

Μπεϋζιανή Ανόρθωση και Ανακατασκευή Εικόνων Υψηλής Ανάλυσης από Άγνωστης  
Υποβάθμισης Δεδομένα Χαμηλής Ανάλυσης

Η  
ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

Υποβάλλεται στην

ορισθείσα από την Γενική Συνέλευση Ειδικής Σύνοψης  
του Τμήματος Πληροφορικής  
Εξεταστική Επιτροπή

από τον

Ιωάννη Χάντα

ως μέρος των Υποχρεώσεων

για τη λήψη

του

ΔΙΔΑΚΤΟΡΙΚΟΥ ΔΙΠΛΩΜΑΤΟΣ ΣΤΗΝ ΠΛΗΡΟΦΟΡΙΚΗ

Οκτώβριος 2008

**This thesis is part of the 03D535 research project, implemented within the framework of the “Reinforcement Programme of Human Research manpower” (PENED) and co-financed by National and Community Funds (25% from the Greek Ministry of Development-General Secretariat of Research and Technology and 75% from E.U.-European Social Fund).**

### **Τριμελής συμβουλευτική επιτροπή:**

- Αριστείδης Λύκας, Αναπληρωτής Καθηγητής Τμήματος Πληροφορικής του Πανεπιστημίου Ιωαννίνων (επιβλέπων).
- Νικόλαος Γαλατσάνος, Καθηγητής Τμήματος Ηλεκτρολόγων Μηχανικών και Τεχνολογίας Υπολογιστών του Πανεπιστημίου Πατρών.
- Ισαάκ Λαγαρής, Καθηγητής Τμήματος Πληροφορικής του Πανεπιστημίου Ιωαννίνων.

### **Επταμελής εξεταστική επιτροπή:**

- Αριστείδης Λύκας, Αναπληρωτής Καθηγητής Τμήματος Πληροφορικής του Πανεπιστημίου Ιωαννίνων (επιβλέπων).
- Νικόλαος Γαλατσάνος, Καθηγητής Τμήματος Ηλεκτρολόγων Μηχανικών και Τεχνολογίας Υπολογιστών του Πανεπιστημίου Πατρών.
- Ισαάκ Λαγαρής, Καθηγητής Τμήματος Πληροφορικής του Πανεπιστημίου Ιωαννίνων.
- Χριστόφορος Νίκου, Λέκτορας Τμήματος Πληροφορικής του Πανεπιστημίου Ιωαννίνων.
- Κωνσταντίνος Μπλέκας, Λέκτορας Τμήματος Πληροφορικής του Πανεπιστημίου Ιωαννίνων.
- Γεώργιος Μουστακίδης, Καθηγητής Τμήματος Ηλεκτρολόγων Μηχανικών και Τεχνολογίας Υπολογιστών του Πανεπιστημίου Πατρών.
- Κωνσταντίνος Μπερμπερίδης, Καθηγητής Τμήματος Μηχανικών Η/Υ και Πληροφορικής του Πανεπιστημίου Πατρών.

# **DEDICATION**

---

To my parents, Costas and Sofia.

## ACKNOWLEDGMENTS

---

I would like to sincerely thank Dr. Nikolaos Galatsanos, Professor at the Department of Electrical and Computer Engineering, University of Patras, for his continuous and tireless supervision of this dissertation, his advice, directive thoughts and inspiration, and his faith in me despite all the difficulties that I encountered. I would also like to thank Dr. Aristidis Likas Associate Professor and Isaak Lagaris, Professor, both at the Department of Computer Science, University of Ioannina, for their valuable counseling and teaching of useful tools during my studies that found application in this dissertation. I would like specially to thank Dr. A. Likas for his unreserved help in the writing of this dissertation. Also, I would like to thank Dr. A. Likas and Dr. Christophoros Nikou, Department of Computer Science, University of Ioannina, for their help in the preparation of the thesis defence.

I would like also to thank Prof. Rafael Molina (Universidad de Granada), Prof Michael Saunders (Stanford University), C. Constantinopoulos, and the PhD candidates A. Mairgiotis and D. Tzikas from the Departments of Computer Science, University Ioannina and S. Mpelekos, PhD candidate from the Department of Physics of University of Athens.

Finally, I would like to thank my wife Jenny for her love and patience during the many years of my studies and my parents Costas and Sofia and my sisters Chrysanthi and Natasa for their help, love and understanding.

# CONTENTS

---

	Page
DEDICATION	iv
LIST OF TABLES	viii
LIST OF FIGURES	ix
ABSTRACT	xi
EKTENHΣ ΠΕΡΙΛΗΨΗ ΣΤΑ ΕΛΛΗΝΙΚΑ	xiii
CHAPTER 1. INTRODUCTION	1
1.1 Image restoration and super-resolution	1
1.2 Imaging model for image restoration	2
1.3 Image restoration: an ill-posed problem	4
1.4 Regularization in image restoration	6
1.4.1 <i>Stochastic regularization</i>	6
1.4.2 <i>Deterministic regularization</i>	8
1.4.3 <i>Estimation of the regularization parameter</i>	10
1.5 Super-resolution: reconstructing a high-resolution image from low-resolution images	12
1.6 Thesis contribution	15
CHAPTER 2. APPROXIMATE BAYESIAN INFERENCE	19
2.1 Expectation Maximization (EM) algorithm: a variational point of view	20
2.2 Approximate Bayesian inference with variational bound maximization	22
2.3 Constrained variational optimization	24
2.4 Bayesian inference for restoration with the SAR prior model	25
CHAPTER 3. SPATIALLY ADAPTIVE IMAGE RESTORATION	27
3.1 Visibility based non-stationary restoration	28
3.2 Half-Quadratic Regularization	29
3.3 Markov Random Fields	33
3.4 Total-variation regularization	36
3.5 Regularization in the Wavelet Domain	40
3.6 Other image priors	44
3.7 Super-resolution methods	44
3.8 Conclusions	46
CHAPTER 4. BAYESIAN IMAGE RESTORATION USING A HIERARCHICAL NON-STATIONARY PRIOR	47
4.1 Introduction	47
4.2 Imaging and image prior model	50
4.3 Maximum a posteriori (MAP) estimation	54
4.4 Bayesian algorithm	55

4.5	Numerical experiments	57
4.6	Conclusions and extensions	60
CHAPTER 5. BAYESIAN IMAGE RESTORATION WITH A PRIOR BASED ON A PRODUCT OF t-DISTRIBUTIONS		67
5.1	Introduction	68
5.2	Imaging and Image Model	69
5.3	Variational Inference	71
5.3.1	<i>The Variational Algorithm for Equivalent Imaging Model</i>	71
5.3.2	<i>The Variational Update Equations</i>	74
5.4	Computational Implementation	77
5.5	Numerical Experiments	80
5.6	Conclusions and suggestions for improvement	82
	Appendix	87
CHAPTER 6. BAYESIAN IMAGE RESTORATION WITH A SPATIALLY ADAPTIVE TOTAL VARIATION PRIOR		90
6.1	Introduction	90
6.2	Imaging and image model	92
6.2.1	<i>Generalized Student's-t image prior</i>	92
6.3	Variational inference with the generalized Student's-t prior	94
6.3.1	<i>A Lower Bound for <math>L(q(\boldsymbol{\varepsilon}, \mathbf{a}), \theta)</math></i>	96
6.4	A Constrained Variational Inference Algorithm	96
6.5	Computational Implementation	100
6.6	Numerical Experiments	102
6.7	Conclusions	104
	Appendix	104
CHAPTER 7. SUPER-RESOLUTION BASED ON FAST REGISTRATION AND MAXIMUM A POSTERIORI RECONSTRUCTION		112
7.1	Introduction	112
7.1	Introduction	112
7.2	Imaging model	113
7.3	Image prior model	115
7.4	Pre-processing step of the super-resolution algorithm	117
7.5	Maximum a posteriori (MAP) reconstruction	119
7.6	Experiments	122
7.7	Conclusions and future work	125
	Appendix	126
CHAPTER 8. CONCLUSIONS AND FUTURE RESEARCH		133

## LIST OF TABLES

---

Table	Page
4.1: ISNR comparisons with the experiments in Table I of [55]	61
4.2: ISNR comparisons with the experiments in Table III of [55]	63
4.3: <i>Bias</i> metric for the MAP and the Bayesian algorithms	63
4.4: <i>Variance</i> metric for the MAP and the Bayesian algorithms	64
5.1: ISNR results comparing the proposed algorithm with the algorithms in [17], [18] and [8] using 3 images, 3 noise levels and Gaussian shaped blur. The <i>ISNR</i> results for the <i>BFO1</i> , <i>BFO2</i> , <i>BMK1</i> and <i>BMK2</i> algorithms are obtained from [8]	84
5.2: ISNR results comparing the proposed algorithm with the algorithms in [17], [18] and [8] using 3 images, 3 noise levels and uniform blur. The <i>ISNR</i> results for the <i>BFO1</i> , <i>BFO2</i> , <i>BMK1</i> and <i>BMK2</i> algorithms are obtained from [8]	85
5.3: <i>ISNR</i> results comparing the proposed algorithm with the algorithms in [17] using 3 images, 3 noise levels and pyramidal blur	87
6.1: <i>ISNR</i> 's for the experiments using uniform blur $9 \times 9$	107
6.2: <i>ISNR</i> 's for the experiments using Gaussian blur with variance 9	108
6.3: <i>ISNR</i> 's for the experiments using pyramidal blur	109
7.1 Original and estimated parameters for the artificially generated images	110



# LIST OF FIGURES

---

Figure	Page
1.1 (a) ‘Lena’ original image, (b) degraded image and (c) the result of direct inverse filtering to the degraded image (without regularization)	6
1.2 (a) ‘Lena’ original image, (b) degraded image, (c) over-regularized restored image, (d) under-regularized restored image and (e) restored image with average regularization	11
1.3 (a)-(d) Low resolution observations, (e) 2x super-resolved image regularized under the SAR image model [133]	14
4.1: The graphical model of the observations	53
4.2: Experiments with ‘Lena’ image, Gaussian blur and $SNR=25,36dB$	65
4.3: Experiments with ‘Cameraman’ image, uniform blur and $SNR= 38.64dB$	66
5.1: a) Degraded “Cameraman” image by uniform 9x9 blur and noise with $BSNR=40dB$ , b) restored image using a stationary Gaussian prior [94] $ISNR=5.76 dB$ , c) restored image using TV-TE $ISNR = 9.07 dB$ , d) restored image using proposed algorithm $ISNR =9.53dB$	86
5.2: Magnitude of frequency responses of the filters used in the prior: horizontal differences ( $Q_1$ ), (b) vertical differences ( $Q_2$ ), (c) $Q_3$ and (d) $Q_4$	86
6.1: Experiment on Cameraman image with uniform $9 \times 9$ blur and $BSNR = 30$ ; $ISNR$ and $VIF$ [120] comparisons: (a) Degraded image, $VIF = 2.01$ (b) restored with spatially invariant prior, [94], $ISNR = 3.29$ and $VIF = 2.22$ , (c) restored image with method in Chapter 5, $ISNR = 5.88$ and $VIF = 4.52$ , (d) restored image with the proposed algorithm, $ISNR = 6.55$ and $VIF = 4.55$	110

6.2: Experiment on Lena image with Gaussian blur (variance 9) and $BSNR = 20$ ; $ISNR$ and $VIF$ [120] comparisons: (a) Degraded image, $VIF = 0.04$ , (b) Restored with spatially invariant prior, [94], $ISNR = 2.32$ and $VIF = 0.31$ , (c) restored image with method in Chapter 5, $ISNR = 2.76$ and $VIF = 0.38$ , (d) restored image with the proposed algorithm, $ISNR = 3.14$ and $VIF = 0.39$	111
7.1.1: Low resolution degraded observation	128
7.1.2: (a) Stationary 2x [133], ( $MSE=195.2$ ), (b) Total Variation, Eq. (7.12) ( $MSE=182.1$ ), (c) Non-stationary MAP 2x super-resolved image ( $MSE= 162.4$ )	129
7.2.1: Low resolution observation	130
7.2.2: Super-resolved images; (a) 2x stationary [133], (b) 2x Total Variation, Eq. (7.12), (c) 2x MAP non-stationary	130
7.3.1: Low resolution observation	131
7.3.2: Super-resolution experiment with real data; super-resolved images; (a) 4x stationary [133], (b) 4x Total Variation, Eq. (7.12), (c) 4x MAP non-stationary	131

# ABSTRACT

---

Giannis Chantas, K. S.

PhD, Computer Science Department, University of Ioannina, Greece. October, 2008.

Title: Bayesian Restoration and Reconstruction of High-Resolution Images from Low-Resolution Images with Unknown Degradations.

Thesis Supervisor: Aristidis Likas.

The research topic of this dissertation is the development of stochastic non-stationary image models as image priors used for regularization in restoration and super-resolution problems. The proposed non-stationary image models lead to spatially adaptive regularization, or in other words, non-uniform regularization along the image which depends on the local spatial activity. Furthermore, working in the stochastic framework we employ the Bayesian methodology to solve these inverse problems and in parallel to estimate the model parameters. Thus, development of Bayesian restoration and reconstruction algorithms is another major issue of this dissertation.

First, we introduce a new hierarchical (two-level) Gaussian non-stationary image prior. This prior assumes that the residuals of the first order differences of the image, in four *different directions*, are Gaussian random variables with zero mean and variance that is *spatially varying*. In this way, the variances manifest the spatial adaptivity mechanism. In order to deal with the resulting over-parameterization of this model, the spatially varying variances are considered random variables (not parameters) and a Gamma hyper-prior is imposed on them, which is *conjugate* to the Gaussian. To learn this model and infer the image we propose two iterative algorithms. The first is based on the maximum a posteriori estimation (MAP) principle and computes explicitly both the image and the spatially varying variances in all four directions. The second is a Bayesian algorithm that marginalizes the “hidden variables”. Also, the marginalization of the hidden variables produces a Student’s-t distribution.

Next, we propose a new Bayesian inference framework for image restoration using a prior in product form. This prior assumes that the outputs of local high-pass filters, (their number is arbitrary), follow again the Student's-t distribution. Then, a Bayesian inference methodology is proposed that bypasses the difficulty of evaluating the normalization constant of product type priors. The methodology is based on a *constrained* variational approximation that uses the outputs of all the local high-pass filters to produce an estimate of the original image. In this manner the use of improper priors is avoided and *all* the parameters of the prior model are estimated from the data.

As a next step, we extend the total-variation prior by introducing a new prior which has a number of novel features. More specifically, we introduce a total-variation (TV) prior with spatially varying regularization parameters. In order to avoid the over parameterization, we introduce a Gamma hyperprior for the spatially adaptive regularization parameters of the local TV priors. We also use this prior in a product form, which means that we assume that the outputs of an arbitrary number of high-pass filters are distributed according to this prior. This gives two novel features to the new prior. First, it is explicitly spatially adaptive and thus it is better suited to capture the salient features of the image. Second, it is in product form and has the ability to enforce simultaneously a number of different properties to the image. If the hidden variables of the second layer are marginalized, the resulting density function has a form similar to a Student's-t distribution; thus, we label it as Generalized Student's-t. Due to the complexity of this model, we resort to the variational approximation for Bayesian inference.

Finally, we present our contribution to the super-resolution problem. We utilize for the first time in the super-resolution problem a hierarchical two-level image prior. Using this non-stationary prior, it is possible to reconstruct high-resolution images without smoothed edges or ringing artifacts in the vicinity of edges. Another contribution to the super-resolution problem is a novel two-step reconstruction algorithm. The first stage of this algorithm is a preprocessing step that approximately registers the degraded low-resolution observations. These "almost-registered" low-resolution observations are used subsequently by an iterative algorithm which simultaneously reconstructs the high-resolution images and finds their registration parameters.

## ΕΚΤΕΝΗΣ ΠΕΡΙΛΗΨΗ ΣΤΑ ΕΛΛΗΝΙΚΑ

---

Ιωάννης Χάντας του Κωνσταντίνου και της Σοφίας.

Διδακτορικό Δίπλωμα, Τμήμα Πληροφορικής, Πανεπιστήμιο Ιωαννίνων, Οκτώβριος, 2008.

Τίτλος: Μπεϋζιανή Ανόρθωση και Ανακατασκευή Εικόνων Υψηλής Ανάλυσης από Αγνώστης Υποβάθμισης Δεδομένα Χαμηλής Ανάλυσης.

Επιβλέπωντας: Αριστείδης Λύκας.

Το ερευνητικό αντικείμενο της διατριβής αυτής σχετίζεται με την ανάπτυξη πρωτότυπων μεθοδολογιών για τα προβλήματα της ανόρθωσης εικόνων (image restoration) και της υπερ-ανάλυσης εικόνων (image super-resolution). Πιο συγκεκριμένα, η διατριβή επικεντρώνεται στη μελέτη χωρικά μεταβαλλόμενων στοχαστικών μοντέλων κατάλληλων για να χρησιμοποιηθούν ως εκ των προτέρων κατανομές (priors) προκειμένου να επιτευχθεί κανονικοποίηση (regularization) στα προβλήματα της ανόρθωσης και της υπερ-ανάλυσης εικόνων. Με τα προτεινόμενα μη-στατικά μοντέλα εικόνας επιτυγχάνεται τοπικά προσαρμοζόμενη κανονικοποίηση, δηλαδή ανομοιόμορφη κανονικοποίηση της εικόνας εξαρτώμενη από την τοπική χωρική δραστηριότητα. Χρησιμοποιώντας τη στοχαστική προσέγγιση για τη μοντελοποίηση των εικόνων, εφαρμόζεται η Μπεϋζιανή μεθοδολογία για τη λύση των αντίστροφων παραπάνω προβλημάτων καθώς και για την εκτίμηση των παραμέτρων του μοντέλου. Κατά συνέπεια, η ανάπτυξη Μπεϋζιανών αλγορίθμων ανόρθωσης και υπερ-ανάλυσης είναι ένα ακόμη βασικό πεδίο έρευνας της διατριβής.

Στη διατριβή αυτή προτείνεται καταρχήν μια νέα ιεραρχική (δύο επιπέδων) εκ των προτέρων κατανομή που είναι Γκαουσιανή και μη-στατική. Αυτή η κατανομή θεωρεί ότι οι πρώτες διαφορές των εικόνων, σε τέσσερις διαφορετικές διευθύνσεις, είναι Γκαουσιανές τυχαίες μεταβλητές με χωρικά μεταβαλλόμενη διακύμανση. Με τον τρόπο αυτό, οι διακυμάνσεις υλοποιούν το μηχανισμό της χωρικής μεταβλητότητας. Για να αντιμετωπιστεί το ζήτημα της υπερ-παραμετροποίησης αυτού του μοντέλου, οι χωρικά μεταβαλλόμενες διακυμάνσεις θεωρούνται τυχαίες μεταβλητές (όχι παράμετροι) που ακολουθούν μια κοινή κατανομή

Γάμμα. Για την εκπαίδευση του μοντέλου και την εκτίμηση της εικόνας προτείνονται δύο επαναληπτικοί αλγόριθμοι. Ο ένας βασίζεται στην αρχή της maximum a posteriori (MAP) εκτίμησης και υπολογίζει άμεσα και την εικόνα και τις χωρικά μεταβαλλόμενες διακυμάνσεις. Ο άλλος είναι ένας Μπεϋζιανός αλγόριθμος που βασίζεται στην περιθωριοποίηση των ενδιάμεσων «κρυμμένων» μεταβλητών.

