0$ -AP result using 8-scale DT-CWT (31.06 dB). **Bottom-right**,  $\ell_0$ -AP result using 6-scale Curvelets (30.85 dB).

only in number but also in type of elementary functions used, significantly improves the result, increasing the PSNR and drastically reducing the isolated elementary functions appearing in the estimated image when only one representation is used. We have chosen these two dictionaries to fairly compare with the results shown in 6.1 using each representation separately. Moreover, using other dictionaries we can further improve the results. For example, using Curvelets and a version of the Steerable Pyramid [131]

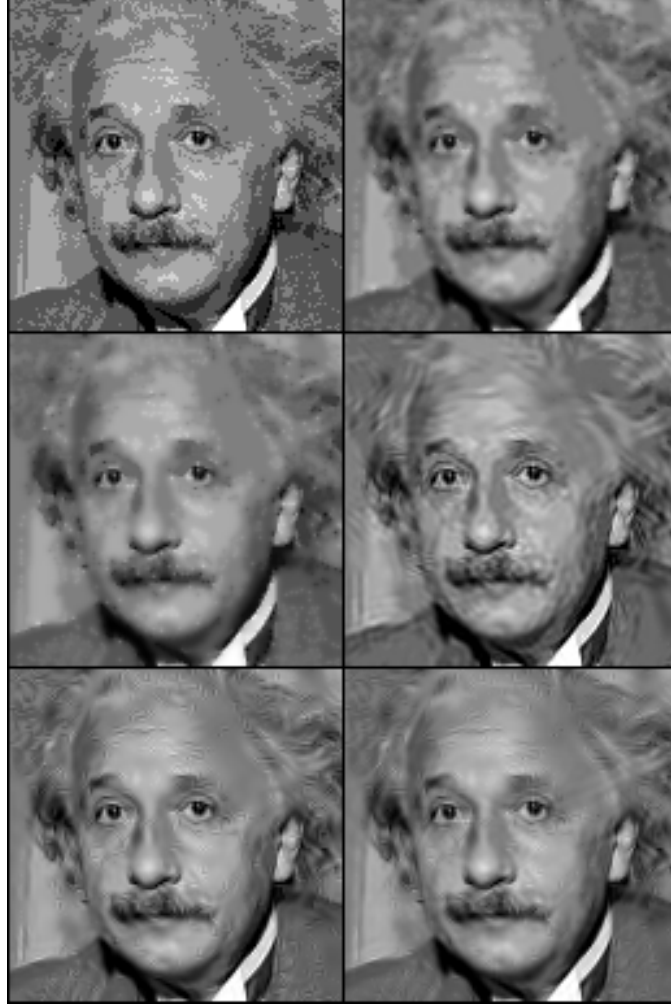


Figure 6.3: **Top-left**, Einstein quantised with 3 bits and cropped to  $128 \times 128$ , (PSNR: 27.98 dB). **Top-right**, RCIR result (30.39 dB). **Centre-left**, CD result (30.44 dB). **Centre-right**, DT+OP result using DT-CWT with 8 scales (30.72 dB). **Bottom-left**,  $\ell_0$ -AP result using DT-CWT with 8 scales (31.21 dB). **Bottom-right**,  $\ell_0$ -AP result using jointly 8-scale DT-CWT and 6-scale Curvelets, with equal scale factor,  $\sqrt{\frac{1}{2}}$  (31.93 dB).

without high-pass-residual, the PSNR is 31.99 dB for *Einstein* and 31.46 dB for *Peppers*.

Note, in Table 6.1, that for a high number of quantisation bits, there is a relative decrease in PSNR when using  $\ell_0$ -AP. However, in these cases this method also manages to remove low-contrast artifacts, improving the visual appearance when enhancing the contrast of the image. Figure 6.5 shows an example. Left panel is a  $32 \times 32$  detail of a smooth area in a photographic

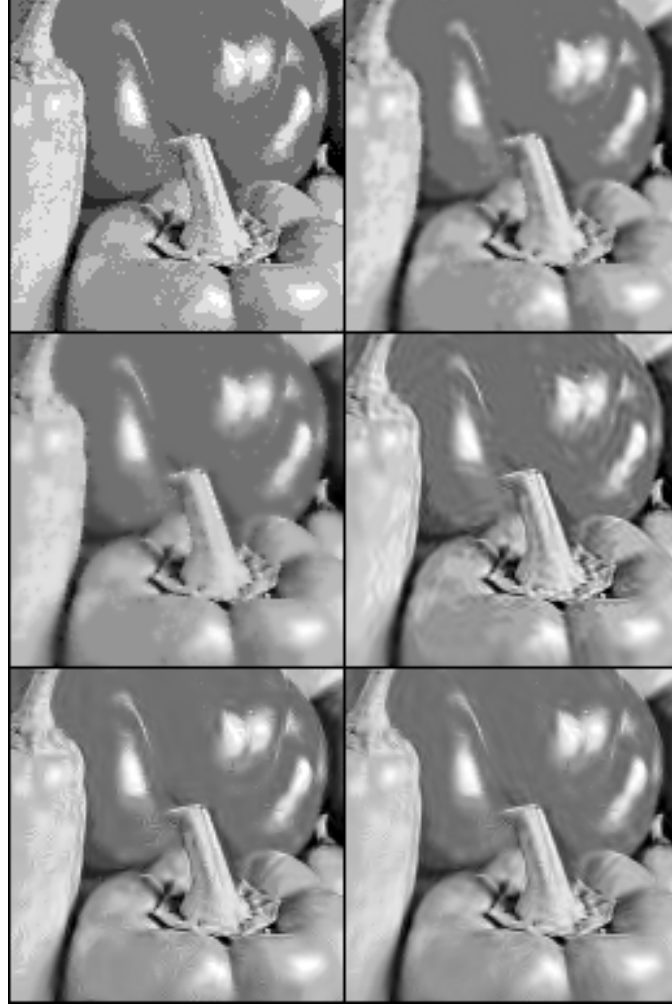


Figure 6.4: **Top-left**, *Peppers* quantised with 3 bits and cropped to  $128 \times 128$ , (PSNR: 28.81 dB). **Top-right**, RCIR result (29.65 dB). **Centre-left**, CD result (29.85 dB). **Centre-right**, DT+OP result using DT-CWT with 8 scales (30.38 dB). **Bottom-left**,  $\ell_0$ -AP result using 8-scale DT-CWT (31.07 dB). **Bottom-right**,  $\ell_0$ -AP result using jointly 8-scale DT-CWT and 6-scale Curvelets, with equal scale factor,  $\sqrt{\frac{1}{2}}$  (31.46 dB).

8-bits image, with a contrast amplification factor around 40 times. Right panel is the same crop in the result after processing with  $\ell_0$ -AP. Note the more natural appearance.

Regarding computation time, because of their simplicity, RCIR and CD are clearly faster. They are iterative methods only requiring a convolution and a projection onto the consistency set at each iteration. The methods based on promoting sparseness are dominated by one analysis and one

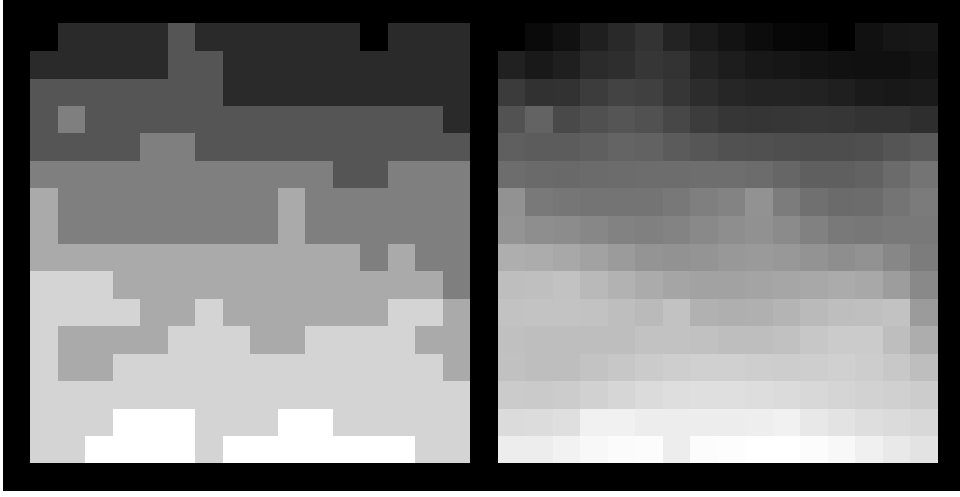


Figure 6.5: **Left**, detail of the sky of a photographic 8-bits image with contrast amplified approximately 40 times. **Right**, same detail after processing with  $\ell_0$ -AP.

synthesis operation per iteration, and they also have to look for the value of the threshold preserving a certain number of coefficients. Thus, both RCIR and CD perform around 10 iterations for each estimation, taking barely half a second in average, for a  $256 \times 256$  image. DT+OP takes between 15 seconds and 5 minutes using DT-CWT with 8 scales, depending on the number of quantisation bits of the observed image. Finally,  $\ell_0$ -AP takes 80 seconds on average using DT-CWT with 8 scales. Times are for our MATLAB® implementatoin on an Intel®, *Core*<sup>TM</sup>2 Duo with 1.66 GHz and 2 GB RAM.

### 6.1.5 Conclusions

We have analysed the performance of  $\ell_0$ -AP applied to removing spatial quantisation artifacts, comparing it to  $\ell_1$ -AP and to other recently published methods (RCIR, DT+OP CD). We can say that the search for the sparsest image within the consistency set made by  $\ell_0$ -AP provides very satisfactory results. In general, we have seen that methods based on promoting sparseness outperform those based on linear filtering operations. However,  $\ell_0$ -GM, which possesses a better energy compaction capacity, provides estimations too concentrated in low frequencies. The discussion about why this happens is left to a future work.

## 6.2 Interpolation of missing pixels

### 6.2.1 Introduction

Missing pixels in images is a common problem both in the capture and transmission of digital images. It is also usual wanting to remove some undesired details in an image (overprinted text, publicity, a disturbing cable in a beautiful landscape, etc.) or to restore images degraded by the pass of years.

In the last 30 years, or even more, many different techniques have been proposed to recover missing pixels (they are usually referred as *in-painting* techniques). On the other hand, texture-synthesis methods can also be used to fill-in missing regions. There are many papers using this latter type of strategies, and we can refer to [132, 133, 134, 34] among others. Unfortunately, the need to manually indicate the areas of the image from where the information needed for the interpolation should be taken, makes them inappropriate in practice. The most successful heuristic strategies combine the edge propagation (using partial differential equations, PDE) with local texture synthesis (e.g., [135, 136, 137, 138]).

There is a fast and very simple method providing comparable results to PDE-based methods [139]. It is based on iteratively combining a filtering linear operation and the non-linear constraint preserving the observed pixels.

Recently, some different strategies based on promoting sparseness have been developed. A good example of this is [20], proposing a formulation based on *Expectation-Maximisation* to approximate the sparsest solution consistent with the observation through optimal minimisation of the  $\ell_1$ -norm.

In this section we compare the performance of our methods ( $\ell_p$ -AP and  $\ell_p$ -GM) adapted to pixel interpolation of missing regions of the image. We see, through application examples, that  $\ell_0$ -GM provides the best results, among them, in MSE sense. We also compare to some of the referred techniques ([20] and [139]).

### 6.2.2 Consistency set

When the degradation is missing pixels in the image, the consistency set is composed by those images resulting in the same observation when missing the same pixels. Then, given a subset of fixed indices,  $I$ , from 1 to  $N$ , and given an observation  $\mathbf{y}$  preserving pixels  $y_i$  from the original image for all

$i \in I$ , we define the consistency set associated to  $\mathbf{y}$ ,  $R_I(\mathbf{y})$ , as:

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

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

### 6.2.3 $\ell_0$ -AP: new strategy for searching the radius

We have experienced that  $\ell_p$ -AP does not provide good interpolation results if we search for the sparsest image within the consistency set. In practice, the quality of the interpolation depends on finding a precise value of  $R$  allowing to minimise the error of our estimation. We call  $R_{opt}$  to this value. Then, what we are looking for is, given a sparseness level  $R_{opt}$ , the projection onto the consistency set of the image with coefficients belonging to the  $\ell_p$ -ball of radius  $R_{opt}$ , and whose pixels in  $I$  are closest, in a MSE sense, to the observed pixels in  $\mathbf{y}$ . To find this optimal value we propose a solution based on maximising the Mean Square Value (MSV) of the interpolated pixels [16]. Intuitively, we see that, for small values of  $R$ , only the more salient features of the observed image will be represented, so the estimation will be too smooth and we will necessarily obtain low MSV in the interpolated areas. When we choose a very high  $R$ , the broken edges caused by missing pixels will be better represented by using a lot of vectors that approximating them at lower scales. This provokes a poor interpolation and, once again, a low MSV. Finally, if we use intermediate values for  $R$ , we expect to have enough functions to represent the main features of the image, but not enough to describe false edges. Because of that, the missing holes will be filled with the appropriate dictionary functions, which will cause a higher MSV in the interpolated areas, and this is a better interpolation.

Bold line in Figure 6.6, corresponding to the left vertical axis, shows the normalised MSE in the estimated pixels for each value of  $R$ , where  $R$  is normalised by  $R_{opt}$ , which corresponds to the minimum (at one) of this curve. Dashed line shows the normalised MSV, corresponding to the right vertical axis. We call  $R_{max}$  to the value of  $R$  where this curve reaches its maximum. Dotted lines indicate the typical deviation of this curve for each value of  $R$ . Dashed-dotted line is the real MSV value of the lost pixels in the original image, which is an upper bound for the MSV of the estimation. All these values are averaged in our test set, using a randomly generated mask where approximately 40% of pixels are lost. For each test 250  $\ell_0$ -AP iterations were executed. The described method proposes to estimate  $R_{opt}$



from the observed value  $R_{max}$ . Then, for this percentage of missing pixels, we have that  $\hat{R}_{opt} = \frac{1}{0.7} R_{max}$ .

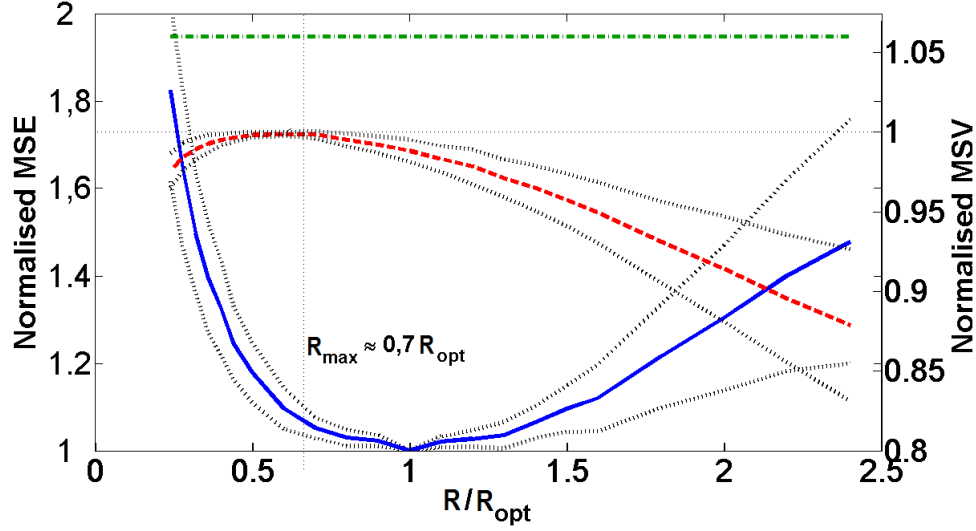


Figure 6.6: **Bold line - left axis**, Mean Square Error of the estimated pixels normalised to the minimum value of this curve, in ordinates, and to the value for which this minimum occurs, in abscissas. **Dashed line - right axis**, Normalised Mean Square Value of the estimated pixels. **Dotted line**, typical deviation for each value in the horizontal axis. **Dashed - dotted line**, Mean Square Value of the original pixels in the missing positions. All curves are averaged in our test set, using a random mask where approximately 40% of the pixels are lost.

#### 6.2.4 Implementation

**Fast-inpainting.** This method [139] simply consists of iteratively applying a linear filtering using the 2D convolving mask:

$$\begin{pmatrix} 0.073235 & 0.176765 & 0.073235 \\ 0.176765 & 0 & 0.176765 \\ 0.073235 & 0.176765 & 0.073235 \end{pmatrix}, \quad (6.2)$$

and a projection onto the consistency set to preserve the observed pixels. In our implementation, the iterations end when the mean square difference of the estimation at some iteration with respect to the one at previous iteration is less than  $10^{-3}$ .

**EM-inpainting.** This method [20] follows a convex relaxation approach. We have used the implementation available in MCALab 8.02 for MATLAB®, which can be downloaded from [140]. We have used the

values of the parameters described in [20]. The representation used is 6-scale Curvelets combined with Local DCT (LDCT) using a block size of  $32 \times 32$ . In [140] one can also find the analysis and synthesis functions for both representations.

**$\ell_p$ -AP and  $\ell_p$ -GM.** To apply  $\ell_p$ -AP to in-painting. We first estimate the optimal radius, for which several interpolations are required, and then we execute the method once again. We have used 100 iterations for each interpolation. Regarding  $\ell_p$ -GM we have experienced that the performance is not significantly improved when  $\beta$  is greater than 0.8. In addition, in this case there are no significant differences when using  $\alpha$  values greater than  $\alpha_0$ . We stop the iterations when the threshold is below 0.1.

As  $\ell_1$ -GM and *EM-inpainting* follow a very similar strategy (promoting sparseness through minimising the  $\ell_1$ -norm), we only present the results of that implementation giving us the best performance for each case.

## 6.2.5 Results and discussion

### 6.2.5.1 Missing random pixels (*filling-in*)

Here we consider that the pixels are independently lost with a given probability. Restoration of this type of degradations is usually known as *filling-in*.

We systematically compare  $\ell_0$ -GM (using SA) to  $\ell_1$ -GM (using also SA) and *Fast-inpainting*. Tables 6.2 and 6.3 show the averaged performance of each method with the images in our test set and for a wide range of the percentage of missing pixels. For  $\ell_0$ -GM, we use two different representations: 6-scale Curvelets alone and combined with LDCT using  $32 \times 32$  blocks. For  $\ell_1$ -GM, we only use the combined representation (providing the best results). For both  $\ell_0$ -GM and  $\ell_1$ -GM, we have used a scale factor of  $\sqrt{0.5}$  for both dictionaries. Note that by minimising the  $\ell_0$ -norm we obtain the best results, except for very high percentages of missing pixels. We see again that using combined incoherent dictionaries improves the quality of the estimation.

Figure 6.7 shows the images corresponding to the compared methods for randomly missing  $\approx 80\%$  of image pixels for *Barbara*. We observe that the method based on minimising the  $\ell_1$ -norm ( $\ell_1$ -GM) is not able to interpolate properly the missing pixels. On the other hand, the linear interpolation is totally unable to recover the texture. Surprisingly,  $\ell_0$ -GM achieves very good results, given the high percentage of missing pixels, both in smooth and texture areas. We can also see that using combined representations removes the artifacts inherent to the representation (isolated selected atoms).

Time per iteration for  $\ell_1$ -GM and  $\ell_0$ -GM are, obviously, very similar

Method	PSNR (dB)				
$\approx$ % missing pixels	10	20	30	40	50
Observation	24.05	20.90	19.13	17.90	16.82
<i>Fast-inpainting</i>	37.72	34.26	32.23	30.67	29.08
$\ell_1$ -GM (Curv.)	42.21	38.14	35.48	33.29	30.84
$\ell_0$ -GM (Curv.)	<i>43.00</i>	<i>39.15</i>	<i>36.67</i>	<i>34.72</i>	<i>32.56</i>
$\ell_0$ -GM (Curv.+LDCT)	<b>43.48</b>	<b>39.65</b>	<b>37.09</b>	<b>35.14</b>	<b>32.96</b>

Table 6.2: PSNR (through MSE average) when restoring the image in our test set after randomly missing different percentages of pixels. The PSNR of the observation has been calculated by using the global mean for lost positions. Both  $\ell_0$ -GM and  $\ell_1$ -GM use 6-scale Curvelets and LDCT with block-size  $32 \times 32$ , and the same scale factor for both of them ( $\sqrt{0.5}$ ).  $\ell_0$ -GM is also presented using only Curvelets. Bold numbers indicate the best method for each percentage, and italic the second best.

Method	PSNR (dB)			
$\approx$ % missing pixels	60	70	80	90
Observation	16.06	15.42	14.78	14.30
<i>Fast-inpainting</i>	27.78	26.48	<i>24.85</i>	<b>22.88</b>
$\ell_1$ -GM (Curv.)	28.75	26.38	22.84	18.54
$\ell_0$ -GM (Curv.)	<i>30.70</i>	<i>28.45</i>	24.79	<i>19.86</i>
$\ell_0$ -GM (Curv.+LDCT)	<b>31.03</b>	<b>28.77</b>	<b>25.14</b>	19.68

Table 6.3: Continuation of Table 6.2 for greater missing pixels percentages.

( $\approx 3$  min.). *Fast-inpainting* takes only around 0.5 seconds per image, so it is an alternative when there are not enough resources to apply  $\ell_0$ -GM.

### 6.2.5.2 Missing pixel areas (in-painting)

Recovery of missing areas in an image is more difficult than interpolating randomly missing pixels. Here we compare the application of  $\ell_0$ -GM to this problem against the methods described previously. Both the test images and the pixels masks used can be found in [140]. We have also downloaded from this page the results of *EM-inpainting*, enforcing the value of observed pixels in order to compare in the same conditions as the other methods<sup>3</sup>.

Figure 6.8 shows a particularly interesting example because missing pixels are localised both in smooth and textured areas. Top-left panel

<sup>3</sup>Note that enforcing the values of the observed pixels necessarily increases the PSNR of the estimation, whereas it may make some artifacts more visible.



Figure 6.7: Visual interpolation example of randomly missing pixels. **Top-left**, Barbara image, cropped to  $128 \times 128$ . **Top-right**, missing  $\approx 80\%$  of the pixels and filling them with the global mean (PSNR: 14.75 dB). **Centre-left**, interpolation made by  $\ell_1$ -GM (23.26 dB). **Centre-right**, result from Fast-inpainting (24.84 dB). **Bottom-left**,  $\ell_0$ -GM result using Curvelets with 6 scales (25.19 dB) **Bottom-right**, interpolation made by  $\ell_0$ -GM combining 6-scale Curvelets and LDCT with block size  $32 \times 32$ , and equal scale factors,  $\sqrt{0.5}$  (25.65 dB).

is the observation, with missing pixels filled in using the mean of the observed pixels. Top-right panel corresponds to *Fast-inpainting* (32.71 dB), and bottom-left to *EM-inpainting* (34.14 dB) using Curvelets and LDCT with  $32 \times 32$  block size. Last panel is  $\ell_0$ -GM result using Curvelets with 6 scales (34.92 dB). Note that, in contrast with the randomly missing pixels case, now the result based on minimising the  $\ell_1$ -norm is better than the



Figure 6.8: **Top-left**, Barbara image where value of missing pixels is the global mean of the observed ones (PSNR: 24.19 dB). **Top-right**, Fast-inpainting result (32.71 dB). **Bottom - left**, EM-inpainting result using 6-scale Curvelets and LDCT with block size  $32 \times 32$  (34.14 dB). **Bottom-right**,  $\ell_0$ -GM result using 6-scale Curvelets (34.92 dB).

one based on iterative linear filtering. Once again,  $\ell_0$ -GM provides the best performance. Figure 6.9 zooms in the second quadrant of the *EM-inpainting* (left) and  $\ell_0$ -GM (right) results. Partial recovery of the lost eye is particularly interesting in  $\ell_0$ -GM, but we also note a much better interpolation of the nose and the mouth.

Figure 6.10 is a practical example of restoration of old degraded photos. The image obtained in [140] was deformed, so we have stretched it with a 1.4 scale factor in the vertical direction. We have also removed last row as a requisite of the analysis and synthesis functions of LDCT, where the number of rows and columns should be multiple of half the block size used). The



Figure 6.9: **Left**, detail of the result of EM-inpainting shown in Figure 6.8 (PSNR: 34.38 dB). **Right**, same for the method  $\ell_0$ -GM (35.13 dB).

compared methods<sup>4</sup> appear in the same order as in Figure 6.8, but  $\ell_0$ -GM is now using 6-scale Curvelets combined with LDCT. Once again, we have enforced the observed values in the *EM-inpainting* result to have similar conditions for the comparison. We see again that  $\ell_0$ -GM is qualitatively better than the other two methods. In contrast with the other methods, the lower thick horizontal line is barely visible in  $\ell_0$ -GM. Also, faces of the girls, particularly the oldest and the youngest, are much better interpolated with  $\ell_0$ -GM. Figure 6.11 shows a detail of the results of *EM-inpainting* and  $\ell_0$ -GM for this image.

### 6.2.6 Conclusions

In this section we have applied the methods derived in this Thesis to the interpolation of missing pixels in the image. We have introduced a new heuristic to find the best radius for  $\ell_p$ -AP, based on maximising the MSV of the estimated pixels. This solution, in general, is less sparse than that obtained with the classical strategy (looking for the sparsest image inside the consistency set).

Our experiments show, however, that using  $\ell_0$ -GM with AS provides much better results. This is consistent with the model of promoting sparse solutions. We have compared to a very efficient method (*Fast-inpainting*) [139] based on combining linear and non-linear operations in

<sup>4</sup>The *EM-inpainting* result available in [140] has a different size to that of the observation and pixels mask, so we have replicated 3 times the first column.

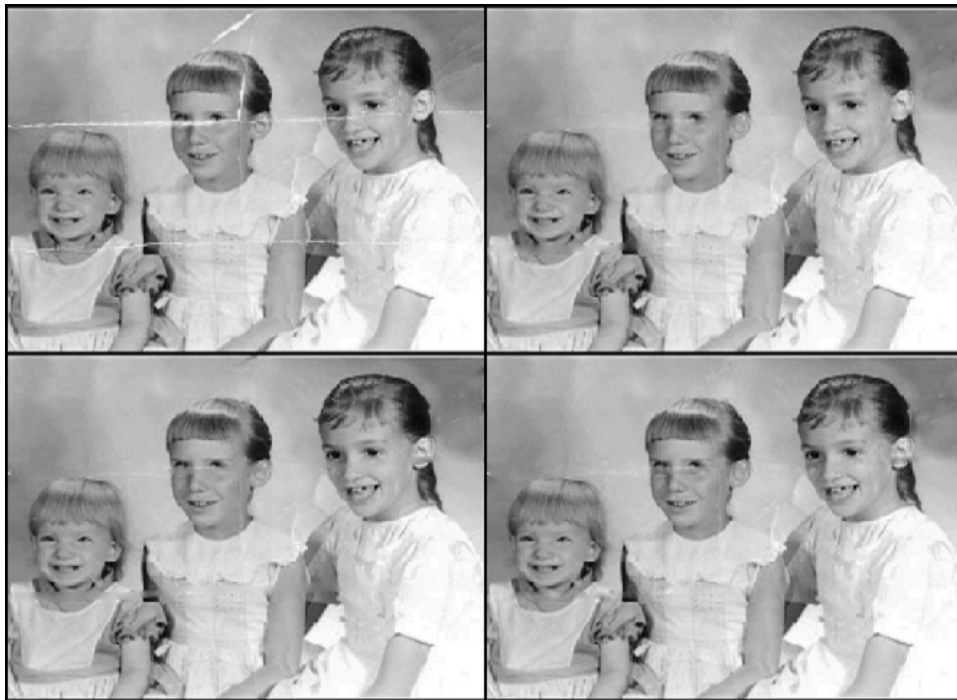


Figure 6.10: **Top-left**, damaged photographic image. **Top-right**, Fast-inpainting result. **Bottom-left**, EM-inpainting result using Curvelets and LDCT. **Bottom-right**, our result using  $\ell_0$ -GM with 6-scale Curvelets and LDCT using  $32 \times 32$  blocks and both scale factors to  $\sqrt{0.5}$ .