Στη συνέχεια, προτείνεται μια νέα Μπεϋζιανή προσέγγιση για ανόρθωση εικόνων στην οποία χρησιμοποιείται μια εκ των προτέρων κατανομή για την εικόνα η οποία έχει μορφή γινομένου. Παρουσιάζεται μια μεθοδολογία που ξεπερνά τη δυσκολία της εκτίμησης της σταθεράς κανονικοποίησης των κατανομών τύπου γινομένου και βασίζεται σε μια variational προσέγγιση με περιορισμούς (constrained variational approximation). Με τον τρόπο αυτό αποφεύγεται η χρήση μη κανονικοποιημένων (improper) κατανομών και όλες οι παράμετροι του μοντέλου εκτιμώνται από τα δεδομένα.

Ως επόμενο βήμα, προτείνεται μια επέκταση της total-variation (TV) εκ των προτέρων κατανομής για την εικόνα μέσω της εισαγωγής χωρικά μεταβαλλόμενων παραμέτρων κανονικοποίησης. Για την αποφυγή της υπερ-παραμετροποίησης, επιβάλλουμε μια Γάμμα κατανομή για τις χωρικά μεταβαλλόμενες παραμέτρους των τοπικών TV κατανομών. Επιπλέον, χρησιμοποιούμε την κατανομή αυτή σε μορφή γινομένου. Αν περιθωριοποιηθούν οι ενδιάμεσες μεταβλητές, η συνάρτηση πυκνότητας πιθανότητας που προκύπτει έχει μορφή όμοια με της Student's-t κατανομής και την ονομάζουμε Γενικευμένη Student's-t. Για το παραπάνω στοχαστικό μοντέλο προτείνεται μια Μπεϋζιανή μεθοδολογία για την εκτίμηση της εικόνας και των παραμέτρων του μοντέλου.

Τέλος, παρουσιάζεται μια νέα προσέγγιση στο πρόβλημα της υπερ-ανάλυσης εικόνων, χρησιμοποιώντας για πρώτη φορά μια ιεραρχική χωρικά προσαρμοζόμενη εκ των προτέρων κατανομή. Το γεγονός αυτό οδηγεί σε ανακατασκευασμένες εικόνες υψηλής ανάλυσης χωρίς ομαλοποιημένες ακμές ή ringing artifacts στην γειτονιά των ακμών. Μια άλλη συνεισφορά στο πρόβλημα της υπερ-ανάλυσης είναι ένας καινοτόμος αλγόριθμος ανακατασκευής δύο σταδίων. Το πρώτο στάδιο του αλγορίθμου είναι ένα βήμα προ-επεξεργασίας όπου γίνεται υπέρθεση (registration) μεταξύ των εικόνων χαμηλής ανάλυσης. Οι προκύπτουσες εικόνες χρησιμοποιούνται στη συνέχεια από έναν επαναληπτικό αλγόριθμο που ταυτόχρονα ανακατασκευάζει την εικόνα υψηλής ανάλυσης και βρίσκει τις παραμέτρους της υπέρθεσης.

# CHAPTER 1. INTRODUCTION

---

- 1.1. Image restoration and super-resolution
  - 1.2. Imaging model for image restoration
  - 1.3. Image restoration: an ill-posed problem
  - 1.4. Regularization in image restoration
  - 1.5. Super-resolution: reconstructing a high-resolution image from low-resolution images
  - 1.6. Thesis contribution
- 

## **1.1 Image restoration and super-resolution**

Digital image restoration is the problem of estimating the original scene of a digital image given a degraded version of it. It is a problem with a long history and emerged when the first imaging systems appeared. The imperfections of the imaging systems and the artifacts that are created when the images are produced motivated the study of methods to restore degraded images. Restoration can be applied to both analog and digital systems. However, the development of digital imaging systems has made digital image restoration more popular than the analog and provoked even more research for this field.

Applications of image restoration are numerous. For example, astronomical images from spacecrafts and telescopes are usually degraded due to atmospheric turbulence and imperfect optics. Photography is another field where image restoration can be useful when the focus is not good or there is motion. Also, because image restoration is strongly related to image reconstruction (where the observations are multiple projections of the image), advances in the first area lead to improvements in the latter. Magnetic Resonance Imaging (MRI), f-MRI, Positron Emission Tomography (PET) and super-resolution are examples of imaging modalities where reconstruction is the goal.

Super-resolution is the problem of reconstructing a high-resolution image from low-resolution images of the same scene. The desire for images with high resolution without the cost of using expensive optical imaging systems has led to the development of image processing techniques that produce a high-resolution image from multiple under-sampled images of the same scene, taken with cheaper optical systems. The key idea of super-resolution is to fuse in one image the information included in multiple low-resolution images. Reconstruction techniques for high-resolution images are very similar to image restoration methodologies. Thus, development of efficient restoration algorithms provides automatically efficient super-resolution algorithms. For this reason, in this dissertation we deal with these problems in parallel.

More specifically, image restoration and super-resolution problems are formulated using the linear equation

$$\mathbf{g} = \mathbf{H}\mathbf{f} + \mathbf{n} ,$$

where  $\mathbf{g}$  are the observations  $\mathbf{f}$  is the original image and  $\mathbf{n}$  is the noise. In image restoration  $\mathbf{H}$  is the blurring operator that blurs the image [5]. In super-resolution it is the operator that produces the degraded low-resolution observations  $\mathbf{g}$ , [105]. In tomographic reconstruction (or computed tomography, CT), the matrix  $\mathbf{H}$  is the Radon transform operator [7] and  $\mathbf{g}$  are the projections that the image is reconstructed from.

In this dissertation we deal with the restoration and super-resolution problems and in the next sections we present an introduction of these problems. Techniques for image restoration make use of models for the degradation mechanism (imaging models) as well as for the images. The assumption of a particular image model corresponds to introduction of prior knowledge of the image structure. This prior knowledge is also needed for super-resolution. Thus, realistic and efficient image models for the restoration problem can also be adapted and applied to super-resolution. Of course, super-resolution is a different problem because there are different parameters in the imaging model that have to be estimated along with the reconstruction of the high-resolution image.

## 1.2 Imaging model for image restoration



Degradation of images is usually caused by two successive operations [1]. The first operation is a spatial degradation which is caused from various factors, i.e. atmospheric turbulence, out of focus camera and/or motion. The result is a blurred version of the original scene. The second operation is a point degradation according to which noise affects the individual pixel gray level. This is caused by various types of noise (shot, thermal) in the detector systems and errors in the recording process due to quantization.

In this dissertation we will assume, as in many other works, space-invariant blurring and additive white Gaussian noise (AWG). Thus, formally, the imaging model is linear and for pixel  $(i, j)$  it is given by:

$$g(i, j) = \sum_{k=1}^N \sum_{l=1}^N h(i-k, j-l) f(k, l) + n(i, j),$$

which can also be written as:

$$\mathbf{g} = \mathbf{h} ** \mathbf{f} + \mathbf{n}. \quad (1.1)$$

The observed image  $\mathbf{g}$ , is an  $N \times N$  matrix, and is produced first by a convolution of the original image  $\mathbf{f}$ , also an  $N \times N$  matrix, with a linear shift-invariant (LSI) low-pass filter  $\mathbf{h}$ . This is also called the point spread function (PSF), since it spreads an impulse to many pixels, and models the spatial degradation mechanism. The operation denoted by  $**$  is the two-dimensional convolution. Then, AWG noise is added, represented by the vector  $\mathbf{n}$ , for which it is assumed that  $\mathbf{n} \sim N(\mathbf{0}, \beta^{-1}\mathbf{I})$  where  $\mathbf{0}$  and  $\mathbf{I}$  are the  $N \times N$  zero and identity matrices respectively. The noise variance  $\beta$  is assumed unknown, in contrast to the PSF  $\mathbf{h}$ , which will be assumed known in this dissertation.

To use a more convenient notation we express the convolution as a matrix-vector multiplication of the  $N^2 \times N^2$  matrix  $\mathbf{H}$  (representing a linear convolution operator) by the vector  $\mathbf{h}$ . The equivalent equation is:

$$\mathbf{g} = \mathbf{H}\mathbf{f} + \mathbf{n}, \quad (1.2)$$

where  $\mathbf{g}$ ,  $\mathbf{f}$  and  $\mathbf{n}$  denote  $N^2 \times 1$  vectors ordered lexicographically.

Due to the spatially invariant nature of the PSF, the matrix  $\mathbf{H}$  is block-Toeplitz. However, in order to make the multiplication of the matrix with a vector fast, we approximate it by a block circulant matrix [1]. The larger the dimension  $N$  is, the better the approximation becomes

[68]. The use of this approximation is the same with the assumption that the convolution in Eq. (1.1) is circular. However, one can avoid the approximation by padding the vectors in the convolution with zeros, and convert any linear convolution to a circular one. The padding will result in a different block-circulant matrix, but will abolish the approximation.

Padding or not, the block-circulant matrix  $\mathbf{H}$  is formed as follows:

$$\mathbf{H} = \begin{bmatrix} H(0) & H(N-1) & \cdots & H(1) \\ H(1) & H(0) & \cdots & H(N-1) \\ \vdots & \vdots & \ddots & \vdots \\ H(N-1) & H(N-2) & \cdots & H(0) \end{bmatrix},$$

where each sub-matrix is a circulant matrix, formed by  $\mathbf{h}$  (padded or not), the LSI filter:

$$H(i) = \begin{bmatrix} h(i,0) & h(i,N-1) & \cdots & h(i,1) \\ h(i,1) & h(i,0) & \cdots & h(i,N-1) \\ \vdots & \vdots & \ddots & \vdots \\ h(i,N-1) & h(i,N-2) & \cdots & h(i,0) \end{bmatrix}.$$

The circulant form leads to more tractable equations and an easy to handle model in the discrete Fourier transform (DFT) domain. This is because the eigenvalues of all circulant matrices are obtained by the DFT of the filter  $\mathbf{h}$  [1]. In other words, the eigenvectors of circulant matrices correspond to the complex exponentials basis functions of the DFT. Precisely, a circulant  $N \times N$  matrix  $\mathbf{A}$  is diagonalized as follows:

$$\mathbf{A} = N^{-1} \mathbf{W} \mathbf{\Lambda} \mathbf{W}^{-1}$$

where  $\mathbf{\Lambda}$  is diagonal with elements (the eigenvalues of  $\mathbf{A}$ ) the DFT coefficients of the filter that corresponds to this matrix. Also,  $N^{-1/2} \mathbf{W}$  is the DFT operator matrix and  $N^{-1/2} \mathbf{W}^{-1}$  its inverse [1]. In the case of a block circulant matrix the same equations hold. The only difference is that instead of the one-dimensional DFT transform, two-dimensional DFT is used. In this dissertation the notation that will be used for simplicity is one-dimensional.

### 1.3 Image restoration: an ill-posed problem

Given the imaging model, we are faced with the problem of estimating the original image  $\mathbf{f}$ , denoted by  $\hat{\mathbf{f}}$ , from the degraded observation  $\mathbf{g}$ . Inverse filtering using the degradation matrix  $\mathbf{H}$  is the most direct but naive way for restoration:

$$\hat{\mathbf{f}} = \mathbf{H}^{-1}\mathbf{g}.$$

In the DFT domain, using the diagonalization of the previous section we get:

$$\hat{\mathbf{f}} = \mathbf{H}^{-1}\mathbf{g} \Rightarrow \hat{\mathbf{f}} = \frac{1}{N} \mathbf{W}\mathbf{\Lambda}_h^{-1}\mathbf{W}^{-1}\mathbf{g} \Rightarrow \mathbf{W}^{-1}\hat{\mathbf{f}} = \mathbf{\Lambda}_h^{-1}\mathbf{W}^{-1}\mathbf{g} \Rightarrow \hat{\mathbf{F}} = \mathbf{\Lambda}_h^{-1}\mathbf{G},$$

where,  $\mathbf{\Lambda}_h$  is a diagonal matrix with the eigenvalues of  $\mathbf{H}$  (the DFT coefficients of  $\mathbf{h}$ ), and  $\hat{\mathbf{F}}, \mathbf{G}$  are the images in the DFT domain. According to the imaging model, noise has been added to the observed image, so using Eq.(1.2) the estimated image will be:

$$\hat{\mathbf{f}} = \mathbf{f} + \mathbf{H}^{-1}\mathbf{n}. \quad (1.3)$$

Therefore, because of the ill-conditioned nature of  $\mathbf{H}$ , the noise in the estimated image will be greatly amplified. To make this clearer, Eq. (1.3) can be written in the frequency domain:

$$\hat{\mathbf{F}}(i) = \frac{\mathbf{G}(i)}{\mathbf{\Lambda}_h(i,i)} = \mathbf{F}(i) + \frac{\mathbf{N}(i)}{\mathbf{\Lambda}_h(i,i)}, \quad i = 1 \dots N, \quad (1.4)$$

where  $\mathbf{F}$  and  $\mathbf{N}$  are the DFT of the image and noise vectors respectively. Notice that the eigenvalues  $\mathbf{\Lambda}_h(i,i)$  must not be zero for the inversion to be possible. However, a disastrous but usual situation appears when large levels of noise correspond to high frequencies, in which case the eigenvalues  $\mathbf{\Lambda}_h(i,i)$  are close to zero. This results in restored images that contain amplified high frequency noise.

The direct estimate of the image through inverse filtering of the observations using  $\mathbf{H}$  leads to noise amplification. In Figure 1.1(c) the result of this operation to the degraded image of Figure 1.1(b) is illustrated. In the literature, inverse problems similar to this are called “ill-posed”. Loosely speaking, this means that small changes to the observed data can cause very large changes to the estimated solution, or in the worst case the original data cannot be recovered (restored in our case), even in the absence of noise (i.e.  $\mathbf{H}$  is not invertible). In order to ameliorate the difficulties of ill-posed problems, the theory of regularization has been introduced which converts an ill-posed problem to a well-posed [1], [127]. This is achieved by constraining the set of admissible solutions using a priori knowledge about the image. The

quality of the solution depends highly on the regularization mechanism that the image model provides. This is why the image model must be accurate and realistic in order to obtain solutions as close as possible to the original image.

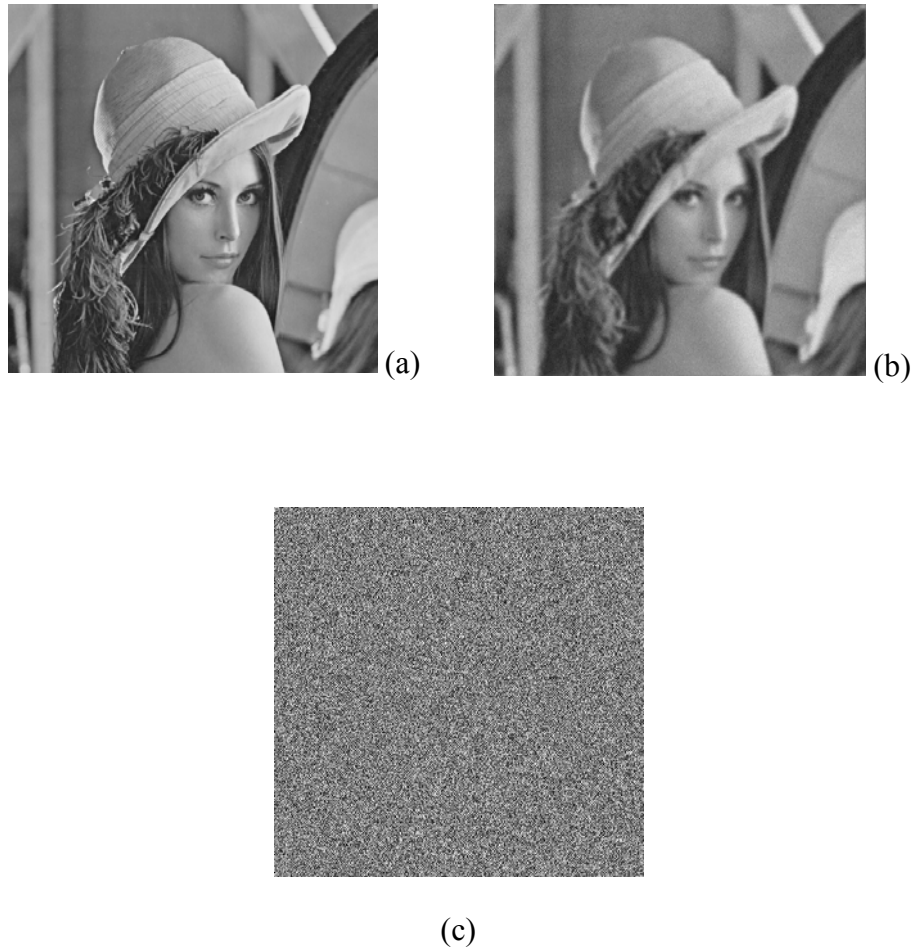


Figure 1.1 (a) ‘Lena’ original image, (b) degraded image and (c) the result of direct inverse filtering to the degraded image (without regularization).

## 1.4 Regularization in image restoration

### 1.4.1 Stochastic regularization

Let us now present a method to enforce regularization to the solution of the restoration problem. According to this, a regularized estimate of the image is the linear minimum mean

square error (LMMSE) estimate [1]. This is based on second order statistics of the image and noise. More specifically, let  $\mathbf{R}_{ff} = E[\mathbf{ff}^T]$  and  $\mathbf{R}_{nn} = E[\mathbf{nn}^T]$  (where  $E[\cdot]$  is the expectation operator) the image and noise covariances, respectively. The covariances can be estimated from the degraded image for  $\mathbf{R}_{ff}$  and from a flat region of the image for  $\mathbf{R}_{nn}$ . The LMMSE estimate which is also called Wiener filter [1] minimizes the expectation:

$$\min_{\hat{\mathbf{f}}} \left\{ E \left[ \left\| \mathbf{f} - \hat{\mathbf{f}} \right\|^2 \right] \right\}.$$

Thus, the estimated image is given by:

$$\hat{\mathbf{f}} = \mathbf{R}_{ff} \mathbf{H}^T \left( \mathbf{H}^T \mathbf{R}_{ff} \mathbf{H} + \mathbf{R}_{nn} \right)^{-1} \mathbf{g}. \quad (1.5)$$

In this case the inversion is better conditioned, and the ill-posed problem is avoided. This can be better seen in the case where the noise has simple covariance and it is circulant:

$$\mathbf{R}_{nn} = \sigma^2 \mathbf{I} = \mathbf{W}^{-1} \sigma^2 \mathbf{I} \mathbf{W}.$$

Thus, the power spectrum  $\mathbf{S}_{nn}$  is easily computed:

$$\mathbf{S}_{nn}(i) = \sigma^2, \quad i = 1, \dots, N.$$

If the process is assumed zero mean, the covariance equals to the correlation matrix. Also, if the covariance matrix is assumed circulant, the equivalent of Eq. (1.5) in the DFT domain is expressed as:

$$\hat{\mathbf{F}}(i) = \frac{\mathbf{S}_{ff}(i) \mathbf{\Lambda}_h^*(i, i)}{\left\| \mathbf{\Lambda}_h(i, i) \right\|^2 \mathbf{S}_{ff}(i) + \sigma^2} \mathbf{G}(i), \quad i = 1, \dots, N, \quad (1.6)$$

where  $S_{ff}(i)$  is the power spectrum of the image [1]. The quantity that multiplies the observed image  $\mathbf{G}(i)$  to obtain the restored  $\hat{\mathbf{F}}(i)$  is the inverse filter. This filter does not amplify the noise because there are small values in the high frequencies. In this way, a regularized solution is obtained.