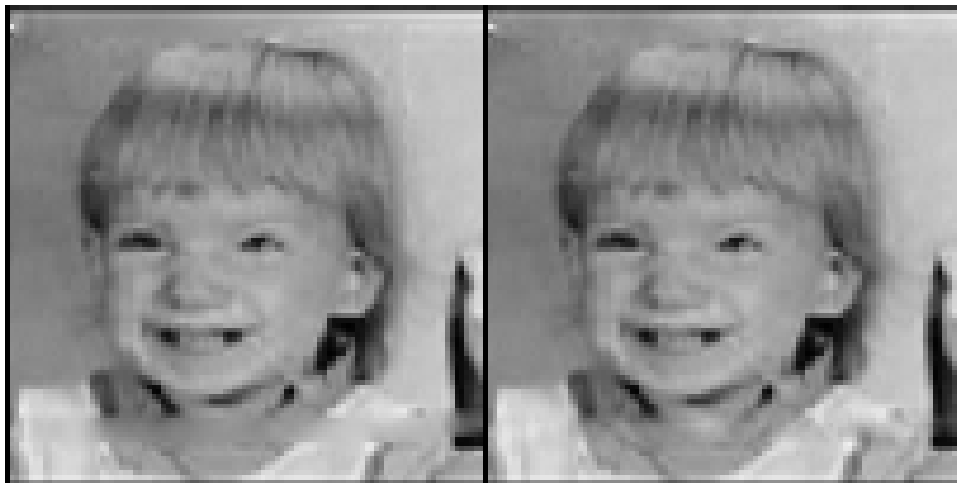


Figure 6.11: **Left**, detail of the result of EM-inpainting shown in Figure 6.10. **Right**, same for  $\ell_0$ -GM.

the image domain. We have also compared to methods based on promoting sparseness through the minimisation of the  $\ell_1$ -norm (*EM-inpainting* [20] and  $\ell_1$ -GM). In all the experiments,  $\ell_0$ -GM has clearly outperformed its competitors.

To conclude, we have observed, again, that methods based on finding suboptimal solutions to the minimisation of the  $\ell_0$ -norm behave substantially better than those minimising optimally the  $\ell_1$ -norm, for usual images and representations. Our results also outperform other heuristics. We should mention the good behaviour of  $\ell_0$ -GM with high randomly missing pixels percentages and with complicated richly textured areas. Using Curvelets clearly contributes to the success of edge/lines interpolation, because its atoms are very appropriate for elongated features. Moreover, we have seen that, usually, combination of Curvelets with LDCT further improves the results.

## 6.3 Spatial-chromatic interpolation in digital camera mosaics

### 6.3.1 Introduction

Most conventional digital cameras are based on the Colour Filter Array technology introduced by [141]. This means that they have a sensor capturing only one colour at each pixel. To reconstruct the complete colour image it is needed, in consequence, to interpolate the non-captured colour components at each position. This process is known as *de-mosaicing*.

There have been proposed very diverse techniques to solve this problem (see [142] for a review). In order to integrate the de-mosaicing as part of the digital image capturing process, to have a fast algorithm is very important, especially when considering the rapidly increasing resolution of CCD sensors. Bilinear interpolation, for example, is very fast, but it processes independently each colour channel, thus ignoring the correlation between them. This results in poor results [143]. On the other hand, high performance iterative methods are too slow, and thus inadequate for real time image capture [144, 145, 146]. However, their good results allow having high-quality images if it is possible to post-process them after the capture. Finally, some existing methods are based on linear filtering taking into account the inter-channel correlation, then trying to reach a good balance between computation and image quality [147, 148, 143, 149].

In this section we explore the performance of  $\ell_0$ -GM applied to de-mosaicing. Given its iterative nature, based on analysis and synthesis



operations with redundant dictionaries, we cannot expect to be competitive with other methods in terms of speed. However, the high quality of  $\ell_0$ -GM results makes an appealing alternative when we can post-process the image after capturing it. Firstly, we compare the results of  $\ell_0$ -GM to other methods proposed in this Thesis, and then to three state-of-the-art methods.

### 6.3.2 Consistency set

The degradation inherent to the Bayer mosaic is, similarly to the in-painting case, in missing "pixels" of the three-folded image formed by the three chromatic channels of the image. Therefore, the set of images consistent with the observation is analogous to that used in previous section, that is, it is formed by all those images preserving the observed colour components. Then, given a set of indices,  $I$ , taken from the interval  $\{1, \dots, 3N\}$ , and given an observation  $\mathbf{y} \in \mathbb{R}^{3N}$ , preserving all pixels  $y_i$  of the original image where  $i \in I$ , we define the consistency set associated to  $\mathbf{y}$ ,  $R_d(\mathbf{y})$  as:

$$R_d(\mathbf{y}) = \{\mathbf{x} \in \mathbb{R}^{3N} : x_i = y_i, \forall i \in I\}.$$

Given an RGB image  $\mathbf{x} \in \mathbb{R}^{3N}$  and a diagonal  $3N \times 3N$  matrix  $\mathbf{D}$  where each element  $d_{ii}$  is 1 if  $i \in I$  and 0 otherwise, the orthogonal projection of a vector  $\mathbf{v}$  onto  $R_d(\mathbf{y})$  is simply  $P_{R_d(\mathbf{y})}(\mathbf{v}) = \mathbf{D}\mathbf{y} + (\mathbf{I} - \mathbf{D})\mathbf{v}$ , where  $\mathbf{I}$  is the  $3N \times 3N$  identity matrix.

### 6.3.3 Additional constraint increasing the spatial-chromatic correlation

As we have said, there exists a strong correlation between the amplitude distributions of the pixels of the three chromatic channels of a RGB image. If we apply our methods independently to each channel the results will be unsatisfactory for not considering this correlation. So we introduce a modification in the sparseness promoting methods, which maintains the methodology and the convergence properties but preserves better the correlation between channels.

To start with, we change the colour space of our images from RGB to YUV. In this new space the correlation between channels is reduced, because it is composed of one luminance and two chrominance channels. As transforming from one colour space to the other and vice versa is a linear operation, this does not affect to the geometrical and convergence properties of the methods, as the composed transformation including the change of colour representation and the redundant transform it is still a tight frame, assuming the latter was already a tight frame.

Chromatic components  $U$  and  $V$  represent differences between colours. They are somehow indicating how correlated they are. We want to keep a high correlation, to avoid classical high frequency colour aliasing artifacts. Thus, the elements of such a representation should be low in amplitude, especially in the high frequencies. Then, we want to promote the smoothness of the  $U$  and  $V$  channels. This is consistent to many colour image representations, which use less information for the chrominance channels than for the luminance. An example is the PAL colour TV system.

The proposed modification is to introduce one more step at each iteration of the algorithm. This new step consists of setting to zero all high-frequency sub-bands of  $U$  and  $V$  wavelet representation. This can be interpreted as an added knowledge to the *a priori* model, so we simultaneously promote sparse images and those having high correlation between chromatic channels in the high spatial frequencies. Note that setting to zero some frequencies is not against promoting sparseness, because the number of significant coefficients is decreased. In addition, this constraint is an orthogonal projection onto a convex set, so the convergence properties of the method are basically unaffected.

A similar idea is described in [148], where a low-pass filter is applied to the frequencies of  $R$  and  $B$  channels. This manages to reduce the colour artifacts of the bilinear interpolation. However, it also smoothes a lot the image, while our method, as it preserves the high frequency details of the luminance, it keeps sharp edges (as shown in subsection 6.3.5).

### 6.3.4 Implementation

**Existing methods.** We have compared to methods having competitive performance with the current state-of-the-art. First method is based on promoting the channel correlation using alternating projections [144]. Second method is an iterative algorithm dealing with representation of colour differences and using an adaptive stopping criterion in this space, with the objective of keeping high correlation between colour channels and removing zipper artifacts [146]. Finally, we also compare with a heuristic method, which is effective to keep channel correlation, although it is computationally inefficient [145]. All these methods are implemented in a MATLAB® package available in the Web page of Professor Xin Li [150].

**$\ell_p$ -GM.** In all the  $\ell_p$ -GM experiments we used  $\alpha = \alpha_0$  and  $\beta = 0.8$ , which provide a good trade-off between computation and quality. We established the stopping threshold for the iterations in 0.01. We used 5 scales for both DT-CWT and Curvelets.

### 6.3.5 Results and discussion

#### 6.3.5.1 $\ell_0$ -GM vs. $\ell_1$ -GM

The results obtained for these application using  $\ell_p$ -AP are worse (and much slower) than those using bilinear interpolation. So we have not included these methods in the comparison.

In our experiments we have also found that  $\ell_0$ -GM provides better results than  $\ell_1$ -GM. Figure 6.12 is an example, where the methods have been applied to the Bayer mosaic with pattern 'GB' built with image 15 of *Eastman Kodak* database [151] (*Lighthouse*). Top-left panel shows a  $64 \times 64$  crop of the original image with a very high-frequency pattern and, thus, particularly difficult to interpolate. The result of top-right panel corresponds to  $\ell_1$ -GM using DT-CWT. PSNR values per channel are: 37.24 dB for R, 39.87 for G, and 37.17 for B. We can still appreciate some colour artifacts. Bottom-left panel is the result of  $\ell_0$ -GM using Curvelets (39.29, 42.27 and 38.20 dB). Finally, bottom-right panel is  $\ell_0$ -GM using DT-CWT (39.59, 41.99 and 39.07 dB). Note that  $\ell_0$ -GM is clearly better than  $\ell_1$ -GM, in terms of removing colour artifacts. Using  $\ell_0$ -GM, PSNR is similar for both DT-CWT and Curvelets, but, on the one hand, this difference is hardly appreciable in these dB levels; and, on the other hand, colour correlation seems better using DT-CWT. In addition, the implementation we used for Curvelets is 4 times slower than the one for DT-CWT, which translates in Curvelets taking around 100 minutes per each full image in [151], whereas only 25 minutes using DT-CWT.

#### 6.3.5.2 $\ell_0$ -GM vs. other methods

We show in Table 6.4 the averaged MSE for each RGB channel obtained by a set of methods and for the images 18, 31, 32, 33, 12, 34, 39, 15, 40, 16, 17 and 19 of [151] (same selection as in [146]). We also averaged the error in  $\nabla E_{ab}^*$  S-CIELab metric [152] and the computation time. We use MSE instead of PSNR because it is a more significant datum in these examples, given the small difference between the methods and the high level of signal-to-noise ratio they achieve. In addition to this, in Tables 6.5 and 6.6 we have detailed these results for every image. Note that, although  $\ell_0$ -GM is slower than the rest, its performance is comparable, being the best or second best in several cases.

However, we want to emphasize that our method is particularly good in high-frequency areas where the de-mosaicing is very complicated. In Figure 6.13 we can see the crop of the results of the methods compared in a specially difficult area of the image 15 of [151] ( $128 \times 128$ ). We can

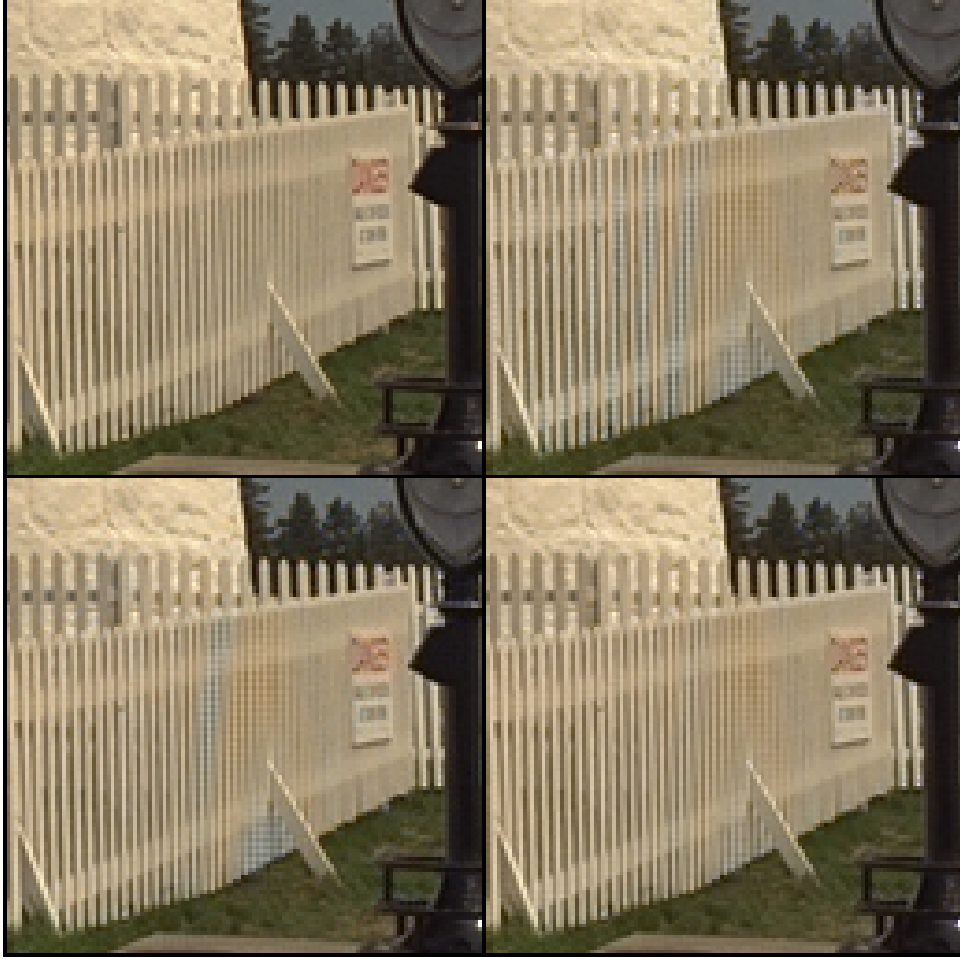


Figure 6.12: Visual example of the comparison between  $\ell_1$ -GM and  $\ell_0$ -GM applied to de-mosaicing. **Top-left**,  $64 \times 64$  detail of image 15 of Eastman Kodak database. **Top-right**, result of de-mosaicing a Bayer mosaic with pattern 'GB' using  $\ell_1$ -GM with DT-CWT. PSNR for channels R, G and B is 37.24 dB, 39.87 dB and 37.17 dB. **Bottom-left**,  $\ell_0$ -GM result using Curvelets (39.29, 42.27 and 38.20 dB). **Top-right**,  $\ell_0$ -GM result using DT-CWT (39.59, 41.99 and 39.07 dB).

appreciate the significant reduction of colour artifacts in our method. We have also reduced the zipper artifacts with respect to [145] and [144]. These artifacts, are due to imposing the observed values, according to the mosaic structure, where the holes are not well-interpolated [146].

### 6.3.6 Conclusions

We have proposed the application of  $\ell_0$ -GM to the problem of interpolating the lost chromatic components after capturing a natural image using a Bayer

Method	MSE			$\nabla E_{ab}^*$ <i>S-CIELab</i>	Time (s.)
	R	G	B		
Lu & Tam [145]	8.67	5.59	11.91	<b>0.78</b>	932.22
Gunturk et al. [144]	7.92	<i>3.61</i>	10.60	0.84	<i>9.40</i>
Li [146]	<b>7.66</b>	<b>3.46</b>	<b>8.61</b>	<i>0.83</i>	<b>2.31</b>
$\ell_0$ -GM	<i>7.68</i>	4.31	<i>9.53</i>	0.93	1238.50

Table 6.4: *MSE and S-CIELab error averaged in the 12 images described in the text, including computation times. Bold numbers indicate the best result for each column, and italic the second best.*

mosaic. We have applied our methods to find sparse approximations to the representation of the image in the YUV colour space. To promote chromatic regularity, we have introduced an extra projection minimising the  $\ell_2$ -norm of the high-frequency components of the U and V channels by setting to zero the high-frequency sub-bands of a redundant linear representation of these channels at each iteration. This does not change the basic geometrical interpretation and convergence properties of the algorithm.

We have seen again that promoting sparseness through a direct local minimisation of the  $\ell_0$ -norm outperforms the optimal minimisation of  $\ell_1$ -norm. We have also seen that  $\ell_0$ -GM is competitive with other methods, although it is very slow. Moreover, we have seen that our method behaves particularly well, clearly better than the others, in difficult high-frequency areas, thus reducing the artifacts significantly. Therefore, we can conclude that it is a promising alternative when the image can be post-processed after capturing them.

The results of  $\ell_0$ -GM still have many zipper artifacts. These appear because we are forcing the observed pixels to not quite well interpolated areas.

## 6.4 Detail increase

### 6.4.1 Introduction

Images often suffer from lost resolution. This could happen, for example, when capturing using photodetectors integrating the incident light; or also when transmitting images through a limited channel.

Detail increase or super-resolution of images consists of the process of obtaining an image or sequence of images with higher-resolution from a set of lower-resolution observations [153]. There are many works approaching

Image	Method	MSE			$\nabla E_{ab}^*$	Time (s.)
		R	G	B	$S\text{-}CIELab$	
18	[145]	3.26	<i>1.51</i>	<b>3.16</b>	<b>0.56</b>	981.77
	[144]	<b>2.69</b>	2.01	4.88	0.62	9.71
	[146]	4.25	1.99	3.49	0.63	1.78
	$\ell_0$ -GM	<i>3.05</i>	<b>1.44</b>	<i>3.24</i>	<i>0.58</i>	1303.50
31	[145]	9.73	6.43	11.23	<b>0.88</b>	980.13
	[144]	<i>7.98</i>	<i>3.46</i>	13.29	0.97	10.09
	[146]	8.15	<b>3.26</b>	<i>9.86</i>	<b>0.88</b>	2.07
	$\ell_0$ -GM	<b>7.21</b>	3.82	<b>8.25</b>	<i>0.93</i>	1368.15
32	[145]	<b>3.06</b>	<b>1.75</b>	<b>4.12</b>	<b>0.51</b>	969.55
	[144]	<i>3.43</i>	<i>2.21</i>	6.59	0.63	8.63
	[146]	3.99	2.49	5.51	0.65	2.39
	$\ell_0$ -GM	4.29	2.23	<i>4.81</i>	<i>0.57</i>	1316.93
33	[145]	20.50	12.17	19.81	<b>1.29</b>	970.12
	[144]	18.28	<i>7.42</i>	23.06	1.40	8.84
	[146]	<i>16.74</i>	<b>6.28</b>	<b>14.48</b>	<i>1.35</i>	3.09
	$\ell_0$ -GM	<b>15.71</b>	8.44	<i>17.88</i>	1.48	1384.49
12	[145]	<i>3.30</i>	2.05	<i>4.06</i>	<b>0.53</b>	971.74
	[144]	<b>3.29</b>	<i>1.79</i>	5.43	<i>0.60</i>	10.21
	[146]	3.57	<b>1.77</b>	<b>3.95</b>	0.58	2.30
	$\ell_0$ -GM	3.40	2.02	4.37	<i>0.60</i>	1428.25
34	[145]	<i>8.27</i>	5.14	<i>7.27</i>	<b>0.73</b>	973.10
	[144]	<b>6.84</b>	<i>3.78</i>	7.61	0.76	9.05
	[146]	8.38	<b>3.29</b>	<b>6.90</b>	<i>0.75</i>	2.06
	$\ell_0$ -GM	9.65	4.39	9.05	0.96	1403.20

Table 6.5: *MSE, S-CIELab and computation time for the 12 images describe in the text and the four methods compared. Bold numbers indicate the best result for each column and method, and italic the second best.*

this problem when there are multiple observations (e.g., video). This is called dynamic super-resolution (e.g., [154, 155, 153]). Here we have focused on the case of having a single observation, called static or single-frame super-resolution case. It is also known as , detail increse, image scaling, interpolation, zooming-in or enlargement.

There are very simple linear methods (such as bilinear, bicubic, etc.), which treat homogeneously every pixel, interpolating their value from a linear combination of the neighbours. Linearity seriously limit the final

Image	Method	MSE			$\nabla E_{ab}^*$	Time (s.)
		R	G	B	$S\text{-}CIELab$	
39	[145]	4.91	3.18	5.50	<i>0.83</i>	976.80
	[144]	4.53	<b>1.67</b>	6.23	0.92	8.73
	[146]	<i>4.34</i>	<i>1.72</i>	<i>4.92</i>	<b>0.82</b>	2.32
	$\ell_0$ -GM	<b>3.30</b>	1.78	<b>3.94</b>	<b>0.82</b>	1385.81
15	[145]	6.80	4.53	8.02	<b>0.74</b>	975.13
	[144]	<i>6.67</i>	<i>2.67</i>	<i>7.71</i>	<b>0.74</b>	9.66
	[146]	<b>5.88</b>	<b>2.51</b>	<b>6.86</b>	<b>0.74</b>	2.24
	$\ell_0$ -GM	7.08	4.12	7.93	<i>0.88</i>	1144.42
40	[145]	6.83	5.12	34.98	<i>0.62</i>	890.25
	[144]	10.57	<b>2.17</b>	<i>7.71</i>	<b>0.61</b>	9.77
	[146]	<b>5.35</b>	<i>2.80</i>	7.80	<i>0.62</i>	2.40
	$\ell_0$ -GM	<i>5.55</i>	3.04	<b>7.44</b>	0.69	1488.72
16	[145]	8.97	6.49	10.04	<i>0.85</i>	832.32
	[144]	<i>7.43</i>	<b>3.26</b>	<b>9.34</b>	<b>0.84</b>	8.29
	[146]	8.08	<i>3.28</i>	<i>9.45</i>	0.86	1.99
	$\ell_0$ -GM	<b>7.22</b>	4.52	10.13	1.10	1495.98
19	[145]	<i>7.72</i>	4.71	<b>8.85</b>	<b>0.85</b>	833.02
	[144]	<b>7.06</b>	<i>4.66</i>	10.95	<i>0.92</i>	9.77
	[146]	8.65	<b>4.52</b>	<i>9.52</i>	0.94	2.10
	$\ell_0$ -GM	9.82	6.01	11.97	1.07	1436.15
19	[145]	20.64	13.96	25.91	<b>0.99</b>	832.75
	[144]	16.33	<i>8.19</i>	<i>24.35</i>	<i>1.05</i>	10.02
	[146]	<b>14.58</b>	<b>7.57</b>	<b>20.59</b>	1.09	3.04
	$\ell_0$ -GM	<i>15.80</i>	9.94	25.38	1.46	1452.55

Table 6.6: Continuation of Table 6.5 for the rest of the 12 images used.

quality. On the one hand there is an excessive blurring, diffusing the edges and avoiding the preservation of details. On the other hand, there are often aliasing artifacts.

These problems have motivated the study of more powerful super-resolution algorithms, based on non-linear techniques. All of them have in common taking advantage of the strong correlation between neighbouring pixels. Some of them use heuristics adapted to the problem (e.g., [156, 157]). Others are based on learning the relationship between observed and original images, as in [158]. However, these methods lack of a mathematical

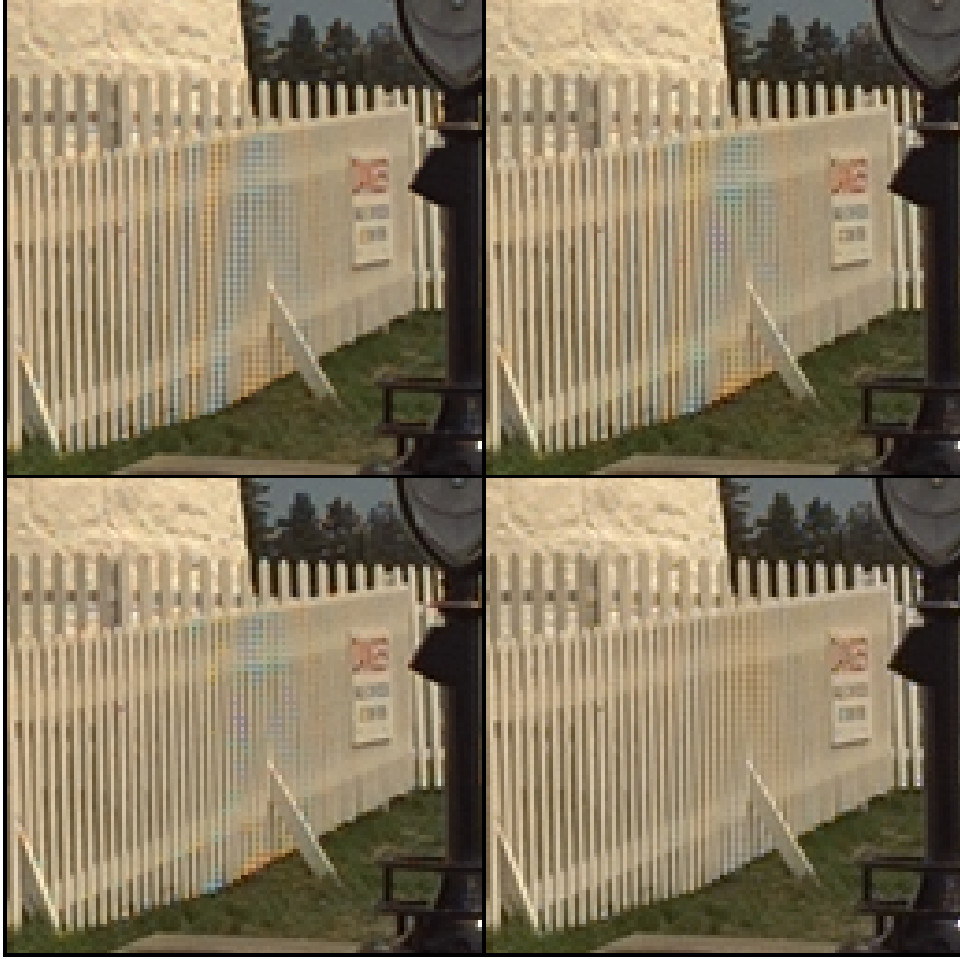


Figure 6.13: Visual comparison between de-mosaicing methods. **Top-left**, result of method in [145] with the image 15 of Eastman Kodak database. We have cropped the image to  $64 \times 64$  to improve the visibility of artifacts. PSNR of R,G and B channels is, respectively, 39.81, 41.57 and 39.09 dB. **Top-right**, [144] result (39.89, 43.87 and 39.26 dB). **Bottom-left**, [146] result (40.39, 44.14 and 39.73 dB). **Bottom-right**,  $\ell_0$ -GM result (39.59, 41.99, 39.07 dB).

model founding their good performance. There are other methods based on setting *a priori* models of the statistics of natural images. For example, [159] establishes a model based on high-kurtosis distributions. Similarly, [160] develops a successful method based on promoting sparseness via minimisation of the  $\ell_1$ -norm. See [161] for recent (2006) and interesting review on non-linear static super-resolution methods.