This convenient formulation in the DFT domain is possible because the image and the noise covariances were assumed circulant, or in other words, they represent stationary processes. The main benefit of a stationary model is the ease of solution in the DFT domain. Unfortunately, these models usually cannot describe efficiently the real world phenomena.

### 1.4.2 Deterministic regularization

At this point we present how deterministic regularization can be used to solve ill-posed problems. In deterministic approaches the criterion used to find the restored image is the minimization of the Euclidian norm:

$$\|\mathbf{g} - \mathbf{H}\mathbf{f}\|_2^2.$$

However, this alone yields the pseudo-inverse solution

$$\hat{\mathbf{f}} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{g},$$

which is ill-posed [127]. To ameliorate this situation, Tikhonov regularization can be used with an addition of a penalty (regularization) term [127]:

$$\min_{\mathbf{f}} \left\{ \|\mathbf{g} - \mathbf{H}\mathbf{f}\|_2^2 + a \|\mathbf{Q}\mathbf{f}\|_2^2 \right\}. \quad (1.7)$$

The regularization parameter  $a$  is a scalar and  $\mathbf{Q}$  is a  $N \times N$  matrix. The penalty term tries to bias the solution obtained by minimizing the first norm, towards to a different solution, which is constrained to have some properties common in images. In most cases this property is that the image must have small energy at high-frequencies and thus be rather smooth [59]. The Laplacian operator could be used as the regularization matrix  $\mathbf{Q}$ . Using one-dimensional notation, it is given by

$$\mathbf{Q} = \begin{bmatrix} 2 & -1 & 0 & \cdots & -1 \\ -1 & 2 & -1 & \cdots & 0 \\ 0 & -1 & 2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -1 & 0 & \cdots & -1 & 2 \end{bmatrix}.$$

Application of this operator to each pixel yields:

$$\mathbf{Q}\mathbf{f}[i] = \mathbf{f}'[i] = -\mathbf{f}[i-1] + 2\mathbf{f}[i] - \mathbf{f}[i+1], \quad i = 1, \dots, N. \quad (1.8)$$

Circularity implies that:  $\mathbf{f}[-1] = \mathbf{f}[N]$ ,  $\mathbf{f}[N+1] = \mathbf{f}[1]$  and it is assumed for the same reasons as in the case of the PSF operator  $\mathbf{H}$  and the covariance matrices  $\mathbf{R}_{ff}$  and  $\mathbf{R}_{nn}$ . If  $\mathbf{Q}$  is viewed as a filter, its frequency response is  $Q(w) = 2(1 - \cos w)$ , which is clearly a high-pass filter.

Bounding the term  $\mathbf{Qf}[i]$  given by Eq. (1.8) to take small values, imposes that neighboring pixels of the image must have similar values. From a mathematical point of view, the Laplacian operator is the discrete analogous of the second derivative operation. Thus, introducing this derivative as a penalty term automatically constrains the restored image to have bounded discontinuities. The solution of the minimization problem of Eq. (1.7) is obtained as:

$$\hat{\mathbf{f}} = (\mathbf{H}^T \mathbf{H} + a \mathbf{Q}^T \mathbf{Q})^{-1} \mathbf{H}^T \mathbf{g}. \quad (1.9)$$

Due to the properties of circulant matrices, Eq. (1.9) can be written in the DFT domain as:

$$\hat{\mathbf{F}}(i) = \frac{\Lambda_{\mathbf{h}}^*(i, i)}{|\Lambda_{\mathbf{h}}(i, i)|^2 + a |\Lambda_{\mathbf{Q}}(i, i)|^2} \mathbf{G}(i) = \frac{|\Lambda_{\mathbf{Q}}(i, i)|^{-2} \Lambda_{\mathbf{h}}^*(i, i)}{|\Lambda_{\mathbf{h}}(i, i)|^2 |\Lambda_{\mathbf{Q}}(i, i)|^{-2} + a} \mathbf{G}(i), \quad i = 1, \dots, N.$$

It is very interesting to notice that Eq. (1.6) is identical to the above equation if the inverse image covariance equals to  $\mathbf{Q}^T \mathbf{Q}$ , or setting the inverse of the image power spectrum equal to the Laplacian operator's inverse eigenvalues:

$$\mathbf{S}_{\text{ff}}(i) = \Lambda_{\mathbf{c}}(i, i)^{-1}, \quad i = 1, \dots, N.$$

From this relation the close connection between stochastic and deterministic approaches becomes obvious. More specifically, the deterministic method can be derived from a corresponding stochastic, by incorporating a prior for the image, based on the *simultaneously auto-regressive* (SAR) prediction model [94]:

$$p(\mathbf{f}) = Z_1^{-1} \exp \left\{ -\frac{1}{2} a \|\mathbf{Qf}\|_2^2 \right\},$$

and following the *maximum a posteriori* (MAP) approach. Due to the additive Gaussian noise in the degradation model, the conditional distribution of the observed image is:

$$p(\mathbf{g} | \mathbf{f}) = Z_2^{-1} \exp \left\{ -\frac{1}{2\sigma^2} \|\mathbf{g} - \mathbf{Hf}\|_2^2 \right\}.$$

In the above equations,  $Z_1$  and  $Z_2$  are normalizing constants. According to Bayes' rule the posterior distribution is:

$$p(\mathbf{f} | \mathbf{g}) = \frac{p(\mathbf{g} | \mathbf{f})p(\mathbf{f})}{p(\mathbf{g})},$$

where  $p(\mathbf{g})$  is the marginal distribution of  $\mathbf{g}$ . It is very convenient to estimate the image by the mode of this density and obtain the maximum a posteriori (MAP) estimation:

$$\hat{\mathbf{f}}_{MAP} = \arg \max_{\mathbf{f}} p(\mathbf{f} | \mathbf{g}) = \arg \max_{\mathbf{f}} p(\mathbf{g} | \mathbf{f})p(\mathbf{f}).$$

This is equivalent to minimizing the negative log-likelihood of the posterior:

$$\hat{\mathbf{f}}_{MAP} = \arg \min_{\mathbf{f}} \{-\log p(\mathbf{g} | \mathbf{f})p(\mathbf{f})\} = \arg \min_{\mathbf{f}} \left\{ \frac{1}{\sigma^2} \|\mathbf{g} - \mathbf{H}\mathbf{f}\|_2^2 + a \|\mathbf{Q}\mathbf{f}\|_2^2 \right\}.$$

The variance parameters can be merged to one, since this has no effect to the solution:

$$\hat{\mathbf{f}}_{MAP} = \arg \min_{\mathbf{f}} \left\{ \|\mathbf{g} - \mathbf{H}\mathbf{f}\|_2^2 + a' \|\mathbf{Q}\mathbf{f}\|_2^2 \right\}, \quad (1.10)$$

where  $a' = a\sigma^2$ . This equation is identical to Eq. (1.7). This demonstrates that regularization in a stochastic framework is achieved with the introduction of an image prior.

#### 1.4.3 Estimation of the regularization parameter

Various ways for estimation of the regularization parameter  $a$  have been proposed in [59]. The simultaneously auto-regressive (SAR) model provides a way to obtain a regularized solution for the restoration problem [59], [94], [58] and [92]. Using this model the image and the regularization parameter can be estimated in the maximum likelihood framework utilizing the *expectation-maximization* EM algorithm, described in Chapter 2. In this way, the parameter is estimated in a rigorous manner.

As it has been shown above, independently of the use of a deterministic or stochastic method, the parameter  $a$  controls the tradeoff between the data fidelity term  $\|\mathbf{g} - \mathbf{H}\mathbf{f}\|_2^2$  and the regularization term  $\|\mathbf{Q}\mathbf{f}\|_2^2$ . Thus, large values of  $a$  give over-regularized images with over-smoothed edges, and small values of  $a$  leads to under-regularized solutions with amplified noise, especially in flat areas. This is illustrated in Figure 1.2, where in Figure 1.2.(c) the over-regularized restored image is shown for large  $a$  and in Figure 1.2.(d) the under-regularized restored image with small  $a$ . It is obvious that in the over-regularized image, the



noise has been suppressed, but unfortunately the same holds for the edges. In the under-regularized case the edges has been preserved but with the disadvantage of noise amplification in the smooth areas. The best value of  $a$  lies between these two extremes. In Figure 1.2.(e) the restored image with average  $a$  is shown. In view of this, it would be desirable for a restoration filter to be spatially adaptive: more regularization in smooth areas of the image and less at the edges. One of the goals of this dissertation is to develop spatially adaptive regularization algorithms.

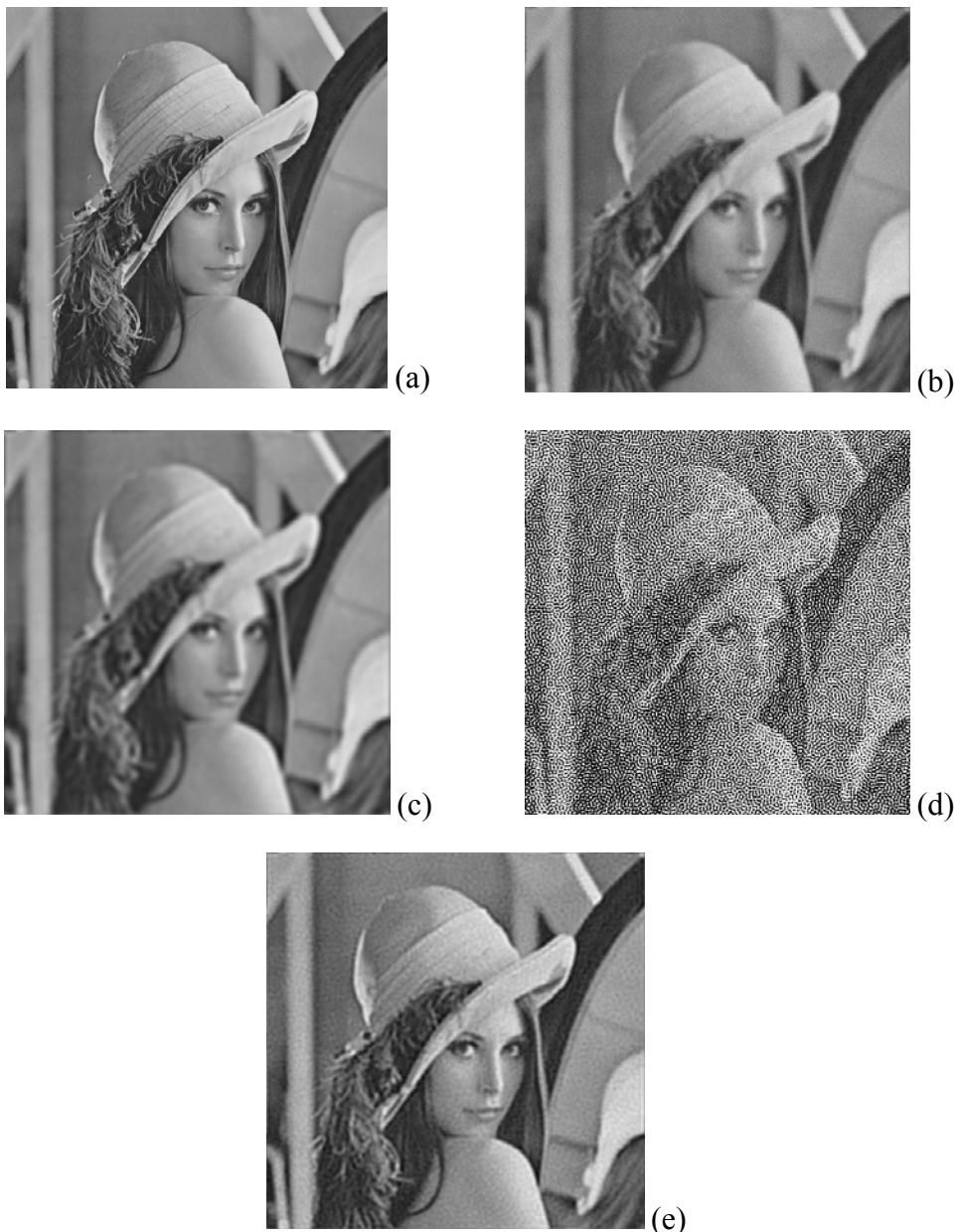


Figure 1.2 (a) ‘Lena’ original image, (b) degraded image, (c) over-regularized restored image, (d) under-regularized restored image and (e) restored image with average regularization.

## 1.5 Super-resolution: reconstructing a high-resolution image from low-resolution images

Super-resolution is the problem of reconstructing a high-resolution image from low-resolution images of the same scene ([105], [48] and [32]). The key idea of super-resolution is to fuse in one image with higher resolution the information included in the multiple low-resolution images. For that, it is required that sub-pixel motion exists between the low-resolution images so as to avoid a redundancy of the information contained in them [67]. The motion can be translation as well as rotation. The imaging model of super-resolution incorporated in this dissertation assumes that the low-resolution images are translated and rotated under-sampled versions of a high resolution image. The problem formulation of super-resolution in this dissertation is similar to that of image restoration [105], in the sense that the imaging model is again linear and similar to that of Eq. (1.2). To present the model we denote as  $d$  the integer decimation factor. In other words, the imaging model assumes a high resolution image of size  $N_H \times 1$ , where  $N_H = dN$  and  $N$  is the size of the low-resolution images. Hence,  $P$  low-resolution images of size  $N \times 1$  are produced by applying the  $PN \times N_H$  degradation operator  $\mathbf{B}$  to the high-resolution image. Then, white Gaussian noise is added at each observation. Let  $\mathbf{y}$  be a  $PN \times 1$  vector, containing the  $P$  low resolution images  $\mathbf{y}_i$ :

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1^T & \mathbf{y}_2^T & \cdots & \mathbf{y}_P^T \end{bmatrix}^T,$$

where  $\mathbf{y}_i$  is a  $N \times 1$  vector, representing a low-resolution image. Using this notation, the observations are given by:

$$\mathbf{y}_i = \mathbf{B}_i \mathbf{x} + \mathbf{n}_i = \mathbf{D} \mathbf{H}_i \mathbf{S}(\delta_i) \mathbf{R}(\theta_i) \mathbf{x} + \mathbf{n}_i,$$

or in one vector

$$\mathbf{y} = \mathbf{B} \mathbf{x} + \mathbf{n},$$

where  $\mathbf{x}$  the (unknown) original  $N_H \times 1$  high-resolution image to be estimated,  $\mathbf{B}$  is the  $PN \times N_H$  degradation matrix and  $\mathbf{n} = \begin{bmatrix} \mathbf{n}_1^T & \mathbf{n}_2^T & \cdots & \mathbf{n}_P^T \end{bmatrix}^T$  the  $PN \times 1$  vector consisting of  $P$   $N \times 1$  additive white noise vectors. This equation has the physical meaning that a low-resolution image is produced by rotating, translating, blurring and decimating the original image  $\mathbf{x}$ . The degradation operator  $\mathbf{B}$  is given by:

$$\mathbf{B} = [\mathbf{B}_1^T \quad \cdots \quad \mathbf{B}_P^T]^T,$$

where  $\mathbf{B}_i = \mathbf{D}\mathbf{H}_i\mathbf{S}(\delta_i)\mathbf{R}(\theta_i)$  for  $i=1,\dots,P$ . The matrix  $\mathbf{D}$  is the known  $N \times N_H$  decimation matrix.  $\mathbf{H}_i$ ,  $i=1\dots P$ , are the shift-invariant  $N_H \times N_H$  blurring convolutional operators, and  $\mathbf{S}(\delta_i)$ , for  $i=1,\dots,P$ , are the  $N_H \times N_H$  shift-invariant shifting operators. Each  $\delta_i$  is a scalar which represents translation. The shift-invariant operators  $\mathbf{S}(\delta_i)$  and  $\mathbf{H}_i$  can be assumed circulant. As in the restoration problem, this is very useful for computational purposes because such matrices can be easily diagonalized in the DFT domain. Finally, the  $N_H \times N_H$  matrix  $\mathbf{R}(\theta_i)$  represents the rotation with angle  $\theta_i$  of each observation relative to the unknown high-resolution image  $\mathbf{x}$ .

Super-resolution is an ill-posed inverse problem [48], in the sense that solution of

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \|\mathbf{y} - \mathbf{B}\mathbf{x}\|_2^2 \quad (1.11)$$

leads to a linear system with ill an conditioned matrix  $\mathbf{B}$ . Inversion of this matrix amplifies the noise in the data, something that happens also in the image restoration. The problem is ill-posed because of the blurring matrices  $\mathbf{H}_i$  and the decimation matrix  $\mathbf{D}$ . The effect of  $\mathbf{H}_i$  is discussed in the previous section. The decimation matrix in the frequency domain is a low-pass ‘sub-folding’ operator [32]. Thus, inverting both matrices leads to high-pass filtering that amplifies the noise.

As in image restoration, the problem of super-resolution can be converted to a well-posed. This is achieved by constraining the solution of Eq. (1.11) by adding a penalty term, for example the Tikhonov regularization term [127]:

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \left( \|\mathbf{y} - \mathbf{B}\mathbf{x}\|_2^2 + a \|\mathbf{Q}\mathbf{x}\|_2^2 \right).$$

where  $\mathbf{Q}$  is the high-pass Laplacian operator. This method can be viewed from a Bayesian perspective as assuming a SAR image prior [133]. Of course, there are other methodologies applying different regularization criteria. The result of this algorithm is shown in Figure 1.3. This makes super-resolution and image restoration related problems, since the regularization criteria are applicable to both. For that reason, in this dissertation the two problems are considered in parallel since the image priors proposed in the literature for the restoration

problem are also applied to super-resolution. However, the difficulty of estimating the registration parameters in super-resolution model makes the latter a more challenging task.