In this chapter we apply  $\ell_p$ -GM to the static super-resolution problem. We see that both versions,  $p = 0$  and  $p = 1$ , behave similarly both in terms of PSNR and visually, although  $\ell_0$ -GM obtains sparser solutions. We



have also compared, for reference, to the nearest neighbour and bilinear interpolation. Our objective in this section is not to propose an alternative to current static super-resolution methods but to analyse the potential of the models and methods used in this Thesis regarding this kind of applications.

### 6.4.2 Consistency set

We name  $L$  the number of pixels averaged to obtain every observed pixel. We note the observed image  $\mathbf{y} \in \mathbb{R}^{N/L}$ . We define a family of indices sets,  $J_i$ , with  $i = \{1, \dots, N/L\}$ , and corresponding to the non-overlapping  $\sqrt{L} \times \sqrt{L}$  block formed by all  $j \in \{1, \dots, N\}$  whose corresponding pixels in the original image,  $\mathbf{x}_0 \in \mathbb{R}^N$ , have been averaged to provide the observed value  $y_i$ . We also define  $\mathbf{x}_0^{J_i}$  as the block formed by pixels of  $\mathbf{x}_0 \in \mathbb{R}^N$  in the positions indicated by  $J_i$ . The consistency set,  $R_a(\mathbf{y})$ , is then formed by those images  $\mathbf{x} \in \mathbb{R}^N$  whose associated blocks,  $\mathbf{x}^{J_i}$ , preserve the observed average. Then we have that:

$$R_a(\mathbf{y}) = \{\mathbf{x} \in \mathbb{R}^N : \langle \mathbf{x}^{J_i} \rangle = y_i, \forall i \in \{1, \dots, N/L\}\},$$

where  $\langle \mathbf{a} \rangle$  denotes the average value of the coefficients of vector  $\mathbf{a}$ . Geometrically, it is easy to check that the projection onto the affine subspace of vectors preserving a given averaged consists of subtracting a constant vector in amplitude, which is the difference between the mean of the vector to be projected and the target mean. Thus, given  $\mathbf{x} \in \mathbb{R}^N$ , we have that  $P_{R_a(\mathbf{y})}(\mathbf{x}) = \mathbf{z}$ , where, for all  $i \in \{1, \dots, N/L\}$  and  $j \in \{1, \dots, N\}$ :

$$z_j^{J_i} = x_j^{J_i} - (\langle \mathbf{x}^{J_i} \rangle - y_i).$$

### 6.4.3 Implementation

$\ell_p$ -GM. We have experienced that, for application to static super-resolution, best results of  $\ell_p$ -GM, for both  $p = 0$  and  $p = 1$ , are obtained when using DT-CWT with only 3 scales. The parameters used are  $\alpha = \alpha_0$  and  $\beta = 0.8$ . Iterations end when the threshold is below 0.01.

### 6.4.4 Results and discussion

Table 6.7 compares methods  $\ell_0$ -GM,  $\ell_1$ -GM, nearest neighbours and bilinear filtering. It reflects the averaged PSNR of the estimation in the images of our test set. To begin with, it is interesting to note the good behaviour of the simplest possible method, nearest neighbours. This is because we are imposing a local mean, what is, in average, the best possible linear strategy

for this degradation. On the other hand, we see that  $\ell_p$ -GM, in both cases  $p = 0$  and  $p = 1$ , behaves quite well. It is curious to observe that, in contrast to the rest of studied degradations, both cases provide very similar results in PSNR terms. This does not mean that the results are strictly similar. The  $\ell_0$ -GM result provides a significantly sparser distribution of coefficients ( $\approx 2.15 \cdot 10^5$  for  $\ell_0$ -GM and  $\approx 2.45 \cdot 10^5$  for  $\ell_1$ -GM, on average). With these results at hand, we conclude that the relative performance of  $\ell_0$ -GM has decreased with respect to previous applications, probably because the function performance vs. achieved sparseness has a maximum (this hypothesis is consistent with the bad results provided by  $\ell_0$ -GM in de-quantizing). In fact, if we use  $\beta$  values closer to 1, the sparseness is further increased but the estimation error is also increased.

Figure 6.14 shows an example, using (*House*) of visual results of the methods. Both  $\ell_1$ -GM and  $\ell_0$ -GM have a sharper visual aspect, and better behaviour in the edges than the linear strategies compared. The aliasing is a significantly reduced (note, for example, the edges of the roof). Finally, we see that, in this case, there is no significant difference, despite the disparity of sparseness, between  $\ell_1$ -GM and  $\ell_0$ -GM.

Method	PSNR (dB)				
	<i>Barbara</i>	<i>Boat</i>	<i>House</i>	<i>Lena</i>	<i>Peppers</i>
Nearest-neighbour	27.09	26.54	30.52	27.68	25.74
Bilinear interpolation	24.16	24.25	25.84	25.41	23.76
$\ell_1$ -GM	<b>27.49</b>	<b>28.84</b>	<i>33.50</i>	<i>30.66</i>	<i>28.06</i>
$\ell_0$ -GM	<i>27.14</i>	<i>28.80</i>	<b>33.53</b>	<b>30.75</b>	<b>28.07</b>

Table 6.7: PSNR (using averaged MSE) obtained in the super-resolution of the methods to recover original size of the images of our test set, when they are averaged in non-overlapping  $2 \times 2$  blocks. Bold numbers indicate the best result for each image, and italic the second best.

### 6.4.5 Conclusions

In this section we have explored the application of  $\ell_0$ -GM and  $\ell_1$ -GM to static super-resolution. Both of them clearly outperform linear methods compared (nearest neighbours and bilinear interpolation). Nevertheless, we have seen that the sparser solution, provided by  $\ell_0$ -GM, is not better than that of  $\ell_1$ -GM. Moreover, for both methods the increase in the performance is inverted when going beyond the sparseness level shown in the experiments. These are preliminary results that need to be improved in the future,

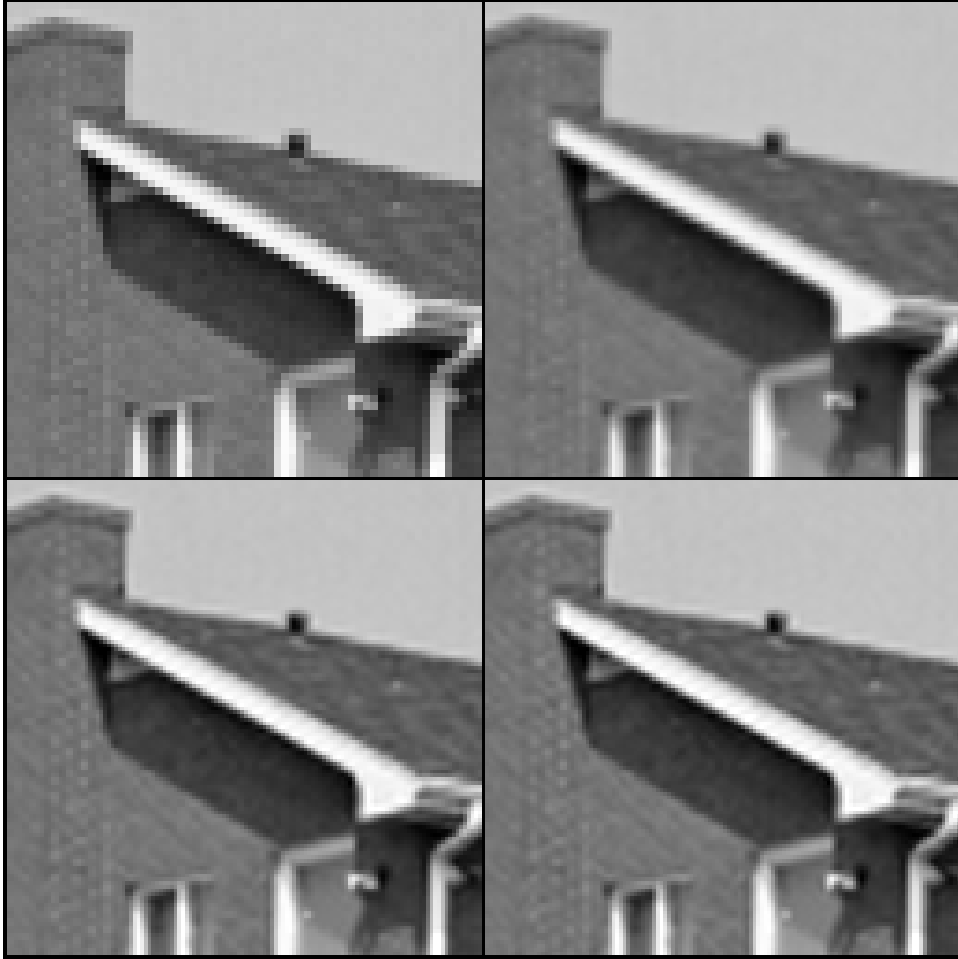


Figure 6.14: **Top-left**, nearest neighbour interpolation (replication) of the resizing of the sub-sampling of the  $2 \times 2$  averaged blocks House image (PSNR: 30.52 dB). We have cropped to  $128 \times 128$  to improve visibility. **Top-right**, result of bilinear interpolation (29.65 dB). **Bottom-left**, result of  $\ell_1$ -GM (33.50 dB). **Bottom-right**, result of  $\ell_0$ -GM (33.53 dB).

through a better understanding of the underlying sparseness promoting mechanisms for this type of applications.



# Chapter 7

## Conclusions and future work

### 7.1 Conclusions

The main conclusion that we obtain from this Thesis is that, although the global optimisation posed by the sparse approximation problem is NP-complex, it is not only possible but relatively simple to find equivalent formulations that allow, using tight frames, the application of well-known optimisation techniques to solve, at least locally, the problem. Although the proposed methods have been recently proposed and used as heuristics, until this work, up to our knowledge, nobody had established an appropriate theoretical framework to obtain them as solutions to classical optimisation problems. This is against the common belief, quite extended through the scientific community in this field, that it is only possible to theoretically derive these kind of methods by using convex approximations to the cost function.

The main objective of this work has not been to obtain the precise theoretical conditions under which the methods find the global optimum. Instead, we have preferred to study, in an intensive way, their practical application in real image processing conditions, both in terms of energy compaction and application to restoration problems.

We have derived two methods to solve the sparse approximation problem. First one is formulated from the equivalent problem of minimising the MSE to the represented image from a vector of synthesis coefficients inside a  $\ell_p$ -ball of given radius. It results in a method based on alternated orthogonal projections, using two sets: 1) the set of vectors that reconstruct perfectly the image; and 2) the  $\ell_p$ -ball of given radius. We have called this method  $\ell_p$ -AP. We have focused on the cases  $p = 0$ , where sub-optimal solutions are achieved, and  $p = 1$ , where we find the global minimum to the convex relaxation problem. In the experiments, we have shown that

$\ell_0$ -AP outperforms  $\ell_1$ -AP under the applied conditions (natural images and useful sparseness levels in typical redundant dictionaries). It also outperforms other iterative techniques based on fixed threshold and also greedy strategies. However, we have seen that the choice of elementary functions is better in  $\ell_1$ -AP than in  $\ell_0$ -AP. In addition, the need of  $\ell_p$ -AP for choosing an appropriate value for a sparseness parameter decreases its applicability.

To overcome these problems, we have derived another method. Firstly, we have reformulated our optimisation problem by finding a continuous and constrained function which is equivalent to the cost function of the sparse approximation problem. Then, we have derived a generalised version of IHT using gradient descent on this new function. We have proved that the convergence point of this method is a local minimum to the sparse approximation problem. Next, we have proposed  $\ell_0$ -GM method, by rewriting the new function as an infinitely sharp cost function convolved with a smoothing kernel, which has allowed us to use a deterministic annealing approach to justify the use of decreasing threshold. We have also derived a similar version of the method using the  $\ell_1$ -norm instead,  $\ell_1$ -GM, whose use is recommended when our estimation is required to have low sparseness levels.

We have experienced that  $\ell_0$ -GM outperforms all compared methods in providing sparse approximation solutions, including  $\ell_0$ -AP and the optimisation of the support given by  $\ell_1$ -AP. Then, we can conclude that methods based on minimising the  $\ell_0$ -norm are better than those based on  $\ell_1$  in the conditions of this study.

It is important to note that the number of coefficients required by  $\ell_0$ -GM to reconstruct perfectly the image tends to the theoretical asymptote as the annealing gets slower. This means a quasi-optimal asymptotic performance in the fidelity-sparseness curve. In addition, solutions with low sparseness approximate quite well the optimal solution for other values when thresholding *a posteriori*. This is very important in practice, because, in this way, we avoid searching for the optimal sparseness level in a first stage, thus simplifying the final implementation of the method and also increasing the efficiency of adjusting the sparseness level afterwards.

Regarding the application to restoration problem, we have seen that is generally easy to adapt these methods to restore images affected by strictly reproducible degradations. We have studied two different kinds of sparseness that can be used as *a priori* model for this problems. On the one hand, the synthesis-based sparseness (SS) assumes that natural images can be expressed as a linear combination of a few elementary functions of a given dictionary. This approximation, though completely valid, does not

have a solid empirical justification. We have proposed instead the use of *a priori* models based on the sparseness of the analysis coefficients (SA), which are based on the observation that the energy of the linear response of redundant wavelet-like dictionaries to natural images is concentrated in a small proportion of coefficients. This allows having a fully justified empirical basis. Adapting our methods to this kind of sparseness does not require to change their conceptual framework.

We have proposed to use  $\ell_0$ -AP, with SS, for removing spatial quantisation artifacts and  $\ell_0$ -GM, with SA, to several interpolation-based problems, as the recovery of missing pixels, the construction of colour images from mosaics and the increase of the detail of the images. We have seen that results based on minimisation of the  $\ell_1$ -norm are worse in most of the studied applications. In addition to this, we have seen that our methods have a similar or superior performance with respect to other existing ones. To the best of our knowledge, nobody has explicitly applied before sparse-promoting techniques to de-quantizing.

Despite good performance, these results made us also realise of some weaknesses of the model. For example, we have seen that  $\ell_0$ -GM provides worse results than  $\ell_0$ -AP when removing spatial quantisation artifacts, being sparser. Moreover,  $\ell_0$ -GM has problems to interpolate regular grids, because it tends to represent the artifacts. We have used heuristics to force the method to get out of these non-favourable local minima. However, we have experienced that this problem is not only due to the method, but also to the too simplistic sparseness model used here.

## 7.2 Future work

We believe that these Thesis opens several interesting venues to explore in the future. Firstly, the marginal statistics of the analysis coefficients can be better approximated by using intermediate norms ( $0 < p < 1$ ) [74, 72]. In this sense, we believe that it is possible, though not trivial, to derive an analogous method to  $\ell_0$ -GM and  $\ell_1$ -GM based on them. For restoration purposes, the use of intermediate norms is more justified as *a priori* model for the analysis coefficients (see, e.g., [29]).

Secondly, we would like to study the use of *a priori* knowledge for synthesis coefficients in a more justified basis. On the one hand, using intermediate norms could be justified as a better compromise between the good behaviour of  $\ell_0$  and the ability of  $\ell_1$  to avoid local minima. On the other hand, we are working on a fully justified Expectation-Maximization-like method based on maximising the likelihood of a model for the synthesis coefficients.

With a more practical orientation, but also with important theoretical impact, we want to explore the problems of the method when dealing with regular grid interpolation. We believe that these problems are caused by the extreme simplicity of the model.

Finally, it would be interesting to study the application of the proposed methods to classical image restoration problems, as additive noise and blurring. This could be approached through a statistical formulation of the sparse approximation problem by searching the *Maximum A Posteriori* (MAP) solution to the restoration problem (as seen, for example, in [72]).



# Appendices



# Appendix A

## Test images set

Most of the experiments in this Thesis has been performed over a standard test set of images of size  $256 \times 256$ , composed by *House*, *Boat*, *Barbara*, *Peppers* and *Lena*. *Boat* and *Barbara* has been cropped from their original size starting by row 200 and column 100 in *Boat*; and by row 150 and columns 50 in *Barbara*. In Figure A.1 we show this set.



Figure A.1: *Test set of images used in this Thesis.*

## Appendix B

### Minimisation of the quadratic error of the reconstruction with a given support.

We prove next that Equation (3.11) solves the pseudo-inverse involved in the problem of minimising the reconstruction error of an image as linear combination of a given set of elementary functions from a redundant dictionary. Given an image  $\mathbf{x} \in \mathbb{R}^N$ , a subset  $I$  of  $R$  indices extracted from the set  $\{1, \dots, M\}$ , and a  $N \times R$  matrix  $\Phi_I$  formed by columns  $\phi_i$  from  $\Phi$ , we want to solve for  $\mathbf{a}_I$ :

$$\hat{\mathbf{a}}_I = \arg \min_{\mathbf{a}_I} \|\Phi_I \mathbf{a}_I - \mathbf{x}\|_2, \quad (\text{B.1})$$

which can be expressed as:

$$\hat{\mathbf{a}}_I = \Phi_I^\# \mathbf{x},$$

where  $\Phi_I^\#$  is the pseudo-inverse of  $\Phi_I$ . We study two possibilities: 1)  $\text{range}(\Phi_I) = R \leq N$ , and 2)  $R > \text{range}(\Phi_I) = N$ .

#### B.1 First case: incomplete subset

When  $\text{range}(\Phi_I) = R \leq N$ , then:

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

The involved matrix inversion is potentially huge. Fortunately, we can use the Taylor expansion of the inverse of a matrix [162], and write:

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

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

$$\mathbf{a}_I^{(k+1)} = \mathbf{a}_I^{(k)} + \Phi_I^T(\mathbf{x} - \Phi_I \mathbf{a}_I^{(k)}). \quad (\text{B.2})$$

Now we define  $\mathbf{S}_I$  as the  $R \times N$  matrix selecting the  $R$  coefficients indicated by set  $I$ . Then,  $\mathbf{S}_I^T$  is the operator expanding a  $R \times 1$  vector into a  $N \times 1$  one by placing each coefficient in its original position and setting the rest to zero. Then,  $\Phi_I = \Phi \mathbf{S}_I^T$  and  $\Phi_I^T = \mathbf{S}_I \Phi^T$ , and substituting in Equation (B.2) we get:

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

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

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

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

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

Let  $\mathbf{D}_I$  be a  $M \times M$  diagonal 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 see that:

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

As the right term only depends on  $\mathbf{D}_I \mathbf{a}^{(k)}$ , and, by construction (following Equation (3.11)),  $\mathbf{a}^{(k)} = \mathbf{D}_I \mathbf{a}^{(k)}$ , for all  $k \geq 0$ , then these iterations are completely equivalent to those of Equation (3.11), being  $\mathbf{a}^{(k)}$  the intermediate result of those iterations.

## B.2 Second case: complete subset

When  $R > \text{range}(\Phi_I) = N$ , then Equation (B.1) has infinite solutions with perfect reconstruction of  $\mathbf{x}$ . The pseudo-inverse provides, among them, the one with minimum Euclidean norm:

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

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

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

that can be calculated using the following iterations:

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

Multiplying by  $\Phi_I^T$ :

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

and substituting  $\Phi_I^T \mathbf{z}_I^{(k)}$  by  $\mathbf{a}_I^{(k)}$  we obtain the Equation (B.2), so the solution is reached using the same iterative method than in the previous case.





## Appendix C

### A Parseval frame formed concatenating two Parseval frames

Let  $\Phi_A$  be a  $N \times M$  matrix with  $M > N$  and  $\Phi_B$  a  $N \times L$  matrix with  $L > N$ . Assume that both matrices are Parseval frames. Then, if we form a new  $N \times (M + L)$  matrix as the union of the columns of  $\Phi_A$  and  $\Phi_B$ , this new matrix is no longer a Parseval frame, because the linear transformation doubles the energy of the original vector. Then, we should modulate each primitive matrix with a scale factor. In order to preserve the energy of the joint transformation, the sum of the square power of these factors should add 1. These two factors act as weight of the relative importance of each matrix in the new frame.

Formally, we define the  $N \times (M + L)$  matrix  $\Phi$  as formed by the union of the columns of  $\Phi_A$  and  $\Phi_B$ , respectively multiplied by two scale factors,  $\sqrt{\gamma_A}$  and  $\sqrt{\gamma_B}$ , such that  $\gamma_A + \gamma_B = 1$ .



# Appendix D

## Publications

This is a list of publications derived from this work:

- L. Mancera, J.A. Guerrero-Coln, J. Portilla. Sparse Approximation via Orthogonal Projections: Beyond Greed and Convexity. *IEEE Transactions on Image Processing* (submitted)
- J. Portilla, L. Mancera. L0-based sparse approximation: Two alternative methods and some applications. *SPIE Optics & Photonics*, edited by SPIE, San Diego (CA), August 2007.
- L. Mancera, J. Portilla. L0-norm-based Sparse Representation through Alternate Projections. *13th International Conference on Image Processing (ICIP'06)*, Atlanta, GE (USA), October 2006.
- L. Mancera, J. Portilla. Image De-Quantizing via Enforcing Sparseness in Overcomplete Representations. *Lecture Notes in Computer Sciences*, vol. **3708**, pp. 411-418. Also in *7th International Conference on Advanced Concepts for Intelligent Vision Systems (ACIVS 2005)*, Antwerp (Belgium), September 2005. (**Present in 2005 JCR Science Edition**)



# Bibliography

- [1] E.P. Simoncelli and B. Olshausen, “Natural image statistics and neural representation,” *Annual Review of Neuroscience*, vol. 24, pp. 1193–1216, May 2001.
- [2] D.L. Ruderman, “The statistics of natural images,” *Network: Computational Neural Systems*, vol. 5, no. 4, pp. 517–548, November 1994.
- [3] S.G. Mallat, “A theory for multiresolution signal decomposition: The wavelet representation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674–693, July 1989.
- [4] H. Barlow, “Redundancy reduction revisited,” *Network: Computational Neural Systems*, vol. 12, no. 3, pp. 241–253, May 2001.
- [5] B.A. Olshausen and D.J. Field, “Sparse coding of sensory inputs,” *Current Opinion in Neurobiology*, vol. 14, pp. 481–487, July 2004.
- [6] B.A. Olshausen and D.J. Field, “Emergence of simple-cell receptive fields properties by learning a sparse code for natural images,” *Nature*, vol. 381, no. 6583, pp. 607–609, June 1996.
- [7] E.P. Simoncelli, W.T. Freeman, E.H. Adelson, and D.J. Heeger, “Shiftable multi-scale transforms,” *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 587–607, March 1992.
- [8] D.L. Donoho, “De-noising by soft thresholding,” *IEEE Transactions on Information Theory*, vol. 41, no. 3, pp. 613–627, May 1995.
- [9] E.J. Candes, M. Rudelson, T. Tao, and R. Vershynin, “Error correction in linear programming,” in *Proceedings 46th Symposium on Foundations of Computer Science*, Pittsburgh, PA, 22–25 October 2005, IEEE Signal Processing Society.

- [10] J. Portilla and L. Mancera, "L0-based sparse approximation: Two alternative methods and some applications," in *SPIE Optics and Photonics*, San Diego, CA, 26-30 August 2007, SPIE.
- [11] S.S. Chen, "Basis Pursuit," *Ph.D. Thesis, Stanford University*, 1995.
- [12] S. Mallat and Z. Zhang, "Matching Pursuit in time-frequency dictionary," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, December 1993.
- [13] T.H. Reeves and N.G. Kingsbury, "Overcomplete image coding using iterative projection-based noise shaping," in *Proceedings of the 9th IEEE International Conference on Image Processing*, Rochester, NY, 23-25 September 2002, vol. 3, pp. 597–600, IEEE Signal Processing Society.
- [14] M.A.T. Figueiredo and R.D. Nowak, "An EM algorithm for wavelet-based image restoration," *IEEE Transactions on Image Processing*, vol. 12, no. 8, pp. 906–916, August 2003.
- [15] N. Kingsbury and T. Reeves, "Redundant representation with complex wavelets: How to achieve sparsity," in *Proceedings of the 10th IEEE International Conference on Image Processing*, Barcelona, Spain, 14-18 September 2003, vol. 1, pp. 45–48, IEEE Signal Processing Society.
- [16] L. Mancera and J. Portilla, "L0-norm-based representation through alternate projections," in *Proceedings 13th IEEE International Conference on Image Processing*, Atlanta, GE, 8-11 October 2006, pp. 2089–2092, IEEE Signal Processing Society.
- [17] I. Daubechies, G. Teschke, and L. Vese, "Iteratively solving linear inverse problems under general convex constraints," *Inverse Problems and Imaging*, vol. 1, no. 1, pp. 29–46, 2007.
- [18] J. Portilla, V. Strela, M. Wainwright, and E.P. Simoncelli, "Image denoising using a Scale Mixture of Gaussians in the wavelet domain," *IEEE Transactions on Image Processing*, vol. 12, no. 11, pp. 1338–1351, November 2003.
- [19] J.L. Starck, "Morphological Component Analysis," in *Proceedings of the SPIE*, San Diego, CA, August 2005, vol. 5914.

- [20] M.J. Fadili and J.L. Starck, “EM algorithm for sparse representation-based image inpainting,” in *Proceedings 12th IEEE International Conference on Image Processing*, Genoa, Italy, 11-14 September 2005, IEEE Signal Processing Society.
- [21] M. Elad and M. Aharon, “Image denoising via sparse and redundant representations over learned dictionaries,” *IEEE Transactions on Image Processing*, vol. 15, no. 12, pp. 3736–3745, December 2006.
- [22] Y. C. Pati, R. Rezaiifar, and P.S. Krishnaprasad, “Orthogonal Matching Pursuit: Recursive function approximation with application to wavelet decomposition,” in *Proceedings 27th Asilomar Conference in Signals, Systems and Computers*, 1-3 November 1993.
- [23] D.L. Donoho, Y. Tsaig, I. Drori, and J.L. Starck, “Sparse solution of undetermined linear equations by Stagewise Orthogonal Matching Pursuit,” *Technical Report*, April 2006.
- [24] L. Mancera and J. Portilla, “Image dequantizing via enforcing sparseness in overcomplete representations,” in *7th International Conference on Advanced Concepts in Intelligent Vision Systems*, Antwerp, Belgium, 20-23 September 2005, vol. LNCS-3708, pp. 411–418, Springer Verlag.
- [25] N.N. Abdelmalek, “An efficient method for the discrete linear L1 approximation problem,” *Mathematical Computation*, vol. 29, no. 131, pp. 844–8502, July 1975.
- [26] S.S. Chen, D.L. Donoho, and M.A. Saunders, “Atomic decomposition by Basis Pursuit,” *SIAM Journal on Signal Processing*, vol. 20, no. 1, pp. 33–61, 1999.
- [27] I.F. Gorodnitsky and B.D. Rao, “Sparse signal reconstruction from limited data using focuss: A re-weighted minimum norm algorithm,” *IEEE Transactions on Signal Processing*, vol. 45, no. 3, pp. 600–616, March 1997.
- [28] I. Daubechies, M. De Fries, and C. De Mol, “An iterative thresholding algorithm for linear inverse problems with a sparsity constraint,” *Communications on Pure and Applied Maths*, vol. 57, pp. 1413–1457, 2004.
- [29] M.A.T. Figueiredo and R.D. Nowak, “A Bound Optimization Approach to wavelet-based image deconvolution,” in *Proceedings 12th*