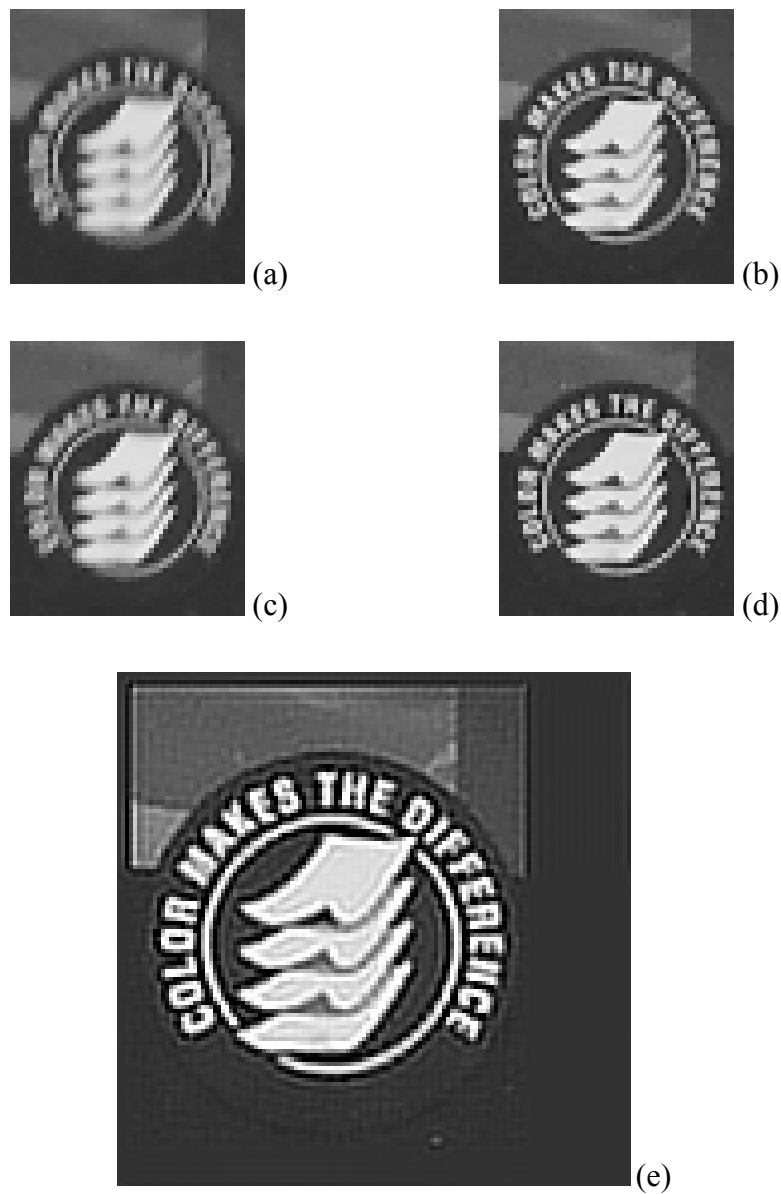


Figure 1.3 (a)-(d) Low resolution observations, (e) 2x super-resolved image regularized under the SAR image model [133].

## 1.6 Thesis contribution

The image restoration and super-resolution problems can be treated in the Bayesian framework, in order to convert both ill-posed inverse problems to well-posed. In a stochastic framework, regularization of the solution is achieved by defining an image prior  $p(\mathbf{f}) \propto \exp\{-aC(\mathbf{f})\}$  and taking the *maximum a posteriori* (MAP) solution using Bayes' rule:

$$\hat{\mathbf{f}} = \arg \min_{\mathbf{f}} \{-\log p(\mathbf{f} | \mathbf{g})\} = \arg \min_{\mathbf{f}} \{-\log p(\mathbf{g} | \mathbf{f}) p(\mathbf{f})\} = \arg \min_{\mathbf{f}} \{\beta \|\mathbf{g} - \mathbf{H}\mathbf{f}\|_2^2 + aC(\mathbf{f})\},$$

where  $C(\mathbf{f})$  is a function that works as a *constraint* to the solution. From a Bayesian perspective, using  $p(\mathbf{f})$  we incorporate prior knowledge for the image and we constrain the solution to follow some specific properties common in real word images, such as large smooth areas and few edges. Consequently, a realistic image prior should be non-stationary in order to adapt to the image local characteristics; for example, large variances at edges and small at smooth areas. In regularization terms, this means that this type of prior leads to spatially adaptive regularization.

The contribution of this dissertation is the development of stochastic non-stationary image models as image priors  $p(\mathbf{f})$  needed for regularization in the restoration and super-resolution problems. The proposed non-stationary image models lead to spatially adaptive regularization, or in other words, non-uniform regularization along the image which depends on the local spatial activity. Furthermore, working in the stochastic framework we employ the Bayesian methodology to solve these inverse problems and in parallel to estimate the model parameters. Thus, development of Bayesian restoration and reconstruction algorithms is another contribution of this dissertation. It must be noted here that, according to the Bayesian methodology, the image is inferred, which means that a distribution is obtained for it and not just a single point estimate as in the MAP approach. This implies that the image is treated as a random variable rather than a parameter.

In Chapter 2 we present the variational Bayesian (VB) methodology for *approximate* inference that is used in this dissertation. Inference means computing the posterior distribution of the model's *hidden* random variables. To make that clear, the term 'hidden' is

analyzed extensively. Also, the *expectation-maximization* (EM) algorithm is presented as a special case of the variational methodology, where the inference is exact.

In Chapter 3 we provide a survey of state-of-the-art spatially adaptive restoration methods and pinpoint some properties and similarities between them. More specifically, it is pointed out that the EM algorithm is a special case of the *majorization-minimization* (MM) class of optimization algorithms, and this explains the fact that stochastic and deterministic methods that employ the EM and the MM algorithms respectively, happen to provide very similar restoration algorithms. To demonstrate this connection, example restoration algorithms are given. Half-quadratic regularization is also presented, which also coincides, under some specific circumstances, with the EM and MM algorithms.

In Chapter 4 we introduce a new hierarchical (two-level) Gaussian non-stationary image prior, as a natural extension of the SAR model presented in section 1.4. This prior assumes that the residuals of the first order differences of the image, in four *different directions*, are Gaussian random variables with zero mean and variance that is *spatially varying*. As a result, these local directional variances capture the image discontinuities with a continuous value model. In this way, the variances manifest the spatial adaptivity mechanism. In order to deal with the resulting over-parameterization of this model, the spatially varying variances are considered as random variables (not parameters) and a Gamma hyper-prior is imposed on them, which is *conjugate* to the Gaussian. The parameters of the imposed hyper-prior control the *degree of non-stationarity* of the imposed image prior. To learn this model and infer the image we propose two iterative algorithms. The first is based on the maximum a posteriori estimation (MAP) principle and computes explicitly both the image and the spatially varying variances in all four directions. The second is a Bayesian algorithm that marginalizes the “hidden variables”. It is interesting that the resulting MAP algorithm is similar with the algorithms proposed in [80], [84], and [78] which are based on heuristic arguments. Also, the marginalization of the hidden variables produces a Student’s-t distribution. Thus, the proposed restoration algorithms are Student’s-t based. Finally, numerical experiments are presented that demonstrate the superiority of the proposed algorithms with respect to other state-of-the-art methods [27], [26], [28].

In Chapter 5 we propose a new Bayesian inference framework for image restoration using a prior in product form [24]. This prior assumes that the outputs of local high-pass filters, (their number is arbitrary), follow the Student’s-t distribution. This means that they are distributed

according to the hierarchical prior of Chapter 4. The main contribution of this chapter is a Bayesian inference methodology that bypasses the difficulty of evaluating the normalization constant of product type priors. The methodology is based on a *constrained* variational approximation that uses the outputs of all the local high-pass filters to produce an estimate of the original image. More specifically, a *constrained expectation step* is used to capture the relationship of the filter outputs of the prior to the original image. In this manner the use of improper priors is avoided and *all* the parameters of the prior model are estimated from the data. Thus, the “trial and error” parameter “tweaking” required in state-of-the-art restoration algorithms (including that of Chapter 4), which makes their use difficult use for non-experts, is avoided. Furthermore, numerical experiments show that the proposed restoration algorithm provides competitive performance compared with previous methods. In this chapter we also propose an efficient Lanczos-based computational framework tailored to the calculations required in our Bayesian algorithm. More specifically, a very large linear system  $Ax = b$  is solved iteratively and the diagonal elements of a matrix  $Q^T A^{-1} Q$  are simultaneously estimated in an efficient manner [24], [25].

In Chapter 6 we extend the TV prior and the related work to estimate the regularization parameter in [8] by introducing a new prior which has a number of novel features. The extension of the TV prior is performed in analogous way with that used to derive the non-stationary prior of Chapters 4 and 5. More specifically, we introduce a TV prior with spatially varying regularization parameters. In order to avoid the over parameterization due to the spatially varying nature of this prior, we introduce a Gamma hyperprior for the spatially adaptive regularization parameters of the local TV priors. Also, we use this prior in a product form, which means that we assume that the outputs of an arbitrary number of high-pass filters are distributed according to this prior. This gives two novel features to the new prior. First, it is explicitly spatially adaptive and thus it is better suited to capture the salient features of the image. Second, it is in product form and has the ability to enforce simultaneously a number of different properties to the image. This prior can use arbitrary linear operators, not just first order differences as TV. Thus, a prior similar to the one in [8] with an exactly calculated partition function is just a special case of it. If the hidden variables of the second layer are marginalized, the resulting density function has a form similar to a Student's-t distribution; thus, we label it as Generalized Student's-t. Due to the complexity of this model, we resort to the variational approximation for Bayesian inference. However, we use two modifications.

First, we derive a quadratic bound to the variational bound, in a manner similar to the methodology used in [8], to bypass the difficulties due to the non-quadratic form of the new prior. Second, we use the constrained variational framework in a manner similar to that proposed in Chapter 5 in order to bypass the problem of computing the partition function of the new prior. In this chapter we also suggest an iterative numerical method to compute the diagonal elements of very large inverse matrices that are necessary to apply the proposed Bayesian algorithm. This numerical method is similar in spirit to the one employed in Chapter 5. However, it is based on conjugate-gradients and not on the Lanczos methodology, and was found empirically to converge faster in this application.

In Chapter 7 we present our contribution to the super-resolution problem. We utilize for the first time in the super-resolution problem the hierarchical two-level image prior presented in Chapters 4 and 5. Using this non-stationary prior, it is possible to reconstruct high-resolution images without smoothed edges or ringing artifacts in the vicinity of edges. Another contribution to the super-resolution problem is a novel two-step reconstruction algorithm. The first stage of this algorithm is a preprocessing step that approximately registers the degraded low-resolution observations. These “almost-registered” low-resolution observations are used subsequently by an iterative algorithm which simultaneously reconstructs the high-resolution images and finds their registration parameters. We propose this sub-optimal two-stage approach in order to speed up the super-resolution algorithm. Thus, the MAP functional is maximized based on coarse estimation of rotation and translation between image pairs. Furthermore, the registration sub-task is based on the Newton-Raphson (NR) algorithm that utilizes analytically calculated first and second order derivatives and exhibits quadratic convergence. Lastly, experiments with both real and synthetic data are conducted that demonstrate the efficacy of the proposed algorithm [29], [30].



## CHAPTER 2. APPROXIMATE BAYESIAN INFERENCE

---

- 2.1. Expectation Maximization (EM) algorithm: a variational point of view
  - 2.2. Approximate Bayesian inference with variational bound maximization
  - 2.3. Constrained variational optimization
  - 2.4. Bayesian inference for restoration with the SAR prior model
- 

In this chapter we present the Bayesian inference methodology used to learn stochastic models. In statistical modeling we assume that the observed data (signals, images, etc.) are produced through a generation process that is modeled using statistical models. Such models are used in image restoration when the problem is formulated in a stochastic framework and the image model is included in the form of a prior distribution  $p(\mathbf{f})$ . Furthermore, another stochastic model, represented by the distribution  $p(\mathbf{g}|\mathbf{f})$ , is assumed for the imaging model that produces the observed degraded image  $\mathbf{g}$ . Thus, the stochastic model  $p(\mathbf{g}, \mathbf{f}) = p(\mathbf{g}|\mathbf{f})p(\mathbf{f})$  of the restoration problem is hierarchical and contains the imaging and the image model. Based on this joint stochastic model, the image  $\mathbf{f}$  is treated as random variable and Bayesian algorithms can be employed to infer it, i.e. to obtain a posterior  $p(\mathbf{f}|\mathbf{g})$ . In addition, estimation of the parameters of the image model, as well as of the imaging model, is simultaneously performed.

Stochastic models include parameters that should be estimated in order to *fit* the model to the observations. A very popular method to estimate these parameters is the *maximum likelihood* (ML) approach [129], defined by:

$$\hat{\boldsymbol{\theta}}_{ML} = \arg \max_{\boldsymbol{\theta}} p(\mathbf{y}; \boldsymbol{\theta}), \quad (3.1)$$

where  $\mathbf{y}$  is the vector of the observations (data generated according to the stochastic model mechanism),  $\boldsymbol{\theta}$  is the vector consisting of the parameters and  $p(\mathbf{y}; \boldsymbol{\theta})$  is the likelihood of the

data. It must be noted that  $p(\mathbf{y}; \boldsymbol{\theta})$  is different from  $p(\mathbf{y} | \boldsymbol{\theta})$ . In the first case,  $\boldsymbol{\theta}$  are just the parameters of the likelihood, while in the latter case  $\boldsymbol{\theta}$  are assumed random variables and  $\mathbf{y}$  is conditioned on them.

## 2.1 Expectation Maximization (EM) algorithm: a variational point of view

A very popular and successful algorithm to obtain the ML solution is the *expectation-maximization* (EM) algorithm, which maximizes the likelihood iteratively by taking advantage of the *hidden* variables in the model, see [99] and [41]. More specifically, the values of the observed random variables  $\mathbf{y}$  are the outcome of a stochastic process. Usually this process depends on the outcome of another preceding process that is not observed (hidden). The outcome of this ‘hidden’ process is represented by random variables that are called hidden. Thus, in this case, the generation of  $\mathbf{y}$  contains two levels/steps and it is described by the joint distribution (complete likelihood) of the hidden  $\mathbf{x}$  and the observed variables  $\mathbf{y}$ :

$$p(\mathbf{y}, \mathbf{x}) = p(\mathbf{y} | \mathbf{x})p(\mathbf{x}),$$

where  $p(\mathbf{x})$  is the prior distribution on the hidden variables  $\mathbf{x}$  and  $p(\mathbf{y} | \mathbf{x})$  is the conditional distribution of the observations given  $\mathbf{x}$ . Note that these distribution may contain parameters that have been omitted for brevity. The above equation shows that the data generation process is hierarchical and contains two steps. In the first step, the variables  $\mathbf{x}$  take values drawn from the distribution  $p(\mathbf{x})$ . Then in the second step, using  $\mathbf{x}$ , values for  $\mathbf{y}$  are drawn from the conditional distribution  $p(\mathbf{y} | \mathbf{x})$ .

Herein, we present the EM algorithm using its variational interpretation [10]. The EM algorithm exploits the existence of hidden variables and the difficult problem of direct optimization of  $L(\boldsymbol{\theta}) = \log p(\mathbf{y}; \boldsymbol{\theta})$  is transformed to convenient successive optimizations of a lower bound  $F(q(\mathbf{x}), \boldsymbol{\theta})$ , which is defined by:

$$F(q(\mathbf{x}), \boldsymbol{\theta}) = L(\boldsymbol{\theta}) - KL(q(\mathbf{x}) || p(\mathbf{x} | \mathbf{y}; \boldsymbol{\theta})) \leq L(\boldsymbol{\theta}),$$

where

$$KL(q(\mathbf{x}) \parallel p(\mathbf{x} | \mathbf{y}; \boldsymbol{\theta})) = \int q(\mathbf{x}) \log \frac{p(\mathbf{x} | \mathbf{y}; \boldsymbol{\theta})}{q(\mathbf{x}; \boldsymbol{\theta})} d\mathbf{x}$$

is the Kullback-Leibler (KL) divergence between an arbitrary function  $q(\mathbf{x})$  and the posterior of the hidden  $p(\mathbf{x} | \mathbf{y}; \boldsymbol{\theta})$ . Inequality holds because KL is always non-negative, and inequality becomes an equality when  $q(\mathbf{x}) = p(\mathbf{x} | \mathbf{y}; \boldsymbol{\theta})$ , i.e. when KL becomes zero. The bound  $F(q(\mathbf{x}), \boldsymbol{\theta})$  is a functional of the function  $q(\mathbf{x})$  and the parameters  $\boldsymbol{\theta}$ .

The EM algorithm iterates between two steps, the expectation step (E-step) and the maximization step (M-step). In the E-step, maximization of  $F$  is performed with respect to  $q(\mathbf{x})$ , and in the M-step with respect to  $\boldsymbol{\theta}$ . The bound is maximized with respect to  $q(\mathbf{x})$ , when  $q(\mathbf{x}) = p(\mathbf{x} | \mathbf{y}; \boldsymbol{\theta})$  and the bound becomes tight. Thus, the E-step and M-step at iteration  $(t+1)$  are

$$\text{E-step: } q^{(t+1)}(\mathbf{x}) = p(\mathbf{x} | \mathbf{y}; \boldsymbol{\theta}^{(t)}), \quad (2.1)$$

$$\text{M-step: } \boldsymbol{\theta}^{(t+1)} = \arg \max_{\boldsymbol{\theta}} F(q^{(t+1)}(\mathbf{x}), \boldsymbol{\theta}). \quad (2.2)$$

In the E-step, a posterior distribution is inferred for  $\mathbf{x}$ , using the model parameters calculated from the previous iteration  $\boldsymbol{\theta}^{(t)}$ . At this point we must note that in the Bayesian inference methodology the hidden random variables are *inferred* and the parameters are *estimated*. In the M-step, the parameters  $\boldsymbol{\theta}$  are updated by maximizing  $F(q^{(t+1)}(\mathbf{x}), \boldsymbol{\theta})$ , which can be written as:

$$F(q^{(t+1)}(\mathbf{x}), \boldsymbol{\theta}) = \int q^{(t+1)}(\mathbf{x}) \log p(\mathbf{x} | \mathbf{y}; \boldsymbol{\theta}) d\mathbf{x} - \int q^{(t+1)}(\mathbf{x}) \log q^{(t+1)}(\mathbf{x}) d\mathbf{x}.$$

Using the above equation the M-Step of Eq. (2.2) can be written also as such:

$$\text{M-Step: } \boldsymbol{\theta}^{(t+1)} = \arg \max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(t)}),$$

where  $Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(t)}) = \langle \log p(\mathbf{y}, \mathbf{x}; \boldsymbol{\theta}) \rangle_{p(\mathbf{x} | \mathbf{y}; \boldsymbol{\theta}^{(t)})}$  is the expectation of the logarithmic complete likelihood with respect to the posterior of the hidden given the observations. This posterior is given by the Bayes' rule

$$p(\mathbf{x} | \mathbf{y}; \boldsymbol{\theta}^{(t)}) = \frac{p(\mathbf{y}, \mathbf{x}; \boldsymbol{\theta}^{(t)})}{p(\mathbf{y}; \boldsymbol{\theta}^{(t)})} = \frac{p(\mathbf{y} | \mathbf{x}; \boldsymbol{\theta}^{(t)})p(\mathbf{x}; \boldsymbol{\theta}^{(t)})}{p(\mathbf{y}; \boldsymbol{\theta}^{(t)})}. \quad (2.3)$$

This equation shows that, in order to apply EM, it is necessary to know explicitly the form of  $p(\mathbf{y}; \boldsymbol{\theta})$ , by integrating out the hidden variables. If the function  $Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(t)})$  is considered as a minorizer of the likelihood, since

$$Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(t)}) \leq L(\boldsymbol{\theta}), \text{ and } Q(\boldsymbol{\theta}^{(t)}; \boldsymbol{\theta}^{(t)}) = L(\boldsymbol{\theta}^{(t)}),$$

then, the EM algorithm can be considered to belong to the general class of the majorization-minimization (MM) algorithms [85]. Hence, the iterative maximization of the minorizer converges to a maximum of the likelihood. In most applications, the ultimate goal is to infer the hidden variables.