- IEEE International Conference on Image Processing*, Genoa, Italy, 11-14 September 2005, vol. 2, pp. 782–785, IEEE Signal Processing Society.
- [30] D.L. Donoho, “For most large undetermined systems of linear equations the minimal L1-norm solution is also the sparsest solution,” *Technical Report*, 2004.
- [31] J.A. Tropp, “Greed is good: Algorithmic results for sparse approximation,” *IEEE Transactions on Information Theory*, vol. 50, no. 10, pp. 2231–2242, October 2004.
- [32] D.L. Donoho, M. Elad, and V.N. Temlyakov, “Stable recovery of sparse overcomplete representations in the presence of noise,” *IEEE Transactions on Information Theory*, vol. 52, no. 1, pp. 6–18, January 2006.
- [33] A.C. Gilbert, M.J. Strauss, J.A. Tropp, and R. Vershynin, “Algorithmic linear dimension reduction in the L1 norm for sparse vectors,” *arXiv:math/0608079*, August 2006.
- [34] O.G. Guleryuz, “On missing data prediction using sparse signal models: A comparison of atomic decomposition with iterated denoising,” in *Proceedings of the SPIE*, San Diego, CA, August 2005, vol. 5914.
- [35] D.C. Youla, “Generalized image restoration by the method of alternating orthogonal projections,” *IEEE Transactions on Circuits and Systems*, vol. CAS-25, no. 9, pp. 694–702, September 1978.
- [36] P. Combettes, “The foundation of set theoretic estimation,” *Proceedings of the IEEE*, vol. 81, no. 2, pp. 182–208, February 1993.
- [37] P. Combettes and V. Wajs, “Signal recovery by proximal forward-backward splitting,” *SIAM Journal on Multiscale Modeling and Simulation*, vol. 4, pp. 1168–1200, 2005.
- [38] I. Daubechies, M. Fornasier, and I. Loris, “Accelerated projected gradient method for linear inverse problems with sparsity constraints,” *Submitted to arXiv:0706.4297v1*, June 2007.
- [39] K.K. Herrity, A.C. Gilbert, and J.A. Tropp, “Sparse approximation via iterative thresholding,” in *Proceedings IEEE International Conference on Acoustics, Speech and Signal Processing*, 14-19 May 2006.



- [40] R.T. Rockafellar, "Monotone operators and the proximal point algorithm," *SIAM Journal of Control and Optimization*, vol. 14, pp. 877–898, 1976.
- [41] R.R. Coifman and M.V. Wickerhauser, "Entropy-based algorithms for best-basis selection," *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 713–718, March 1992.
- [42] I. Daubechies, *Ten Lectures on Wavelets*, Cambridge University Press, 1992.
- [43] N. Kingsbury, "Complex wavelets for shift invariant analysis and filtering of signals," *Journal of Applied and Computational Harmonic Analysis*, vol. 10, no. 3, pp. 234–253, May 2001.
- [44] A. Cohen, I. Daubechies, O.G. Guleryuz, and M.T. Orchard, "Of the importance of combining wavelet-based non-linear approximation with coding strategies," *IEEE Transactions on Information Theory*, vol. 48, no. 7, pp. 1895–1921, July 2001.
- [45] D.L. Donoho J.L. Starck, M. Elad, "Redundant multiscale transforms and their application for Morphological Component Analysis," *Journal of Advances in Imaging and Electron Physics*, vol. 123, pp. 287–384, 2004.
- [46] D.L. Donoho and M. Elad, "On the stability of Basis Pursuit in the presence of noise," *Signal Processing*, vol. 86, pp. 511–532, 2006.
- [47] S. Weisberg, *Applied Linear Regression*, Wiley, New York, 1980.
- [48] C. Daniel and F.S. Wood, *Fitting Equations to Data: Computer Analysis of Multifactor Data*, Wiley, New York, 1980.
- [49] T. Hastie, R. Tibshirani, and J.H. Friedman, *Elements of Statistical Learning*, Springer-Verlag, New York, 2001.
- [50] R.A. DeVore and V.N. Temlyakov, "Some remarks on greedy algorithms," *Advanced Computational Mathematics*, vol. 12, pp. 213–227, 1996.
- [51] V.N. Temlyakov, "Greedy algorithms and m-term approximation," *Journal of Approximation Theory*, vol. 98, pp. 117–145, 1999.
- [52] V.N. Temlyakov, "Weak greedy algorithms," *Advances in Computational Mathematics*, vol. 5, pp. 173–187, 2000.

- [53] R. Gribonval and P. Vandergheynst, "On the exponential convergence of Matching Pursuits in quasi-incoherent dictionaries," *IEEE Transactions on Information Theory*, vol. 52, no. 1, pp. 255–261, January 2006.
- [54] V.N. Temlyakov, "Nonlinear methods of approximation," *Foundations of Computational Mathematics*, vol. 3, pp. 33–107, 2003.
- [55] M. Andrieu, L. Rebollo-Neyra, and E. Sagianos, "Backward-Optimized Orthogonal Matching Pursuit approach," *IEEE Signal Processing Letters*, vol. 11, no. 9, pp. 705–708, September 2004.
- [56] G.Z. Zarabulut, L. Moura, D. Panario, and A. Yongacoglu, "Flexible Tree-Search based Orthogonal Matching Pursuit algorithm," in *30th IEEE International Conference on Acoustic, Speech, and Signal Processing*, Philadelphia, PA, 18-23 March 2005, pp. 673–676, IEEE Signal Processing Society.
- [57] C. La and M.N. Do, "Tree-Based Orthogonal Matching Pursuit algorithm for signal reconstruction," in *Proceedings 13th IEEE International Conference on Image Processing*, Atlanta, GE, 8-11 October 2006, IEEE Signal Processing Society.
- [58] P. Jost, P. Vandergheynst, and P. Frossard, "Tree-based pursuit: Algorithm and properties," *IEEE Transactions on Signal Processing*, vol. 54, no. 12, pp. 4685.
- [59] S.F. Cotter and B.D. Rao, "Application of tree-based searches to Matching Pursuit," in *Proceedings of the IEEE International Conference on Acoustic, Speech and Signal Processing*, Salt Lake City, UT, 7-11 May 2001, vol. 6, pp. 3933–3936.
- [60] B. Efron, T. Hastie, I. Jonhstone, and R. Tibshirani, "Least Angle Regression," *Annual Statistics*, vol. 32, no. 2, pp. 407–499, 2004.
- [61] J. Mairal, M. Elad, and G. Sapiro, "Sparse representation for color image restoration," *IEEE Transactions on Image Processing*, vol. 18, no. 1, pp. 53–69, January 2007.
- [62] C. Distanto, A. Leone, L. My, M. Rizzello, and P. Siciliano, "A video compression algorithm based on Matching Pursuit integrated into a wireless embedded sensor node compliant with IEEE 1451.1 standard architecture," in *Proceedings of the 2nd IASTED International Conference*, Cambridge, MA, 8-10 November 2004.

- [63] A. Rahmoune, P. Vandergheynst, and P. Frossard, "MP3D: Highly scalable video coding scheme based on Matching Pursuit," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Montreal, Canada, 17-21 May 2004, vol. 3, pp. 133–136.
- [64] J.L. Lin, W.L. Hwang, and S.C. Pei, "Video compression based on Orthonormal Matching Pursuits," in *Proceedings International Symposium on Circuits and Systems*, pp. 4–8.
- [65] R. Figueras i Ventura, P. Vandergheynst, P. Frossard, and A. Cavallaro, "Color image scalable coding with Matching Pursuit," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Montreal, Canada, 17-21 May 2004, vol. 3, pp. 53–56.
- [66] R. Gribonval, "Sparse decomposition of stereo signals with Matching Pursuit and application to blind separation of more than two sources from a stereo mixture," in *Proceedings of International Conference on Acoustic, Speech and Signal Processing*, Orlando, FL, 13-17 May 2002.
- [67] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society B.*, vol. 58, no. 1, pp. 267–288, 1996.
- [68] J.A. Tropp, "Just relax: Convex programming methods for identifying sparse signals in noise," *IEEE Transactions on Information Theory*, vol. 52, no. 3, pp. 1030–1051, March 2006.
- [69] I.W. Selesnick, R.V. Slyke, and O.G. Guleryuz, "Pixel recovery via L1 minimization in the wavelet domain," in *Proceedings of the 11th IEEE International Conference on Image Processing*, Singapore, 24-27 October 2004, IEEE Signal Processing Society.
- [70] M.F. Duarte, M.B. Wakin, and R.G. Baraniuk, "Fast reconstruction of piecewise smooth signals from random projections," in *Online Proceedings of Workshop on Signal Proc. with Adaptive Sparse Structured Representations (SPARS)*, 2005.
- [71] D.L. Donoho, "Compressed Sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, April 2006.

- [72] M.A.T. Figueiredo, R.D. Nowak, and S.J. Wright, "Gradient projection for sparse reconstruction: Application to Compressed Sensing and other inverse problems," *IEEE Journal of Selected Topics in Signal Processing: Special Issue on Convex Optimization Methods for Signal Processing*, vol. 1, no. 4, 2007.
- [73] R.D. Nowak and M.A.T. Figueiredo, "Overcomplete image coding using iterative projection-based noise shaping," in *Proceedings of the 35th Asilomar Conference on Signals, Systems and Computers*, Monterrey, CA, 2001.
- [74] E.P. Simoncelli and E.H. Adelson, "Noise removal via Bayesian wavelet coring," in *Proceedings of the 3rd IEEE International Conference on Image Processing*, Lausanne, Switzerland, 16-19 September 1996, vol. 1, pp. 379–382, IEEE Signal Processing Society.
- [75] A. Perez and R.C. Gonzalez, "An iterative thresholding algorithm for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 9, no. 6, pp. 742–751, November 1987.
- [76] M. Elad, "Why simple shrinkage is still relevant for redundant representations?," *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5559–5469, December 2006.
- [77] A. Chambolle, R.A. DeVore, N.Y. Lee, and B.J. Lucier, "Nonlinear wavelet image processing: Variational problems, compression and noise removal through wavelet shrinkage," *IEEE Transactions on Image Processing*, vol. 7, no. 3, pp. 319–335, March 1998.
- [78] P. Moulin and J. Liu, "Analysis of multiresolution image de-noising schemes using generalized-Gaussian and complexity priors," *IEEE Transactions on Information Theory, Special Issue on Multiscale Analysis*, vol. 45, no. 3, pp. 909–919, April 1999.
- [79] D.L. Donoho and I.M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 851, no. 3, pp. 425–455, 1994.
- [80] T. Adeyemi and M.E. Davies, "Sparse representation of images using overcomplete complex wavelets," in *IEEE SP 13th Workshop on Statistical Signal Processing*, Bordeaux, France, 17-20 July 2006, pp. 805–809.
- [81] M. Elad, B. Matalon, J. Shtok, and M. Zibulevsky, "A wide-angle view at iterated shrinkage algorithm," in *SPIE Optics and Photonics*, San Diego, CA, 26-30 August 2007, SPIE.

- [82] M. Elad, “Shrinkage for redundant representations,” in *Workshop of Signal Processing with Adaptive Sparse Structured Representations (Spars05)*, 2005.
- [83] L.A. Karlovitz, “Construction of nearest points in the  $L_p$ ,  $p$  even and  $L_1$  norms,” *Journal of Approximation Theorey*, vol. 3, pp. 123–127, 1970.
- [84] T. Blumensath, M. Yaghoobi, and M.E. Davies, “Iterative hard thresholding and  $L_0$  regularisation,” in *Proceedings IEEE International Conference on Acoustics, Speech and Signal Processing*, 15-20 April 2007.
- [85] R.M. Figueras i Ventura and E.P. Simoncelli, “Statistically driven sparse image approximation,” in *Proceedings 14th IEEE International Conference on Image Processing*, San Antonio, TX, 16-19 September 2007, IEEE Signal Processing Society.
- [86] H.Y. Gao and A.G. Bruce, “Waveshrink with firm shrinkage,” *Statistica Sinica*, vol. 7, pp. 855–874, 1997.
- [87] M. Fornasier and H. Rauhut, “Iterative thresholding algorithms,” *To appear in Applied Computational Harmonic Analysis*, July 2007.
- [88] M. Fornasier, “Accelerated iterative thresholding algorithms,” *Technical Report*, 2007.
- [89] O.G. Guleryuz, “Nonlinear approximation based image recovery using adaptive sparse reconstructions and iterated denoising: Part I - theory,” *IEEE Transactions on Image Processing*, vol. 15, no. 3, pp. 539–554, March 2006.
- [90] O.G. Guleryuz, “Nonlinear approximation based image recovery using adaptive sparse reconstructions and iterated denoising: Part II - applications,” *IEEE Transactions on Image Processing*, vol. 15, no. 3, pp. 555–571, March 2006.
- [91] M.J. Fadili and J.L. Starck, “Sparse representation-based image deconvolution by iterative thresholding,” in *Astronomical Data Analysis (ADA’06)*, Marseille, France, September 2006.
- [92] J. Bobin, J.L. Starck, J. Fadili, Y. Moudden, and D.L. Donoho, “Morphological Component Analysis: an adaptive thresholding strategy,” *IEEE Transactions on Image Processing*, vol. 16, no. 11, pp. 2675–2681, November 2007.