Note that the parameters in the posterior are considered fixed when the function  $Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(t)})$  is maximized in the M-step. Thus, the EM algorithm maximizes the marginal likelihood by incorporating the hidden variables in the iterative scheme, although they are not explicitly apparent in the likelihood. This leads to a more convenient optimization scheme than the direct optimization of the likelihood, and also to inference of the hidden variables. A reasonable estimate of the hidden variables may be the mode of the posterior probability:

$$\hat{\mathbf{x}} = \arg \max_{\mathbf{x}} p(\mathbf{x} | \mathbf{y}; \boldsymbol{\theta}). \quad (2.4)$$

## 2.2 Approximate Bayesian inference with variational bound maximization

In many cases, the likelihood

$$p(\mathbf{y}; \boldsymbol{\theta}) = \int p(\mathbf{y}, \mathbf{x}; \boldsymbol{\theta}) d\mathbf{x},$$

is not directly known, because the integral is intractable. Without the explicit form of the marginal likelihood the E-step (Eq. (2.2)) of the EM algorithm is not applicable [129]. However, there are cases where the integral can be calculated analytically with respect to a subset of the hidden variables. For example, the set of the hidden variables  $\mathbf{x} = \{\mathbf{x}_1, \mathbf{x}_2\}$  consists of the subsets  $\mathbf{x}_1$  and  $\mathbf{x}_2$  that are jointly distributed,

$$p(\mathbf{y}, \mathbf{x}) = p(\mathbf{y}, \mathbf{x}_1, \mathbf{x}_2),$$

but the double integral

$$p(\mathbf{y}; \boldsymbol{\theta}) = \int \int p(\mathbf{y}, \mathbf{x}_1, \mathbf{x}_2) d\mathbf{x}_1 d\mathbf{x}_2,$$

can be only evaluated with respect to  $\mathbf{x}_1$  or  $\mathbf{x}_2$ , but not both simultaneously.

A number of approximations have been used to compute this integral, see for example [90], [99] and [10]. Herein, we focus on the variational methodology, according to which the marginal likelihood can be approximately maximized, by maximizing iteratively an approximate bound instead of the exact bound  $F$  used in the EM algorithm. This can be achieved by altering the EM algorithm, as described in the previous section, with the assumption that the inferred posteriors are independent:

$$q(\mathbf{x}_1, \mathbf{x}_2) = q(\mathbf{x}_1)q(\mathbf{x}_2).$$

This is the mean-field approximation [10] and makes the bound approximate:

$$F(q(\mathbf{x}_1), q(\mathbf{x}_2), \boldsymbol{\theta}) = \log p(\mathbf{y} | \hat{\mathbf{x}}_1; \boldsymbol{\theta}) p(\hat{\mathbf{x}}_1; \boldsymbol{\theta}) + KL(q(\mathbf{x}_1)q(\mathbf{x}_2) \| p(\mathbf{x}_1, \mathbf{x}_2 | \mathbf{y}; \boldsymbol{\theta})),$$

$$F(q(\mathbf{x}_1), q(\mathbf{x}_2), \boldsymbol{\theta}) \approx F(q(\mathbf{x}_1, \mathbf{x}_2), \boldsymbol{\theta}).$$

The variational algorithm contains two steps. In the first step, the VE-step, the bound is maximized with respect to  $q(\mathbf{x}_1)$  and  $q(\mathbf{x}_2)$ ,

$$[q^{(t+1)}(\mathbf{x}_1), q^{(t+1)}(\mathbf{x}_2)] = \arg \max_{q(\mathbf{x}_1), q(\mathbf{x}_2)} F(q(\mathbf{x}_1), q(\mathbf{x}_2), \boldsymbol{\theta}^{(t)}).$$

In the VM-step, the bound is maximized with respect to the parameters:

$$\boldsymbol{\theta}^{(t+1)} = \arg \max_{\boldsymbol{\theta}} F(q^{(t+1)}(\mathbf{x}_1), q^{(t+1)}(\mathbf{x}_2), \boldsymbol{\theta}).$$

The closed form solutions for the posteriors in the VE-step are:

$$q^{(t+1)}(\mathbf{x}_2) = \frac{\exp\left\{-\langle \log p(\mathbf{y}, \mathbf{x}_1, \mathbf{x}_2; \boldsymbol{\theta}) \rangle_{q^{(t)}(\mathbf{x}_2)}\right\}}{\int \exp\left\{-\langle \log p(\mathbf{y}, \mathbf{x}_1, \mathbf{x}_2; \boldsymbol{\theta}) \rangle_{q^{(t)}(\mathbf{x}_2)}\right\} d\mathbf{x}_1},$$

and

$$q^{(t+1)}(\mathbf{x}_1) = \frac{\exp\left\{-\langle \log p(\mathbf{y}, \mathbf{x}_1, \mathbf{x}_2; \boldsymbol{\theta}) \rangle_{q^{(t)}(\mathbf{x}_1)}\right\}}{\int \exp\left\{-\langle \log p(\mathbf{y}, \mathbf{x}_1, \mathbf{x}_2; \boldsymbol{\theta}) \rangle_{q^{(t)}(\mathbf{x}_1)}\right\} d\mathbf{x}_2}.$$

where  $\langle \cdot \rangle_q$  denotes the expectation with respect to an arbitrary distribution  $q$ . The VM-step can be written in more detail:

$$\boldsymbol{\theta}^{(t+1)} = \arg \max_{\boldsymbol{\theta}} \langle \log p(\mathbf{y}, \mathbf{x}_1, \mathbf{x}_2; \boldsymbol{\theta}) \rangle_{q^{(t)}(\mathbf{x}_1)q^{(t)}(\mathbf{x}_2)}.$$

This algorithm maximizes approximately the marginal likelihood with respect to the parameters. The approximation depends on the accuracy of the posterior independence assumption. The EM algorithm is a special case of the variational methodology, in which there is no approximation (for example independence assumption) and the bound is tight in the VE-step, i.e.

$$F(q^{(t+1)}(\mathbf{x}_1), q^{(t+1)}(\mathbf{x}_2), \boldsymbol{\theta}^{(t)}) = L(\boldsymbol{\theta}^{(t)}).$$

The result for two sets of hidden variables can be generalized for  $M$  hidden variables  $\mathbf{x}_1, \dots, \mathbf{x}_M$ , [10]. The updates in this case are:

$$q^{(t+1)}(\mathbf{x}_i) = \frac{\exp \left\{ - \langle \log p(\mathbf{y}, \mathbf{x}_1, \dots, \mathbf{x}_M; \boldsymbol{\theta}) \rangle_{\prod_{k \neq i} q^{(t)}(\mathbf{x}_k)} \right\}}{\int \exp \left\{ - \langle \log p(\mathbf{y}, \mathbf{x}_1, \dots, \mathbf{x}_M; \boldsymbol{\theta}) \rangle_{\prod_{k \neq i} q^{(t)}(\mathbf{x}_k)} \right\} d\mathbf{x}_i},$$

$$\boldsymbol{\theta}^{(t+1)} = \arg \max_{\boldsymbol{\theta}} \langle \log p(\mathbf{y}, \mathbf{x}_1, \dots, \mathbf{x}_M; \boldsymbol{\theta}) \rangle_{\prod_k q^{(t)}(\mathbf{x}_k)}.$$

### 2.3 Constrained variational optimization

There are cases where it is convenient to have functions  $q(\mathbf{x}_i; \lambda_i)$ ,  $i = 1, \dots, M$ , in a particular parametric form (e.g. Gaussian) with parameters  $\lambda_i$ . Doing this, we constrain the inferred posteriors to follow the form of these functions. In this way, the parameters  $\lambda_i$  of the approximate posteriors are amenable to estimation. In this case, in the VE-step the maximization is performed with respect to the parameters  $\lambda_i$  instead of the function  $q(\mathbf{x}_i)$ . Also, this methodology leads to closed form updates in the VB algorithm that work as constraints to the approximate posteriors.

## 2.4 Bayesian inference for restoration with the SAR prior model

At this point, it is interesting to present the EM algorithm as an example, that is employed to solve the SAR prediction model [94] for the restoration problem. Under this, the image is assumed as the hidden variable  $\mathbf{f}$  of the model with Gaussian prior:

$$p(\mathbf{f}) \propto a^{\frac{N-1}{2}} \exp\left\{-\frac{1}{2}a\|\mathbf{Q}\mathbf{f}\|_2^2\right\},$$

where  $a$  is the precision (inverse variance) of the Gaussian, and  $\mathbf{Q}$  is the Laplacian operator.

The observation is the degraded image  $\mathbf{g}$  with conditional

$$p(\mathbf{g}|\mathbf{f}) = \beta^{\frac{N}{2}} \exp\left\{-\frac{\beta}{2}\|\mathbf{g} - \mathbf{H}\mathbf{f}\|_2^2\right\},$$

where  $\beta$  is the precision of the Gaussian noise distribution.

The EM algorithm performs i) the inference of the posterior of the image (E-step) and ii) estimation of the parameters (M-Step). Thus, the E-step update at iteration  $t$  is given by:

$$q^{(t)}(\mathbf{f}) = p(\mathbf{f}|\mathbf{g}) = \frac{p(\mathbf{g}|\mathbf{f})p(\mathbf{f})}{p(\mathbf{g})} = N\left(\mathbf{m}^{(t)}, (\mathbf{C}^{(t)})^{-1}\right).$$

This result shows that the posterior is a Normal distribution with mean  $\mathbf{m}^{(t)} = \beta^{(t)}\mathbf{C}^{(t)}\mathbf{H}^T\mathbf{g}$  and covariance  $\mathbf{C}^{(t)} = (\beta^{(t)}\mathbf{H}^T\mathbf{H} + a^{(t)}\mathbf{Q}^T\mathbf{Q})^{-1}$ .

In the M-step the parameters  $a^{(t)}$  and  $\beta^{(t)}$  are estimated by maximizing the expectation of the complete likelihood with respect to the posterior:

$$[\beta^{(t)}, a^{(t)}] = \arg \max_{[\beta, a]} \langle p(\mathbf{g}|\mathbf{f})p(\mathbf{f}) \rangle_{q^{(t)}(\mathbf{f})}$$

This results in the following the updates:

$$\beta^{(t)} = \frac{N}{\text{trace}\{\mathbf{C}^{(t)}\mathbf{H}^T\mathbf{H}\}}, \text{ and } a^{(t)} = \frac{N}{\text{trace}\{\mathbf{C}^{(t)}\mathbf{Q}^T\mathbf{Q}\}}.$$

The traces of the matrices can be calculated conveniently in the DFT domain. When the algorithm converges, the restored image is taken to be the mean  $\mathbf{m}$  of the posterior  $p(\mathbf{f}|\mathbf{g})$ .

Finally, note that the similarity between the linear equation solved to obtain  $\mathbf{m}^{(t)}$  in the E-step and that of Eq. (1.9) obtained in a deterministic framework is obvious.



# CHAPTER 3. SPATIALLY ADAPTIVE IMAGE RESTORATION

- 
- 3.1. Visibility based non-stationary restoration
  - 3.2. Half-quadratic regularization
  - 3.3. Markov Random Fields
  - 3.4 Total-variation regularization
  - 3.5. Regularization in the Wavelet Domain
  - 3.6. Other image priors
  - 3.7. Super-resolution methods
  - 3.8. Conclusions
- 

In this chapter, we review the literature and the efforts for development of realistic image models and methods for adaptive regularization that preserves edges and also eliminates noise in smooth image areas. The methods that have been proposed are based either on a stochastic or a deterministic formulations. It must be noted here that the deterministic formulation is equivalent to the stochastic, because, as it was shown in section 1.4, the roughness penalty of the image in the deterministic framework can be interpreted in a stochastic framework as the assignment of a particular image prior. However, the stochastic formulation offers the great advantage of a rigorous way to estimate the model parameters. An example of this is the EM algorithm that solves the simultaneously auto-regressive (SAR) prediction model [94] presented in section 1.4.3. However, the SAR model is not a spatially adaptive and cannot model efficiently the image local characteristics.

First, in section 3.1 we present a heuristic method for adaptive regularization. Then, a very popular methodology for deterministic regularization will be presented in section 3.2 which is

the half-quadratic regularization methodology. Markov-random fields (MRFs) are also very popular for probabilistic image modeling and they have been applied successfully in image restoration, as we will see in section 3.3. TV-based regularization is also a very successful and popular methodology and it is presented in 3.4. In section 3.5 regularization methods in the wavelet domain are presented. Lastly, conclusions and remarks about the similarities and differences of the reviewed methodologies are given, as well as their merits and drawbacks.

### 3.1 Visibility based non-stationary restoration

In [1] a measure of spatial detail is defined by a noise masking function  $M(\mathbf{f})(i)$  at pixel  $i$  which depends on  $\mathbf{f}$ . The visibility function  $v(i)$  which expresses the relationship between the noise visibility and the masking function, is defined experimentally. In [76] the masking function is set to be the local variance of the pixel and according to [1] the visibility function  $v$  is defined as:

$$v(i) = \frac{1}{\theta M(i) + 1}, \quad i = 1, \dots, N,$$

where  $\theta$  is a scale parameter that depends on the image. This function goes to zero when the local variance goes to infinity (pixel belongs to an edge) and to one when the variance is close to zero (pixel belongs to a smooth region). Finally, as a result of an extended analysis [76] in contrast to Euclidean norm as in Eq. (1.7), a weighted norm is introduced for the penalty term:

$$a \|\mathbf{Q}\mathbf{f}\|_{\Lambda} = a \mathbf{f}^T \mathbf{Q}^T \Lambda \mathbf{Q} \mathbf{f},$$

where  $\Lambda$  is a  $N \times N$  diagonal matrix with elements:  $\Lambda(i, i) = v^2(i)$ ,  $i = 1, \dots, N$ . The matrix  $\mathbf{Q}$  is recommended to be a high-pass filter and the example of the  $p$ -th order (discrete) derivative operator is given, for example for  $p = 2$  it is the Laplacian operator.

The resulting linear system  $(\mathbf{H}^T \mathbf{H} + a \mathbf{Q}^T \Lambda \mathbf{Q}) \hat{\mathbf{f}} = \mathbf{H}^T \mathbf{g}$  is solved iteratively with a constraint gradient descent algorithm. Also, the visibility function is evaluated using the restored image. This leads to an iterative scheme of estimating the image  $\mathbf{f}$  and the elements of  $\Lambda$ :

$$\mathbf{f}^{(t+1)} = (\mathbf{H}^T \mathbf{H} + a \mathbf{Q}^T \Lambda^{(t)} \mathbf{Q})^{-1} \mathbf{H}^T \mathbf{g},$$

$$\Lambda^{(t+1)}(i,i) = \frac{1}{\theta M(\mathbf{f}^{(t+1)})(i) + 1}.$$

Notice now that the restoration filter in a sense “adapts” to the image spatially via the different diagonal elements of  $\Lambda$ , which depend on the local characteristics of the image. The main drawback of this method is that its parameters  $a$  and  $\theta$  must be specified in advance.

Katsaggelos and Kang [82] proposed a similar methodology, with the estimation of the regularization parameter being a part of the restoration algorithm. They also incorporated spatial adaptivity for the data fidelity term, which corresponds to the noise, and the resulting iteration was

$$\mathbf{f}^{(t+1)} = (\mathbf{H}^T \Lambda_1^{(t)} \mathbf{H} + a \mathbf{Q}^T \Lambda^{(t)} \mathbf{Q})^{-1} \mathbf{H}^T \mathbf{g},$$

where now  $\Lambda_1^{(t)}$  is also a diagonal matrix that adapts to the noise local characteristics. Lastly, they provided an alternative method to estimate the diagonal matrices in every iteration.

### 3.2 Half-Quadratic Regularization

Half-quadratic (HQ) regularization has been developed with the goal to provide an edge-preserving restoration method. Herein, we follow the demonstration of the HQ methodology in [62] and [31]. According to that, the image is estimated by minimizing a function analogous to the one in Eq. (1.10). The estimate is given by minimization of a non-quadratic function:

$$\hat{\mathbf{f}} = \arg \min_{\mathbf{f}} \{J_1(\mathbf{f})\}, \quad (3.1)$$

where  $J_1(\mathbf{f}) = \|\mathbf{g} - \mathbf{H}\mathbf{f}\|_2^2 + aJ(\mathbf{f})$  and  $J(\mathbf{f})$  is a function of  $\mathbf{f}$  that enforces a regularized solution. In the MAP approach this implies the prior for the image to be  $p(\mathbf{f}) \propto \exp\{-aJ(\mathbf{f})\}$ .  $J(\mathbf{f})$  is computed as:

$$J(\mathbf{f}) = \sum_{i=1}^N \varphi([\mathbf{D}_x \mathbf{f}]_i) + \varphi([\mathbf{D}_y \mathbf{f}]_i), \quad (3.2)$$

where  $\mathbf{D}_x$  and  $\mathbf{D}_y$  are the horizontal and vertical difference operators. Thus,  $[\mathbf{D}_x \mathbf{f}]_i$  and  $[\mathbf{D}_y \mathbf{f}]_i$  are the horizontal and vertical difference of the image at location  $i$ . The function

$\varphi(t)$  is called the potential function and is specified to penalize the gradient of the image and simultaneously to preserve the edges. In [31] necessary conditions about  $\varphi(t)$  are given in order for the minimizer of  $J_1(\mathbf{f})$  (Eq. (3.2)) to converge and in addition to provide an image where the edges are preserved.

To make the half-quadratic regularization mechanism clear, we follow the demonstration of it in [31]. We first write the gradient

$$\nabla J_1(\mathbf{f}) = \mathbf{H}^T \mathbf{H} \mathbf{f} - \mathbf{H}^T \mathbf{g} - a(\mathbf{D}_x^T \mathbf{B}_x(\mathbf{f}) \mathbf{D}_x + \mathbf{D}_y^T \mathbf{B}_y(\mathbf{f}) \mathbf{D}_y) \mathbf{f},$$

where  $\mathbf{B}_x(\mathbf{f})$  and  $\mathbf{B}_y(\mathbf{f})$  are diagonal matrices that depend on  $\mathbf{f}$  and their elements are given by

$$\mathbf{B}_x(\mathbf{f})(i, i) = \frac{\varphi'([\mathbf{D}_x \mathbf{f}]_i)}{2[\mathbf{D}_x \mathbf{f}]_i} \quad \text{and} \quad \mathbf{B}_y(\mathbf{f})(i, i) = \frac{\varphi'([\mathbf{D}_y \mathbf{f}]_i)}{2[\mathbf{D}_y \mathbf{f}]_i}.$$

Finding the minimum by setting the gradient equal to zero ( $\nabla J(\mathbf{f}) = \mathbf{0}$ ) leads to a non-linear equation. However, at the minimum  $\mathbf{f}_{\min}$  it holds that

$$\mathbf{f}_{\min} = \left( \mathbf{H}^T \mathbf{H} + a(\mathbf{D}_x^T \mathbf{B}_x(\mathbf{f}_{\min}) \mathbf{D}_x + \mathbf{D}_y^T \mathbf{B}_y(\mathbf{f}_{\min}) \mathbf{D}_y) \right)^{-1} \mathbf{H}^T \mathbf{g}.$$

Based on this result, by fixing the matrices  $\mathbf{B}_x$  and  $\mathbf{B}_y$  ignoring their dependence on  $\mathbf{f}$ , the following iterative scheme can be applied

$$\mathbf{f}^{(t+1)} = \left( \mathbf{H}^T \mathbf{H} + a(\mathbf{D}_x^T \mathbf{B}_x(\mathbf{f}^{(t)}) \mathbf{D}_x + \mathbf{D}_y^T \mathbf{B}_y(\mathbf{f}^{(t)}) \mathbf{D}_y) \right)^{-1} \mathbf{H}^T \mathbf{g}, \quad (3.3)$$

where  $\mathbf{B}_x(\mathbf{f}^{(t)})$  and  $\mathbf{B}_y(\mathbf{f}^{(t)})$  are the matrices formed using the image estimate  $\mathbf{f}^{(t)}$  of the previous iteration  $t$ . Note that the above equation is linear and the non-quadratic minimization problem turns into successive quadratic problems. What remains is to prove that this algorithm converges to a solution of problem (3.1). To do this, in [31] an ‘‘augmented’’ function of  $J_1$  is introduced, using the function  $J^*(\mathbf{f}, \mathbf{b}_x, \mathbf{b}_y)$ , where  $\mathbf{b}_x, \mathbf{b}_y$  are vectors of the auxiliary variables  $b_x(i), b_y(i)$ ,  $i = 1, \dots, N$  introduced to ‘‘augment’’ the function  $J(\mathbf{f})$ . The exact form of  $J^*(\mathbf{f}, \mathbf{b}_x, \mathbf{b}_y)$  depends on the form of  $\varphi(t)$  and it is not presented here for the

moment. It has been shown that  $J^*(\mathbf{f}, \mathbf{b}_x, \mathbf{b}_y)$  shares the same minimum with  $J(\mathbf{f})$ ; thus, minimization of the latter leads to minimization of the former. The steps for the iterative minimization algorithm of  $J^*(\mathbf{f}, \mathbf{b}_x, \mathbf{b}_y)$  are

$$\mathbf{f}^{(t+1)} = \arg \min_{\mathbf{f}} J^*(\mathbf{f}, \mathbf{b}_x^{(t)}, \mathbf{b}_y^{(t)}) \Rightarrow \mathbf{f}^{(t+1)} = \left( \mathbf{H}^T \mathbf{H} + a \left( \mathbf{D}_x^T \mathbf{B}_x^{(t)} \mathbf{D}_x + \mathbf{D}_y^T \mathbf{B}_y^{(t)} \mathbf{D}_y \right) \right)^{-1} \mathbf{H}^T \mathbf{g}, \quad (3.4)$$

$$[\mathbf{b}_x^{(t+1)}, \mathbf{b}_y^{(t+1)}] = J^*(\mathbf{f}^{(t+1)}, \mathbf{b}_x, \mathbf{b}_y) \Rightarrow$$

$$b_x^{(t+1)}(i) = \frac{\varphi'([\mathbf{D}_x \mathbf{f}^{(t)}]_i)}{2[\mathbf{D}_x \mathbf{f}^{(t)}]_i} \quad \text{and} \quad b_y^{(t+1)}(i) = \frac{\varphi'([\mathbf{D}_y \mathbf{f}^{(t)}]_i)}{2[\mathbf{D}_y \mathbf{f}^{(t)}]_i}, \quad (3.5)$$

where  $\mathbf{B}_x^{(t)}$  and  $\mathbf{B}_y^{(t)}$  are the diagonal matrices with elements the vectors  $\mathbf{b}_x^{(t)}$  and  $\mathbf{b}_y^{(t)}$ . This algorithm is identical to that of Eq. (3.3) and it is proven to converge to a solution of Eq. (3.1). The function that is minimized is called half-quadratic because if the auxiliary variables are kept fixed then  $J^*(\mathbf{f}, \mathbf{b}_x, \mathbf{b}_y)$  is quadratic with respect to  $\mathbf{f}$ . Furthermore, it is straightforward to minimize the function with respect to the auxiliary variables. This leads to an algorithm containing linear equations. In [69] and [1] the convergence properties of the half-quadratic algorithm are studied.

At this point an example of the potential function  $\varphi(t)$  is provided. An interesting algorithm results when  $\varphi(t) = \log(1 + \theta t^2)$  [31]. In this case the auxiliary variables are updated according to

$$b_x^{(t)}(i) = \frac{1}{1 + \theta([\mathbf{D}_x \mathbf{f}]_i)^2} \quad \text{and} \quad b_y^{(t)}(i) = \frac{1}{1 + \theta([\mathbf{D}_y \mathbf{f}]_i)^2}.$$

Moreover, if  $\mathbf{D}_x = \mathbf{D}_y = \mathbf{Q}$ , where  $\mathbf{Q}$  the Laplacian operator, the algorithm would be similar to the empirical method of section 3.1. This result demonstrates that the auxiliary variables must not be considered simply as parameters introduced for the relaxation of the minimization problem, but in addition as the manifestation of the spatial adaptivity mechanism. Notice also that we obtain the same algorithm following two different paths, the empirical and the half-quadratic. Next, we present a third path, the stochastic modeling approach.

In [22] Champagnat and Idier have shown that when half-quadratic regularization is viewed in a stochastic framework, the above minimization algorithm is equivalent to an EM algorithm. More specifically, if the function  $J_1(\mathbf{f})$  to be minimized can be viewed as the negative logarithm of the posterior of the data given the observations:

$$J_1(\mathbf{f}) \propto -\log p(\mathbf{f} | \mathbf{g}) \propto -\log p(\mathbf{g} | \mathbf{f}) p(\mathbf{f}),$$

where  $p(\mathbf{g} | \mathbf{f}) \propto \exp\left\{-\frac{\beta}{2}\|\mathbf{g} - \mathbf{H}\mathbf{f}\|_2^2\right\}$  is the data likelihood with  $\beta$  the noise precision and  $p(\mathbf{f}) \propto \exp\{-\alpha J(\mathbf{f})\}$ . Then, the auxiliary variables  $\mathbf{b}_x, \mathbf{b}_y$  can be viewed as hidden variables with prior  $p(\mathbf{b}_x, \mathbf{b}_y)$  and the minimum with respect to these variables function  $J_{\min}^*(\mathbf{f}, \hat{\mathbf{b}}_x, \hat{\mathbf{b}}_y)$  can be viewed as the expectation:

$$J_{\min}^*(\mathbf{f}, \hat{\mathbf{b}}_x, \hat{\mathbf{b}}_y) \propto \left\langle -\log p(\mathbf{g}, \mathbf{f}, \mathbf{b}_x, \mathbf{b}_y) \right\rangle_{p(\mathbf{b}_x, \mathbf{b}_y | \mathbf{f})},$$

where  $p(\mathbf{b}_x, \mathbf{b}_y | \mathbf{f})$  the posterior of the hidden variables given the image. Thus, evaluation of  $\hat{\mathbf{b}}_x, \hat{\mathbf{b}}_y$  according to Eq. (3.5) is included in the E-step. In the M-step the above expectation is minimized with respect to  $\mathbf{f}$ , which is achieved by application of Eq. (3.4).

In light of the above, we can write the augmented half-quadratic function:

$$J^*(\mathbf{f}, \mathbf{b}_x, \mathbf{b}_y) = \|\mathbf{g} - \mathbf{H}\mathbf{f}\|_2^2 + a \sum_{i=1}^N \left( b_x(i) ([\mathbf{D}_x \mathbf{f}]_i)^2 + \psi(b_x(i)) + b_y(i) ([\mathbf{D}_y \mathbf{f}]_i)^2 + \psi(b_y(i)) \right)$$

and note that  $\psi(w)$  plays the role of the prior (the logarithm) for the variables  $b_x(i)$  and  $b_y(i)$ . Of course, the form  $\psi(w)$  depends on  $\varphi(t)$ . It is interesting to see that the form of  $\psi(w)$  for  $\varphi(t) = \log(1 + \theta t^2)$  is  $\psi(w) = w - \log w - 1$ , and this is analogous to the logarithm of a Gamma prior. This result indicates that in this case  $p(\mathbf{f})$  is a Student's-t distribution [19] (with  $\nu = 1, \lambda = \theta$ ) because it is given by the integral of a product of a Normal and a Gamma prior for the precision of the normal:

$$p(\mathbf{f}) = \int p(\mathbf{f} | \mathbf{b}_x, \mathbf{b}_y) p(\mathbf{b}_x, \mathbf{b}_y) d\mathbf{b}_x d\mathbf{b}_y \propto \left(1 + ([\mathbf{D}_x \mathbf{f}]_i)^2\right)^{-1} \left(1 + ([\mathbf{D}_y \mathbf{f}]_i)^2\right)^{-1}.$$

This agrees with the conclusion in [22] that optimization with the EM algorithm in the Bayesian framework under a Student's-t prior is equivalent to half-quadratic regularization for a certain potential function. This fact holds more generally for scale mixtures of Gaussians (SGM). Student's-t, Laplace and hyperbolic distributions are just special cases of GSMs. Thus, it is not surprising that in Chapter 6 the use of the Student's-t distribution as prior leads to a similar restoration algorithm. Also, in [22] it is shown that the auxiliary variables of the half-quadratic criteria can be interpreted as the line process of a CGMRF (see next section).

The half-quadratic algorithm in [31] converges when the potential function is convex and the matrix  $\mathbf{H}$  in the term  $\|\mathbf{g} - \mathbf{H}\mathbf{f}\|_2^2$  is full-rank [39]. In [7] Delaney and Bresler propose a modified half-quadratic algorithm that converges even when the potential function is not convex and/or the matrix  $\mathbf{H}$  is not full-rank. They apply their algorithm to computed tomography (CT), a problem which shares an analogous linear imaging model with that of restoration (Eq. (1.2)) but the linear operator  $\mathbf{H}$  in this case represents the Radon transform. Thus, the image is reconstructed from the projections  $\mathbf{g}$ , with half-quadratic criteria used to constraint the solution.

### 3.3 Markov Random Fields

Markov random fields (MRF) are stochastic models that have been used extensively as image models in image processing and generally in computer vision [111], [20] and [86]. The main advantage of these models is the flexible mechanism that they provide to define image priors. For the image restoration problem this is achieved by the introduction of potential functions that provide a regularized solution, much in the same way with the half-quadratic regularization presented in section 3.2. This flexibility is mainly due to the fact that the stochastic optimization algorithms that are utilized do not depend much on the choice of the potential functions. Notice that this is not the case for half-quadratic deterministic minimization algorithms which can be applied only when the potential function meets some criteria and when the form of the function to be minimized is quadratic with the auxiliary variables fixed. However, the disadvantage of the stochastic algorithms is the large computational cost.

The work that introduced MRFs in image restoration was that of Geman & Geman [61]. The image  $\mathbf{f}$  is regarded as a two-dimensional  $N \times N$  MRF. This means that each pixel  $\mathbf{f}(i, j)$  is a node with conditional probability given all the other nodes,

$$P(\mathbf{f}(i, j) | \mathbf{f}(m, n), \forall (m, n) \neq (i, j)),$$

equal to the conditional probability given only the neighboring nodes, i.e.

$$P(\mathbf{f}(i, j) | \mathbf{f}(m, n), \forall (m, n) \neq (i, j)) = P(\mathbf{f}(i, j) | \mathbf{f}(m, n), \forall (m, n) \in N(i, j)),$$

where  $N(i, j)$  is the set of nodes that are neighbors to  $(i, j)$ . A typical such neighborhood may include the four pixels around the pixel  $(i, j)$ . Thus, in order to define an MRF for the image, it is sufficient to define the local conditional probabilities.

In [61] the overall probability for the image  $p(\mathbf{f})$  is written as follows using the Hammersley-Clifford theorem:

$$p(\mathbf{f}) = \frac{1}{Z} \exp\left\{-\frac{1}{T} U(\mathbf{f})\right\} = \frac{1}{Z} \exp\left\{-\frac{1}{T} \sum_{\forall C \in \mathbb{S}} V_C(\mathbf{f})\right\},$$

where  $\mathbb{S}$  is the set of cliques of the image<sup>1</sup>,  $V_C(\mathbf{f})$  is a function of the nodes that belong only to clique  $C$  and it is called the *potential function*. The normalizing constant  $Z$  is the *partition function* and the parameter  $T$  is called *temperature*. This theorem provides a powerful tool to define MRFs in a straightforward manner through the definition of potential functions. For example, in [61] only cliques with neighboring pixels are used, i.e.  $V_C(\mathbf{f}) = 0$  unless  $C = \{r, s\}$ , and the following potential is defined:

$$V_{r,s} = \begin{cases} 1, & \text{if } \mathbf{f}_r = \mathbf{f}_s \\ 0, & \text{if } \mathbf{f}_r \neq \mathbf{f}_s \end{cases},$$

where  $r, s$  denote neighboring pixels. The role of this potential is to enforce high probability for neighboring pixels to have similar values.

---

<sup>1</sup>Loosely speaking, a clique is a set of neighboring nodes. There is a hierarchy of cliques. At the first level, each clique consists of one node. At the second level, the clique contains a node and its neighbors. Cliques of higher levels seldom appear in applications of MRFs.



However, this is not a realistic model in areas with edges. So, a line process has been included in the MRF with the introduction of the binary variables  $\mathbf{l}$ , indicating whether there is an edge between neighboring pixels. To incorporate this property to the model the following Gibbs joint distribution has been proposed:

$$p(\mathbf{f}, \mathbf{l}) = \frac{1}{Z} \exp \left\{ -\frac{1}{T} U(\mathbf{f}, \mathbf{l}) \right\} = \frac{1}{Z} \exp \left\{ -\frac{1}{T} \sum_{\forall C \in \mathcal{S}} V_C(\mathbf{f}, \mathbf{l}) \right\},$$

where now  $V_C(\mathbf{f}, \mathbf{l}) = V_C(\mathbf{f} | \mathbf{l}) + V_C(\mathbf{l})$ . The first potential  $V_C(\mathbf{f} | \mathbf{l})$  defines the local dependencies between the pixels given information of the presence of ‘breaks’ between them. This information is provided by a binary valued line process variables: zero for no edge and one for edge. The term  $V_C(\mathbf{l})$  works as a penalty on images with an excessive amount of edges.

Given the observed image the goal is to find the restored image according to the rule:

$$[\hat{\mathbf{f}}, \hat{\mathbf{l}}] = \arg \max_{[\mathbf{f}, \mathbf{l}]} p(\mathbf{f}, \mathbf{l} | \mathbf{g}) = \arg \max_{[\mathbf{f}, \mathbf{l}]} p(\mathbf{g}, \mathbf{f}, \mathbf{l}).$$

In [61] this is achieved with the a stochastic relaxation algorithm that uses the Gibbs sampler as a tool to draw samples from  $p(\mathbf{g}, \mathbf{f}, \mathbf{l})$ . A nice feature of the sampling procedure is that there is no need to know explicitly the partition function of the MRF.

In the methods based on compound Gauss-Markov random fields (CGMRF) [72], [74], [73], [34] and [92], the potential function  $V_C(\mathbf{f} | \mathbf{l})$  is quadratic, which makes the distribution  $p(\mathbf{f} | \mathbf{l})$  Gaussian. Moreover, the line process parameters may be discrete or continuous. We give an example by defining the prior  $p(\mathbf{f}, \mathbf{l})$  as in [92], but using one-dimensional notation for brevity:

$$-\log p(\mathbf{f}, \mathbf{l}) = \text{const} + \sum_{i=1}^N (f(i) - f(i+1))^2 (1 - l(i)) + \beta l(i),$$

where  $\beta$  is a scalar and  $l(i)$  is the line process parameter that is 0 if there is no edge between the  $f(i)$  and  $f(i+1)$  pixels and 1 if there is. The term  $\beta l(i)$  works as a penalty and prohibits the parameter  $l(i)$  to be always one across the image, because otherwise the MAP

estimator would yield all these parameters equal to one. Given  $l(i)$ , a Gaussian prior for the image is obtained:

$$p(\mathbf{f}|\mathbf{l}) \propto \exp\left\{-\frac{1}{2\sigma^2}\mathbf{f}^T \mathbf{A}(\mathbf{l})\mathbf{f}\right\},$$

where  $\mathbf{A}(\mathbf{l})$  a matrix that depends on the values of the line process parameters. Simulated annealing is usually used to perform maximization of these models given the observed degraded image, which is usually time-consuming. This type of minimization algorithm does not require a closed form for the partition function of the prior something that would be necessary in a deterministic algorithm.

In [72], [74], [73], [34] the line process is continuous. In this case, the minimization of the posterior  $p(\mathbf{f},\mathbf{l}|\mathbf{g})$  reminds strongly the minimization of Eq. (3.1) in the half-quadratic regularization case. This indicates the relation that exists between the CGMRF models and the half-quadratic criteria as shown in [22].

An interesting model is proposed in [21] where the Gaussian function of the CGMRF is replaced by a generalized Gaussian distribution (GGD). In this case, the line process is not incorporated to the model because of the edge preservation properties of the GGD. This distribution has also been used in other works for image restoration as it is mentioned in section 3.5.