- [93] S. Fischer and G. Cristobal, "Minimum entropy transform using gabor wavelets for image compression," in *Proceedings of the 11th International Conference on Image Analysis and Processing (ICIAP'01)*, Palermo, Italy, 26-28 September 2001.
- [94] G.H. Mohimani, M. Babaie-Zadeh, and Christian Jutten, "Fast sparse representation based on smoothed L0 norm," in *7th International ICA Conference (ICA 2007)*, London, UK, 9-12 September 2007.
- [95] M.A.T. Figueiredo, J.B. Dias, J.P. Oliveira, and R.D. Nowak, "On Total Variation de-noising: A new Majorization-Minimization algorithm and an experimental comparison with wavelet de-noising," in *Proceedings 13th IEEE International Conference on Image Processing*, Atlanta, GE, 8-11 October 2006, IEEE Signal Processing Society.
- [96] M.A.T. Figueiredo, J.M. Bioucas-Dias, and R.D. Nowak, "Majorization-Minimization algorithms for wavelet based image restoration," *IEEE Transactions on Image Processing*, vol. 16, no. 12, pp. 2980–2991, 2007.
- [97] I. Drori, "Fast L1 minimization by iterative thresholding for multidimensional nmr spectroscopy," *Submitted to EURASIP Journal on Advances in Signal Processing*, August 2007.
- [98] C. Rozell, D. Johnson, R. Baraniuk, and B. Olshausen, "Locally competitive algorithms for sparse approximation," in *Proceedings 14th IEEE International Conference on Image Processing*, San Antonio, TX, 16-19 September 2007, IEEE Signal Processing Society.
- [99] J. Haupt and R.D. Nowak, "Compressive sampling vs. conventional imaging," in *Proceedings 13th IEEE International Conference on Image Processing*, Atlanta, GE, 8-11 October 2006, pp. 1269–1272, IEEE Signal Processing Society.
- [100] D.L. Donoho and P.B. Stark, "Uncertainty principles and signal recovery," *SIAM Journal of Applied Mathematics*, vol. 49, no. 3, pp. 906–931, 1989.
- [101] D.L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *Technical Report*, June 1999.
- [102] E.J. Candes and J. Romberg, "Practical signal recovery from random projections," in *Proceedings of the SPIE XI Conference on Wavelet Applications in Signal and Image Processing*, 2004, vol. 5914.

- [103] E. Candes and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, December 2005.
- [104] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, February 2006.
- [105] E. Candes, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1208–1223, August 2006.
- [106] E. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5406–5425, December 2006.
- [107] M. Elad and A.M. Bruckstein, "A generalized uncertainty principle and sparse representation in pair of bases," *IEEE Transactions on Information Theory*, vol. 48, no. 9, pp. 2558–2567, September 2002.
- [108] R. Gribonval and M. Nielsen, "Sparse representations in unions of bases," *IEEE Transactions on Information Theory*, vol. 49, no. 12, pp. 3320–3325, December 2003.
- [109] D.L. Donoho and M. Elad, "Optimally sparse representation from overcomplete dictionaries via L1-norm minimization," *Proceedings of the National Academy of Sciences, USA*, vol. 100, no. 5, pp. 2197–3002, 2003.
- [110] J.J. Fuchs, "On sparse representation on arbitrary redundant bases," *IEEE Transactions on Information Theory*, vol. 50, no. 6, pp. 1341–1344, June 2004.
- [111] S. Kunis and H. Rauhut, "Random sampling of sparse trigonometric polynomials II - Orthogonal Matching Pursuit versus Basis Pursuit," *Submitted to arXiv:math/0604429v2*, February 2007.
- [112] A. Bruckstein, D.L. Donoho, and M. Elad, "From sparse solutions of system of equations to sparse modeling of signal and images," *To appear in SIAM Review*, 2007.
- [113] J.A. Tropp, I.S. Dhillon, R.W. Heath, and T. Strohmer, "Designing structured tight frames via an alternating projection method," *IEEE*

- Transactions on Information Theory*, vol. 51, no. 1, pp. 188–209, January 2005.
- [114] E.J. Candes and D.L. Donoho, “Curvelets - a surprisingly effective nonadaptive representation for objects with edges,” in *Curves and Surfaces IV*, L.L. Schumaker et al., Ed. Vanderbilt University Press, Nashville, TN, 1999.
- [115] J.A. Guerrero-Colón, L. Mancera, and J. Portilla, “Image restoration using space-variant Gaussian Scale Mixtures in overcomplete pyramids,” *IEEE Transactions on Image Processing*, vol. 17, no. 1, pp. 27–41, January 2008.
- [116] N. Kingsbury, “Web page,” <http://www-sigproc.eng.cam.ac.uk/ngk/>.
- [117] E. Candes, L. Demanet, D.L. Donoho, and L. Ying, “Curvelab,” <http://www.curvelab.org/>.
- [118] S. Kirkpatrick, C. Gelatt, and M. Vecchi, “Optimization by simulated annealing,” *Science*, vol. 220, no. 4598, pp. 671–681, May 1983.
- [119] A. Blake and A. Zisserman, “Graduated non-convexity,” in *Visual Reconstruction*, MA MIT Press, Cambridge, Ed. 1987.
- [120] F. Rooms, W. Philips, and J. Portilla, “Parametric psf estimation via sparseness maximization in the wavelet domain,” in *Wavelet Applications in Industrial Processing II*, F. Truchetet and O. Laligant, Eds. November 2004, vol. 5607, pp. 26–33, Proceedings of the SPIE.
- [121] Z. Wang, G. Wu, H.R. Sheikh, E.P. Simoncelli, E.H. Yang, and A.C. Bovik, “Quality-aware images,” *IEEE Transactions on Image Processing*, vol. 15, no. 6, pp. 1680–1689, June 2005.
- [122] M. Elad, P. Milanfar, and R. Rubinstein, “Analysis versus synthesis in signal priors,” *Inverse Problems*, vol. 23, no. 3, pp. 947–978, June 2007.
- [123] Z. Xiong, M.T. Orchard, and Y. Zhang, “A deblocking algorithm for jpeg compressed images using overcomplete wavelet representations,” *IEEE Transactions on Circuit and Systems for Video Technology*, vol. 7, no. 2, pp. 443–437, April 1997.
- [124] V.K. Goyal, M. Vetterli, and N.T. Thao, “Quantized overcomplete expansions in rn: Analysis, synthesis and algorithms,” vol. 44, no. 1, pp. 16–31, January 1998.



- [125] H. Paek, R. Kim, and S. Lee, "On the pocs-based postprocessing technique to reduce the blocking artifacts in transform coded images," *IEEE Transactions on Circuit and Systems for Video Technology*, vol. 8, no. 3, pp. 358–367, June 1998.
- [126] J. Mateos, A.K. Katsaggelos, and R. Molina, "A Bayesian approach to estimate and transmit regularization parameters for reducing blocking artifacts," *IEEE Transactions on Image Processing*, vol. 9, no. 7, pp. 1200–1215, July 2000.
- [127] X. Li, "Improved wavelet decoding via set theoretic estimation," *IEEE Transactions on Circuit and Systems for Video Technology*, vol. 15, no. 1, pp. 108–112, January 2005.
- [128] L. Moisan J.M. Morel A. Desolneux, S. Ladjal, "Dequantizing image orientation," *IEEE Transactions on Image Processing*, vol. 11, no. 10, pp. 1129–1140, October 2002.
- [129] Y.H. Chan and Y.H. Fung, "A regularized constrained iterative restoration algorithm for restoring color-quantized images," *Signal Processing*, vol. 85, no. 7, pp. 1375–1387, July 2005.
- [130] D. Keysers, C.H. Lambert, and T.M. Breuel, "Color image dequantization by constrained diffusion," in *Proceedings of the SPIE*, January 2006, vol. 6058.
- [131] E.P. Simoncelli, "The Steerable Pyramid: A flexible architecture for multi-scale derivative computation," in *Proceedings of the 2nd IEEE International Conference on Image Processing*, Washington DC, 23-26 October 1995, vol. 3, pp. 444–447, IEEE Signal Processing Society.
- [132] A. Hirani and T. Totsuka, "Combining frequency and spatial domain information for fast interactive image noise removal," in *Proceedings of the ACM SIGGRAPH 1996*, 1996.
- [133] J.S. De Bonet, "Multiresolution sampling procedure for analysis and synthesis of texture images," in *Proceedings of the ACM SIGGRAPH 1997*, 1997.
- [134] J. Portilla and E.P. Simoncelli, "Texture model based on joint statistics of complex wavelet coefficients," *International Journal of Computer Vision*, vol. 40, no. 1, pp. 49–71, 2000.

- [135] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, "Image inpainting," in *Proceedings of the SIGGRAPH 2000*, New Orleans, USA, July 2000.
- [136] M. Bertalmio, L. Vese, G. Sapiro, and S. Osher, "Simultaneous structure and texture image inpainting," *IEEE Transactions on Image Processing*, vol. 12, no. 8, pp. 882–889, August 2003.
- [137] H. Chen and I. Hagiwara, "Image reconstruction based on combination of wavelet decomposition, inpainting and texture synthesis," in *International Conferences in Central Europe on Computer Graphics, Visualization and Computer Vision (WSCG'06)*, Plzen, Czech Republic, January 30 - February 3 2006.
- [138] J.H. Kim, S.H. Lee, and N.I. Cho, "Bayesian image interpolation based on the learning and estimation of higher bandwavelet coefficients," in *Proceedings 13th IEEE International Conference on Image Processing*, Atlanta, GE, 8-11 October 2006, pp. 1269–1272, IEEE Signal Processing Society.
- [139] M.M. Oliveira, B. Bowen, R. McKenna, and Y.S. Chang, "Fast digital image inpainting," in *Proceedings of the International Conference on Visualisation, Imaging and Image Processing (VIIP 2001)*, Marbella, Spain, 3-5 September 2001.
- [140] J. Fadili, "Web page," <http://www.greyc.ensicaen.fr/jfadili/>.
- [141] B.E. Bayer, "Color imaging array," *U.S. Patent 3971065*, 1976.
- [142] B.K. Gunturk, J. Glotzbach, Y. Altunbasak, R.W. Schafer, and R.M. Meresereau, "De-mosaicking: Color filter array interpolation in single chip digital cameras," *IEEE Signal Processing Magazine, Special Issue on Color Image Processing*, 2005.
- [143] J. Portilla, D. Otaduy, and C. Dorronsoro, "Low-complexity linear demosaicing using joint spatial-chromatic image statistics," in *Proceedings of the 12th International Conference on Image Processing*, Genoa, Italy, 11-14 September 2005.
- [144] B.K. Gunturk and R.M. Mersereau, "Color plane interpolation using alternating projections," *IEEE Transactions on Image Processing*, vol. 11, no. 9, pp. 997–1013, 2002.

- [145] W. Lu and Y.P. Tan, "Color filter array demosaicing: new methods and performance measures," *IEEE Transactions on Image Processing*, vol. 12, no. 10, pp. 1194–1210, 2003.
- [146] X. Li, "Demosaicing by successive approximation," *IEEE Transactions on Image Processing*, vol. 14, no. 3, pp. 370–379, March 2005.
- [147] S.C. Pei and I.K. Tam, "Effective color interpolation in ccd color filter arrays using signal correlation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 6, pp. 503–512, 2003.
- [148] D. Alleysson, S. Susstrunk, and J. Herault, "Linear demosaicing inspired by the human visual system," *IEEE Transactions on Image Processing*, vol. 14, no. 4, pp. 1–12, April 2005.
- [149] N.X. Lian and Y.P. Tan, "An efficient and effective color filter array de-mosaicking method," in *Proceedings 14th IEEE International Conference on Image Processing*, San Antonio, TX, 16-19 September 2007, IEEE Signal Processing Society.
- [150] X. Li, "Web page," <http://www.csee.wvu.edu/~xinl/>.
- [151] 40 scanned images, "Eastman kodak(c) photographic color image database," 1993.
- [152] M.D. Fairchild, *Color Appearance Models*, Addison-Wesley, 1997.
- [153] D. Barreto, L.D. Alvarez, R. Molina, A.K. Katsaggelos, and G.M. Callico, "Region-based super-resolution for compression," *Multidimensional Systems and Signal Processing, special issue on papers presented at the I International Conference in Super Resolution (Hong Kong, 2006)*, vol. 18, no. 2.
- [154] C.A. Segall, A.K. Katsaggelos, R. Molina, and J. Mateos, "Bayesian resolution enhancement of compressed video," *IEEE Transactions on Image Processing*, vol. 13, no. 7, pp. 898–911, July 2004.
- [155] S. Farsiu, M. Elad, and P. Milanfar, "Video-to-video dynamic super-resolution for grayscale and color sequences," *EURASIP Journal on Applied Signal Processing, Special Issue on Superresolution Imaging*, pp. 1–15, 2006, Article ID 61859.

- [156] C.B. Atkins, C.A. Bouman, and J.P. Allebach, “Optimal image scaling using pixel classification,” in *Proceedings of the 6th International Conference on Image Processing*, Thessaloniki, Greece, 7-10 October 2001, pp. 864–867.
- [157] S. Battiato, G. Gallo, and F. Stanco, “Smart interpolation by anisotropic diffusion,” in *Proceedings of 12th International Conference on Image Analysis and Processing*, Barcelona, Spain, 2003.
- [158] D.D. Muresan and T.W. Parks, “Adaptively quadratic (aqua) image interpolation,” *IEEE Transactions on Image Processing*, vol. 13, no. 5, pp. 690–698, May 2004.
- [159] M.F. Tappen, B.C. Russell, and W.T. Freeman, “Exploiting the sparse derivative prior for super-resolution and image demosaicing,” in *3rd International Workshop on Statistical and Computational Theories of Vision*, 2003.
- [160] S. Farsiu, M. Elad, and P. Milanfar, “Multi-frame demosaicing and super-resolution of color images,” *IEEE Transactions on Image Processing*, vol. 15, no. 1, pp. 141–159, January 2006.
- [161] J.D. van Ouwerkerk, “Image super-resolution survey,” *Image and Vision Computing*, vol. 24, no. 10, pp. 1039–1052, October 2006.
- [162] G.H. Golub and C.F. Van Loan, *Matrix Computations*, John Hopkins University Press.