### 3.4 Total-variation regularization

Total-variation (TV) regularization was first introduced by Rudin, Osher and Fatemi [102] for the image denoising problem. Later, the TV regularizer became popular in other image processing problems [3], [134], [103], [57], including restoration. According to TV-based image restoration, the restored image is obtained by the rule:

$$\hat{\mathbf{f}} = \arg \min_{\mathbf{f}} \left\{ \|\mathbf{g} - \mathbf{H}\mathbf{f}\|_2^2 + aTV(\mathbf{f}) \right\}, \quad (3.6)$$

where  $a$  is the regularization parameter and  $TV(\mathbf{f})$  is the total-variation regularizer that is added to penalize the roughness of the image, and is defined as follows:

$$TV(\mathbf{f}) = \sum_{i=1}^N \sqrt{[\mathbf{D}_x \mathbf{f}]_i^2 + [\mathbf{D}_y \mathbf{f}]_i^2},$$

where  $\mathbf{D}_x$  and  $\mathbf{D}_y$  are the horizontal and vertical first order difference operators. The TV penalty term has the ability to provide edge-preserving regularization, meaning that it constrains the solution of Eq. (3.6) to give images without amplified noise in smooth areas and simultaneously preserve their edges [46].

There have been many efforts to solve Eq. (3.6), especially in the continuous domain [23] using variational optimization theory where the estimated image is the solution of a non-linear partial differential equation (PDE). The PDE is the result of the application of the Euler-Lagrange transformation to the optimization problem of minimizing an integral with respect to a function (the image).

An interesting approach to solve Eq. (3.6) is that in [18]. The authors propose the majorization-minimization (MM) methodology [85], according to which the problem of minimizing a non-quadratic function is transformed to successive minimizations of quadratic surrogate functions, i.e. majorizers of the non-quadratic function that can be minimized linearly. The iterative scheme is described as follows:

$$\mathbf{f}^{(t+1)} = \arg \min_{\mathbf{f}} \{Q(\mathbf{f} | \mathbf{f}^{(t)})\}, \quad (3.7)$$

where  $Q(\mathbf{f} | \mathbf{f}^{(t)}) = \|\mathbf{g} - \mathbf{H}\mathbf{f}\|_2^2 + \frac{a}{2} \sum_{i=1}^N \frac{([\mathbf{D}_x \mathbf{f}]_i^2 + [\mathbf{D}_y \mathbf{f}]_i^2)}{\sqrt{[\mathbf{D}_x \mathbf{f}^{(t)}]_i^2 + [\mathbf{D}_y \mathbf{f}^{(t)}]_i^2}}$  is the majorizer of the function to

be minimized. This means that

$$Q(\mathbf{f} | \mathbf{f}^{(t)}) \geq \|\mathbf{g} - \mathbf{H}\mathbf{f}\|_2^2 + aTV(\mathbf{f}), \text{ and } Q(\mathbf{f}^{(t)} | \mathbf{f}^{(t)}) = \|\mathbf{g} - \mathbf{H}\mathbf{f}^{(t)}\|_2^2 + aTV(\mathbf{f}^{(t)}).$$

where these two equations can be verified by noticing that the inequality

$$\sqrt{x} \leq \sqrt{x_0} + \frac{x - x_0}{2\sqrt{x_0}},$$

holds for every positive  $x$  and  $x_0$  and the equality holds when  $x = x_0$  and we replace  $([\mathbf{D}_x \mathbf{f}]_i^2 + [\mathbf{D}_y \mathbf{f}]_i^2)$  for  $x$  and  $[\mathbf{D}_x \mathbf{f}^{(t)}]_i^2 + [\mathbf{D}_y \mathbf{f}^{(t)}]_i^2$  for  $x_0$  (neglecting the constant terms).

Thus,  $Q(\mathbf{f}|\mathbf{f}^{(t)})$  meets the criteria to be a majorizer according to the MM methodology, and hence the iterative scheme described by Eq. (3.7) converges to a solution of (3.6). Solution of (3.7) corresponds to solution of the linear system:

$$\mathbf{f}^{(t+1)} = \left( \mathbf{H}^T \mathbf{H} + a \left( \mathbf{D}_x^T \mathbf{W}^{(t)} \mathbf{D}_x + \mathbf{D}_y^T \mathbf{W}^{(t)} \mathbf{D}_y \right) \right)^{-1} \mathbf{H}^T \mathbf{g},$$

where  $\mathbf{W}^{(t)}$  is a diagonal matrix with elements  $\mathbf{W}^{(t)}(i,i) = \frac{1}{2\sqrt{[\mathbf{D}_x \mathbf{f}^{(t)}]_i^2 + [\mathbf{D}_y \mathbf{f}^{(t)}]_i^2}}$  and it

manifests the spatially adaptive regularization mechanism that enforces less regularization at the edges (small  $\mathbf{W}^{(t)}(i,i)$ ) and greater in smooth areas (large  $\mathbf{W}^{(t)}(i,i)$ ).

It is interesting to note that the MM methodology is very similar to the half-quadratic. Both transform the non-linear minimization problem to successive linear problems, because both  $Q(\mathbf{f}|\mathbf{f}^{(t)})$  and  $J^*(\mathbf{f}, \mathbf{b}_x, \mathbf{b}_y)$  are quadratic with respect to  $\mathbf{f}$  when the rest variables are kept fixed. Also, taking into account that the EM algorithm belongs to the general category of MM algorithms, and that the half-quadratic algorithm is strongly connected with the EM [22], it can be also concluded that MM optimization is strongly related to the half-quadratic approach.

To demonstrate this, let us give an example where the MM methodology is applied in order to minimize the function

$$J_1(\mathbf{f}) = \|\mathbf{g} - \mathbf{H}\mathbf{f}\|_2^2 + aJ(\mathbf{f})$$

where  $J(\mathbf{f}) = \sum_{i=1}^N \varphi([\mathbf{D}_x \mathbf{f}]_i) + \varphi([\mathbf{D}_y \mathbf{f}]_i)$  and  $\varphi(t) = \log(1 + \theta t^2)$ . Using the inequality [85]

$$\log x \leq \log x_0 + \frac{x - x_0}{x_0},$$

the majorizer of  $J_1(\mathbf{f})$  is constructed:

$$Q(\mathbf{f}|\mathbf{f}^{(t)}) = \|\mathbf{g} - \mathbf{H}\mathbf{f}\|_2^2 + a\theta \sum_{l=x,y} \sum_{i=1}^N \left( \frac{([\mathbf{D}_l \mathbf{f}]_i^2)}{1 + \theta [\mathbf{D}_l \mathbf{f}^{(t)}]_i^2} + \log \left( 1 + \theta [\mathbf{D}_l \mathbf{f}^{(t)}]_i^2 \right) - 1 \right).$$

This majorizer is *very similar* to the function  $J^*(\mathbf{f}, \mathbf{b}_x, \mathbf{b}_y)$  in the half-quadratic case as demonstrated in section 3.2 for this potential  $\varphi(t)$ , when  $\left(1 + \theta \left[ \mathbf{D}_x \mathbf{f}^{(t)} \right]_i^2\right)^{-1}$  and  $\left(1 + \theta \left[ \mathbf{D}_y \mathbf{f}^{(t)} \right]_i^2\right)^{-1}$  are replaced with  $b_x(i)$  and  $b_y(i)$ , respectively. Moreover, MM minimization gives the same restoration algorithm as this half-quadratic example.

In some methods, for example [3] and [57], a variant of the TV regularizer was used, based on the  $l_1$  norm:

$$TV_{ni}(\mathbf{f}) = \sum_{i=1}^N \left( \left| \left[ \mathbf{D}_x \mathbf{f} \right]_i \right| + \left| \left[ \mathbf{D}_y \mathbf{f} \right]_i \right| \right),$$

which is called *non-isotropic* TV. In [3] and [57] linear and quadratic programming optimization methods are proposed to solve both equations:

$$\hat{\mathbf{f}} = \arg \min_{\mathbf{f}} \left\{ \left\| \mathbf{g} - \mathbf{Hf} \right\|_2^2 + a \sum_{i=1}^N \left( \left| \left[ \mathbf{D}_x \mathbf{f} \right]_i \right| + \left| \left[ \mathbf{D}_y \mathbf{f} \right]_i \right| \right) \right\},$$

$$\hat{\mathbf{f}} = \arg \min_{\mathbf{f}} \left\{ \left\| \mathbf{g} - \mathbf{Hf} \right\|_1 + a \sum_{i=1}^N \left( \left| \left[ \mathbf{D}_x \mathbf{f} \right]_i \right| + \left| \left[ \mathbf{D}_y \mathbf{f} \right]_i \right| \right) \right\},$$

where in the second case  $\left\| \mathbf{g} - \mathbf{Hf} \right\|_1$  denotes the  $l_1$  norm which can be used to model non-Gaussian noise.

It is interesting to see how we can perform the minimization of the first objective function (for Gaussian noise) using the MM methodology. First, we have to introduce a majorizer, which is achieved using the following inequality:

$$|x| = \sqrt{x^2} \leq \sqrt{x_0^2} + \frac{x^2 - x_0^2}{2|x_0|}.$$

Thus, the resulting majorizer is:

$$Q(\mathbf{f} | \mathbf{f}^{(t)}) = \left\| \mathbf{g} - \mathbf{Hf} \right\|_2^2 + \frac{a}{2} \sum_{l=x,y} \sum_{i=1}^N \frac{\left( \left[ \mathbf{D}_l \mathbf{f} \right]_i^2 \right)}{\left| \left[ \mathbf{D}_l \mathbf{f}^{(t)} \right]_i \right|},$$

which is minimized by solving the linear system

$$\mathbf{f}^{(t+1)} = \left( \mathbf{H}^T \mathbf{H} + \frac{a}{2} (\mathbf{D}_x^T \mathbf{W}_x^{(t)} \mathbf{D}_x + \mathbf{D}_y^T \mathbf{W}_y^{(t)} \mathbf{D}_y) \right)^{-1} \mathbf{H}^T \mathbf{g},$$

with the elements of the diagonal matrices  $\mathbf{W}_x^{(t)}$  and  $\mathbf{W}_y^{(t)}$  given by

$$\mathbf{W}_x^{(t)}(i, i) = \frac{1}{\left| \left[ \mathbf{D}_x \mathbf{f}^{(t)} \right]_i \right|}, \quad \mathbf{W}_y^{(t)}(i, i) = \frac{1}{\left| \left[ \mathbf{D}_y \mathbf{f}^{(t)} \right]_i \right|}.$$

Note also that this algorithm has a Bayesian interpretation and can be derived by assuming a Laplacian image prior and applying the EM algorithm, as in [52] where a Laplacian prior was used in order to obtain sparse solutions for supervised learning<sup>2</sup>.

In order to estimate the TV regularization parameter  $a$ , Lagrange multipliers are employed in [102]. In [17] a Bayesian method is proposed to estimate this parameter. According to this method the solution of Eq. (3.7) is equivalent to the MAP approach:

$$\hat{\mathbf{f}} = \arg \min_{\mathbf{f}} \left\{ -\log p(\mathbf{g} | \mathbf{f}) p(\mathbf{f}) \right\},$$

where  $p(\mathbf{g} | \mathbf{f}) \propto \exp \left\{ -\frac{\beta}{2} \|\mathbf{g} - \mathbf{H}\mathbf{f}\|^2 \right\}$  is (as usual) the data likelihood with noise variance  $\beta$  and  $p(\mathbf{f}) \propto \exp \left\{ -aTV(\mathbf{f}) \right\}$  is the TV prior for the image. In [18],  $\beta$  is assumed known and  $a$  is integrated out, thus it is not estimated.

In [8] both the noise and regularization parameters are estimated in a Bayesian framework. Precisely, the *variational* EM (VEM) methodology is utilized where the image is treated as random variable and a distribution is inferred for the image. The parameters are estimated in the maximization step of the VEM algorithm. More details are given in Chapter 6, where VEM is also employed to solve an improved TV image model of [8].

### 3.5 Regularization in the Wavelet Domain

Wavelet-based methods have been very successful for image denoising, see for example [96] and [119]. However, their application to the image restoration is not as straightforward because of the convolutional operator that appears when the restoration imaging model is

---

<sup>2</sup> This corresponds to edge-preserving regularization in image restoration

formulated in the wavelet domain. This formulation is based on the representation of the image  $\mathbf{f}$  via the wavelet transform coefficients  $\boldsymbol{\theta}$ , i.e.  $\mathbf{f} = \mathbf{W}\boldsymbol{\theta}$ , where  $\mathbf{W}$  is a matrix representing the inverse discrete wavelet transform (DWT). Thus, if the wavelet basis is orthogonal,  $\mathbf{W}^T$  represents the inverse of  $\mathbf{W}$ . Then, the image is estimated by first minimizing the penalized log-likelihood:

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} \left\{ \|\mathbf{g} - \mathbf{H}\mathbf{W}\boldsymbol{\theta}\|_2^2 + aC(\boldsymbol{\theta}) \right\},$$

where  $C(\boldsymbol{\theta})$  is the penalty term on the coefficients, and then taking  $\hat{\mathbf{f}} = \mathbf{W}\hat{\boldsymbol{\theta}}$ . The function  $C(\boldsymbol{\theta})$  has a particular form to induce *sparseness*, which means that it enforces the majority of the coefficients to have small values. Using the terminology of spatially adaptive methods presented previously, this approach provides edge-preserving regularization. In a MAP framework, the prior for the coefficients is given by:

$$p(\boldsymbol{\theta}) \propto \exp\{-aC(\boldsymbol{\theta})\}, \quad (3.8)$$

and in most cases independence is assumed, i.e.  $C(\boldsymbol{\theta}) = \sum_i C(\boldsymbol{\theta}_i)$ .

This formulation is adopted in many early works on wavelet-based regularization; see for example [9], [10], [43], [100]. For a review of these works (and related works) the interested reader is referred to [16] and [55]. We focus here on recent works that have been introduced independently ([38], [55], [44], [45], [54], [88], [123], [119]) and they can be interpreted as MM algorithms, as shown in [53].

In [53] the MM methodology is applied to the term  $\|\mathbf{g} - \mathbf{H}\mathbf{W}\boldsymbol{\theta}\|_2^2$  and the minimization problem of Eq. (3.8) is transformed to a sequence of minimizations problems given by

$$\boldsymbol{\theta}^{(t+1)} = \arg \min_{\boldsymbol{\theta}} \left\{ \|\boldsymbol{\theta} - \boldsymbol{\varphi}^{(t)}\|_2^2 + aC(\boldsymbol{\theta}) \right\}, \quad (3.9)$$

with  $\boldsymbol{\varphi}^{(t)} = \boldsymbol{\theta}^{(t)} + \mathbf{W}^T \mathbf{H}^T (\mathbf{g} - \mathbf{H}\mathbf{W}\boldsymbol{\theta}^{(t)})$ . Notice that this vector is the previous estimate of  $\boldsymbol{\theta}^{(t)}$  minus the gradient of the term  $\|\mathbf{g} - \mathbf{H}\mathbf{W}\boldsymbol{\theta}^{(t)}\|_2^2$ . In this way, inversion of large matrices that include both  $\mathbf{H}$  and  $\mathbf{W}$  is avoided. This MM algorithm coincides with the *iterative shrinkage/thresholding* (IST) algorithm proposed in [38] and [54].

If  $C(\boldsymbol{\theta})$  can be written as  $C(\boldsymbol{\theta}) = \sum_i C(\boldsymbol{\theta}_i)$ , Eq. (3.9) can be solved independently for each coefficient  $\boldsymbol{\theta}_i$ :

$$\boldsymbol{\theta}_i^{(t+1)} = \arg \min_{\boldsymbol{\theta}_i} \left\{ \|\boldsymbol{\theta}_i - \boldsymbol{\Phi}_i^{(t)}\|_2^2 + aC(\boldsymbol{\theta}_i) \right\}.$$

Furthermore, in case  $C(\boldsymbol{\theta}_i)$  has a convenient form for optimization, such as  $C(\boldsymbol{\theta}_i) = |\boldsymbol{\theta}_i|^p$  for  $1 \leq p \leq 2$  (notice that this corresponds to a Generalized Gaussian distribution), closed form solution of  $\boldsymbol{\theta}_i^{(t+1)}$  can be obtained for some values of  $p$ . For example, in the Gaussian case ( $p = 2$ ) the update rule takes the form:

$$\boldsymbol{\theta}_i^{(t+1)} = \frac{\boldsymbol{\Phi}_i^{(t)}}{1+a}.$$

Solutions for other values of  $p$  are presented in [33]. In [38] the convergence properties of the IST algorithm with GGD prior are analyzed.

The distribution of each coefficient  $\boldsymbol{\theta}_i$  can be considered as a (univariate) Gaussian scale mixture (GSM) [6]

$$p(\boldsymbol{\theta}_i) = \int_0^{+\infty} \frac{1}{\sqrt{2\pi z}} \exp\{-a\boldsymbol{\theta}_i^2 z^{-1}\} p(z) dz,$$

where  $p(z)$  is the multiplier density. The integral represents the sum of infinite zero mean Gaussian densities with  $p(z)$  representing the weight of each Gaussian that is added. In [33], an MM approach to solve Eq. (3.9) under the GSM prior is presented. When the marginal density has the form  $p(\boldsymbol{\theta}_i) \propto \exp\{-aC(\boldsymbol{\theta}_i)\}$  and belongs to the GSM family (for example GGD), then a (quadratic) majorizer of  $-\log p(\boldsymbol{\theta}_i)$  is

$$a \frac{C'(\boldsymbol{\theta}_i^*)}{\boldsymbol{\theta}_i^*} \boldsymbol{\theta}_i^2 + \text{const} \geq aC(\boldsymbol{\theta}_i).$$

This can be plugged into Eq. (3.9) and obtain an MM algorithm. Notice that this majorizer is similar to that in the half-quadratic case (see section 3.2) where the auxiliary variables are



updated using the term  $\frac{\varphi'(t)}{t}$ . This result is validated as follows. The majorizer can be written using the likelihood:

$$-a \frac{p(\boldsymbol{\theta}_i^*)}{\boldsymbol{\theta}_i^* p'(\boldsymbol{\theta}_i^*)} \boldsymbol{\theta}_i^2 + const \geq -\log p(\boldsymbol{\theta}_i),$$

and this is identical to Eq. (8) of [16], where an EM algorithm is derived for image restoration using GSM priors. Also, remember from section 3.2 that the half-quadratic methodology is actually an EM algorithm under GSM priors. Based on these observations, it is concluded that the MM algorithm with GSM prior is equivalent to the half-quadratic methods.

In [33], an algorithm is also given where the data-likelihood term is not majorized, unlike the prior term. The result is a generalized MM algorithm in which a linear system is solved at each iteration.

Alternatives to the MAP approach have been also used in image restoration and denoising, see for example [100], [64], [110] and [121]. An interesting work that employs GSM for image restoration is presented in [64]. In this, a multivariate GSM is used where the vector  $\boldsymbol{\theta}$  of the coefficients being GSM distributed as

$$p(\boldsymbol{\theta}) = \int_0^{+\infty} \frac{1}{(2\pi)^{N/2} |z\mathbf{C}|^{1/2}} \exp\{-a\boldsymbol{\theta}^T (\mathbf{C}z)^{-1} \boldsymbol{\theta}\} p(z) dz,$$

where  $\mathbf{C}$  is the  $N \times N$  covariance matrix of  $\boldsymbol{\theta}$ . Instead of using the MAP approach, the authors propose the Bayesian Least Squares (BLS) estimator.

There are problems that have similar formulation with restoration in the wavelet domain and can be expressed by the equation

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} \left\{ \|\mathbf{g} - \mathbf{A}\boldsymbol{\theta}\|_2^2 + aC(\boldsymbol{\theta}) \right\},$$

where the matrix  $\mathbf{A}$  has replaced  $\mathbf{HW}$ . The LASSO criterion (least absolute shrinkage and selection operator), for example, used for robust regression is similar to the above formulation with  $C(\boldsymbol{\theta}) = \|\boldsymbol{\theta}\|_1$  and  $\mathbf{A}$  is the design matrix of the regression problem. Furthermore, in [44] and [45] blurring is absent ( $\mathbf{H} = \mathbf{I}$ ) and the columns of  $\mathbf{W}$  represent a redundant dictionary, i.e. a basis that forms the solution. Moreover,  $C(\boldsymbol{\theta})$  leads to sparse representation, which

means it is desirable to make most of the coefficients of the dictionary as small as possible. For  $C(\boldsymbol{\theta}) = \|\boldsymbol{\theta}\|_1$  this problem is the *basis pursuit* denoising problem [35].

### 3.6 Other image priors

The image priors presented so far are applied to the outputs of pre-selected filters, such as first order differences and the filters of the wavelet transform. There are works where the filters are learned from a large training set of images, for example the Fields of Experts (FOE) training algorithm [114]. In this dissertation we focus on methods that use a priori fixed filters. In [122] an application of the FOE is presented.

Finally, image priors which are based on the Huber function:

$$\rho_T(x) = \begin{cases} x^2, & |x| \leq T \\ T^2 + 2(|x| - T), & |x| > T \end{cases}$$

have been proposed, see for example [104]. Using such priors leads to the minimization of

$$\|\mathbf{g} - \mathbf{H}\mathbf{f}\|_2^2 + \sum_{i=1}^N \left( \rho_T \left( [\mathbf{D}_x \mathbf{f}]_i^2 \right) + \rho_T \left( [\mathbf{D}_y \mathbf{f}]_i^2 \right) \right)$$

in order to obtain an estimate of  $\mathbf{f}$ . This can be achieved using the MM methodology by exploiting a quadratic bound for the Huber function:

$$\rho_T(x) \leq N_T(x | x^{(t)}) = \begin{cases} x^2, & |x| \leq T \\ \frac{T^2}{|x^{(t)}|} x^2 + |x^{(t)}| T - T, & |x| > T \end{cases}$$

where equality hold when  $x = x^{(t)}$ .

### 3.7 Super-resolution methods

Recent surveys and articles for super-resolution can be found in [105], [48] and the edited book [32]. Many methodologies have been applied to the super-resolution problem. An important category of methods formulates this problem as an ill-posed image reconstruction problem [40] and introduces prior information (regularization) to reconstruct the high-resolution image [105]. This renders the image model presented in this chapter, applicable

also to super-resolution. In this dissertation, this formulation is adopted and the development of effective image priors for the restoration problem leads inevitably to development of effective super-resolution algorithms. Super-resolution viewed as learning problem has also been recently considered in [56] and [75].

Recent efforts based on the regularized reconstruction methodology for the super-resolution problem are described in [118]-[113]. In what follows, we will concentrate on the regularized reconstruction point of view. Regularized reconstruction can be also viewed as a maximum a posteriori (MAP) approach by assuming an appropriate probability density for the error in the assumed imaging model and an appropriate prior for the image [40]. Thus, in what follows we will not distinguish between these two approaches.

In [118] the problem of reconstructing high-resolution frames from compressed video is examined using a Bayesian formulation based on a Gaussian simultaneously autoregressive (SAR) stationary image prior. In [1] a methodology is proposed based on the theory of projections onto convex sets [51]. In [48], color images and demosaicing are considered, and regularization (image priors) based on the  $L_1$ -norm is proposed in order to avoid the shortcomings of  $L_2$ -norm based regularization. Furthermore, non-Gaussian measurement errors are considered. It was shown that  $L_1$ -norm minimization yields better results in the case of inaccuracies in the imaging model. In [48], a computationally fast method is proposed based on the  $L_1$ -norm assuming known integer pixel displacements between frames. However, in [48] and [48] the parameters that define the regularization term are chosen empirically. In [133] an expectation-maximization (E-M) algorithm and a maximum a posteriori algorithm (MAP) are presented for simultaneous registration, restoration and interpolation for super-resolution. Nevertheless, a stationary SAR prior is used in both formulations in [133]. In [65] different degradations are assumed in each low resolution observation. However,  $L_2$ -norm based stationary regularization is used. In [113] an interesting statistical performance analysis is presented that offers insight into the fundamental bottlenecks limiting the performance of super-resolution algorithms.

Kanemura *et al* [77] used a CGMRF to model the image and applied approximate Bayesian inference for the super-resolution problem to infer the high-resolution image. All the parameters of the model, including the registration parameters are estimated in a rigorous

manner. However, time-consuming techniques are used for the estimation of some parameters which are avoided in our method presented in Chapter 6.

### **3.8 Conclusions**

From the literature survey on spatially adaptive image restoration, we first conclude that all the image priors lead to spatially adaptive restoration filters, which enforces spatially adaptive regularization. The only exception to this is the Gaussian prior which leads to spatially invariant filters. The spatial adaptivity mechanism is manifested by the auxiliary variables in the deterministic (half-quadratic) or the hidden variables in the stochastic framework (EM algorithm).

As described in section 3.2, regularization based on half-quadratic criteria is equivalent to the EM algorithm under a GSM prior. Student's-t and the Laplace distributions belong to the GSM family. We presented an example and showed that the EM algorithm with a Student's-t prior coincides with an algorithm by applying the half-quadratic methodology with a specific potential function. Also, we obtained the same algorithm by applying the MM methodology to this potential. For the Laplace prior, we obtained the same algorithms by applying EM and MM. From all these, we conclude that half-quadratic, EM and MM are very similar methodologies (at least for GSM-based prior) that view the problem from different perspectives.

It is also clear that there is not sufficient work on the estimation of the proposed model parameters. The Bayesian methodology used herein provides an elegant way for statistical modeling and estimation. In this dissertation we employ the Bayesian inference methodology to both estimate the model parameters and infer the image.

# CHAPTER 4. BAYESIAN IMAGE RESTORATION USING A HIERARCHICAL NON-STATIONARY PRIOR

---

- 4.1. Introduction
  - 4.2. Imaging and image prior model
  - 4.3. Maximum a posterior (MAP) estimation
  - 4.4. Bayesian algorithm
  - 4.5. Numerical Experiments
  - 4.6. Conclusions and extensions
- 

In this chapter we present image restoration algorithms based on the Bayesian framework and a new hierarchical spatially adaptive image prior. The proposed prior has the following two desirable features. First, it models the image discontinuities in different directions with a model which is continuously valued. Thus, it generalizes the on/off (binary) line process idea used in previous image prior definitions within the context of Markov Random Fields (MRF). Second, the proposed Bayesian framework with the hierarchical prior has been shown to be successful in generating sparse representations in other signal processing problems. Using the proposed hierarchical prior, two restoration algorithms are derived. The first is based on the maximum a posteriori (MAP) principle, and the second on the Bayesian methodology. Numerical experiments are presented that compare the proposed algorithms among themselves and with previous stationary and non-stationary MRF-based algorithms. These experiments demonstrate the advantages of the proposed prior.

## 4.1 Introduction

In this chapter we formulate the image restoration problem in the Bayesian framework because it offers many advantages, such as a systematic and flexible way for regularization, and a rigorous framework for estimation of the model parameters, as it was discussed in Chapter 3. The novel prior that we introduce for regularization ameliorates some of the disadvantages and difficulties of the methods presented in Chapter 3. In the next paragraphs we present these disadvantages.

The disadvantage of an unrealistic stationary prior appears in many Bayesian formulations for the image restoration problem, where the image prior is based on a Gaussian stationary assumption for the residuals of the local image differences, see section 1.4.3. Such stationary models are seriously handicapped because they do not provide the flexibility to model the spatially varying structure of images in edge and texture areas. In other words, such priors enforce smoothness uniformly across the entire image and correspond to uniform “regularization”.

Compound Gauss-Markov models provide a way to avoid uniform “regularization”, see section 3.3. In these models the image is assumed to be generated by a two-level process. The first level represents the correlations of adjacent pixels of the image. The second level contains a binary process used to capture the image variations (edges). In other words, when the line process between two pixels is “on” smoothness is not enforced between them while when it is “off” smoothness is enforced. From an image modeling point of view the binary (on/off) nature of the line process that is used is insufficient to capture the image variations of most natural images. More specifically, edges of different strengths and “degrees of sharpness” are present in natural images and a binary model is limited since it inevitably introduces quantization in representing them.

The methods based on the error visibility idea use a continuous (non-binary) model to capture the visibility of the image artifacts, see section 3.1. Since the visibility of the artifacts is related to the variation structure of the image, these methods use a continuous model for the image variations. However, their main shortcoming is that quantification of visibility is not rigorous but rather heuristic. Thus, the estimation of all the necessary parameters is not based on a systematic framework derived from a rigorous model. As a result such models are cumbersome to use, and suboptimal.

In this chapter, we introduce a methodology that ameliorates the difficulties of the above mentioned methods. In particular, we propose a hierarchical (two-level) Gaussian non-

stationary image prior [27]. This prior assumes that the residuals of the first order differences of the image in four *different directions* are Gaussian random variables with zero mean and variance that is *spatially varying*. As a result these local directional variances capture the image discontinuities with a continuous value model and can be thought as “continuous line processes”, in contrast to the CGMRFs models. In order to deal with the resulting over-parameterization of this model, the spatially varying variances are considered as random variables (not parameters) and a Gamma hyper-prior is imposed on them. The parameters of the imposed hyper-prior control the mean and the variance of the residual variances and in a sense control the *degree of non-stationarity* of the imposed image prior.

Another aspect of this image model is that it enforces *sparse* first order directional differences in the image using the same Bayesian mechanism as in sparse kernel based regression and basis selection; see for example [128] and [132]. Sparse signal representations have found extensive applications in inverse problems and are becoming very important area of research for many signal and image processing applications [132].

To learn this model and infer the image we propose two iterative algorithms. The first is based on the maximum a posteriori estimation (MAP) principle and computes explicitly both the image and the spatially varying variances in all four directions. The second is a Bayesian algorithm that marginalizes the “hidden variables”, see for example [12]. At this point we would like to make two observations. First, unlike MRF based models, the generative graphical model that stems from the proposed prior in this chapter *does not contain cycles*, thus, learning and inference based on it is easy. Second, we obtain as MAP estimates of the inverse local variances the spatially adaptive regularization weights in [80], [84], and [78] which were previous obtained based on heuristic arguments.

We provide numerical experiments where we compared the proposed restoration algorithms with two different versions of the classical Wiener filter [70], the constrained least squares approach with spatially adaptive constraints, [76], [84], [78] and [80], previous Bayesian algorithms based on the stationary SAR model in [92], as well as on the CGMRF that use “binary line processes” [92]. We also compared the proposed algorithms among themselves in terms of both the bias and the variance of the inferred restored image using a Monte-Carlo simulation. Our experimental results are encouraging and demonstrate the advantages of both the proposed new prior and the employed Bayesian methodology.

The rest of this chapter is organized as follows: in section 4.2 we present the imaging model and the proposed image prior model. In section 4.3 we present the MAP-based restoration algorithm. In section 4.4 we derive the Bayesian marginalization based algorithm. In section 4.5 we present our numerical experiments, and finally in section 4.6 we provide conclusions and directions for future research.

## 4.2 Imaging and image prior model

A linear imaging model is assumed. The  $N \times 1$  vector  $\mathbf{g}$ , represents the observed degraded image which is obtained by

$$\mathbf{g} = \mathbf{H}\mathbf{f} + \mathbf{n}, \quad (4.1)$$

where  $\mathbf{f}$  is the (unknown) original image,  $\mathbf{H}$  is a  $N \times N$  known convolution matrix and  $\mathbf{n}$  is additive white noise. We assume Gaussian statistics for the noise given by  $\mathbf{n} \sim N(\mathbf{0}, \beta^{-1}\mathbf{I})$  where  $\mathbf{0}$  is a  $N \times 1$  vector with zeros and  $\mathbf{I}$  the  $N \times N$  identity matrix respectively, and  $\beta^{-1}$  the noise variance which is assumed unknown.

For the image prior model we assume that the first order differences of the image  $\mathbf{f}$  in four directions,  $0^\circ$ ,  $90^\circ$ ,  $45^\circ$  and  $135^\circ$  respectively, are given by:

$$\begin{aligned} \varepsilon^1(i, j) &= f(i, j) - f(i, j+1), \varepsilon^2(i, j) = f(i, j) - f(i+1, j), \\ \varepsilon^3(i, j) &= f(i, j) - f(i+1, j+1), \text{ and } \varepsilon^4(i, j) = f(i, j) - f(i-1, j+1), \end{aligned} \quad (4.2)$$

with  $\varepsilon^k(i, j)$   $k=1,2,3,4$ , the difference residuals for the image location  $(i, j)$ . The above equations can be also written in matrix vector form for the entire image as  $\mathbf{Q}^k\mathbf{f} = \boldsymbol{\varepsilon}^k$ ,  $k=1,2,3,4$  where  $\mathbf{Q}^k$  is the  $N \times N$  directional difference operators for  $N \times 1$  images. Without loss of generality, in what follows, for convenience, we will use one dimensional notation, in other words, we assume  $\boldsymbol{\varepsilon}^k = [\varepsilon_1^k \varepsilon_2^k, \dots, \varepsilon_N^k]^T$ . We assume that the residuals have Gaussian statistics according to  $\varepsilon_i^k \sim N\left(0, (a_i^k)^{-1}\right)$ , for  $i=1,2 \dots N$  and  $k=1,2,3,4$ , where  $a_i^k$  the inverse variance of  $\varepsilon_i^k$  and  $N$  the size of the image.



For the inverse variance  $a_i^k$  we introduce the notation  $\mathbf{A}^k = \text{diag} \{a_1^k, a_2^k, \dots, a_N^k\}$  an  $N \times N$  diagonal matrix and  $\tilde{\mathbf{A}} = \text{diag} \{\mathbf{A}^1, \mathbf{A}^2, \mathbf{A}^3, \mathbf{A}^4\}$  a  $4N \times 4N$  diagonal matrix and  $\tilde{\mathbf{a}} = [\mathbf{a}^1, \mathbf{a}^2, \mathbf{a}^3, \mathbf{a}^4]^T$  a  $4N \times 1$  vector. Also for the errors we use the notation  $\tilde{\boldsymbol{\epsilon}} = [\boldsymbol{\epsilon}^1, \boldsymbol{\epsilon}^2, \boldsymbol{\epsilon}^3, \boldsymbol{\epsilon}^4]^T$ . We assume that the errors in each direction and at each pixel location are independent. This is based on the assumption that at each pixel location an edge can occur at any direction independently of what happens in adjacent pixels. This assumption makes subsequent calculations tractable. Thus, the joint density for the errors is Gaussian and is given as

$$p(\tilde{\boldsymbol{\epsilon}}; \tilde{\mathbf{a}}) \propto \prod_{k=1}^4 \prod_{i=1}^N (a_i^k)^{1/2} \exp\left(-0.5 \left( (\boldsymbol{\epsilon}^k)^T \mathbf{A}^k \boldsymbol{\epsilon}^k \right)\right) = \prod_{k=1}^4 \prod_{i=1}^N (a_i^k)^{1/2} \exp\left(-0.5 \left( \tilde{\boldsymbol{\epsilon}}^T \tilde{\mathbf{A}} \tilde{\boldsymbol{\epsilon}} \right)\right).$$

To relate  $\tilde{\boldsymbol{\epsilon}}$  with the image  $\mathbf{f}$  we define the  $4N \times N$  operator  $\tilde{\mathbf{Q}} = [(\mathbf{Q}^1)^T, (\mathbf{Q}^2)^T, (\mathbf{Q}^3)^T, (\mathbf{Q}^4)^T]^T$ . Then, the relation between the image and the residuals is  $\tilde{\boldsymbol{\epsilon}} = \tilde{\mathbf{Q}}\mathbf{f}$ . Based on this relation and  $p(\tilde{\boldsymbol{\epsilon}}; \tilde{\mathbf{a}})$  we can define an *improper prior* for the image  $\mathbf{f}$ . This prior is given by:

$$\begin{aligned} p(\mathbf{f}; \tilde{\mathbf{a}}) &\propto \prod_{k=1}^4 \prod_{i=1}^N (a_i^k)^{1/8} \exp\left(-0.5 \left( (\tilde{\mathbf{Q}}\mathbf{f})^T \tilde{\mathbf{A}} \tilde{\mathbf{Q}}\mathbf{f} \right)\right) \\ &= \prod_{k=1}^4 \prod_{i=1}^N (a_i^k)^{1/8} \exp\left(-0.5 \left( (\mathbf{Q}^k\mathbf{f})^T \mathbf{A}^k \mathbf{Q}^k\mathbf{f} \right)\right). \end{aligned} \quad (4.3)$$

This prior is termed *improper* since it is not scaled to integrate to 1. For a proper Gaussian the normalizing constant as a function of the spatially varying variances  $a_i^k$  cannot be of the form in Eq. (4.3), since  $\tilde{\mathbf{Q}}$  is not a square matrix,

$$\text{Det}\{\tilde{\mathbf{Q}}^T \tilde{\mathbf{A}} \tilde{\mathbf{Q}}\} \neq C \prod_{k=1}^4 \prod_{i=1}^N (a_i^k)^{\frac{1}{8}},$$

where  $C$  is a constant, even though  $\tilde{\mathbf{A}}$  is diagonal. However, improper priors are used on a routinely basis with success in Bayesian modeling [13]. More specifically, the prior in equation (4.3) is obtained by assuming that all the elements of the diagonal matrix  $\tilde{\mathbf{A}}$  are equal to their geometric mean  $\left(\prod_{i=1}^N \prod_{k=1}^4 a_i^k\right)^{1/4N}$ . This implies, because

$\tilde{\mathbf{Q}}^T \tilde{\mathbf{A}} \tilde{\mathbf{Q}} = \sum_{k=1}^4 (\mathbf{Q}^k)^T \mathbf{A}^k \mathbf{Q}^k$  is a  $N \times N$  matrix, that  $(\det\{\tilde{\mathbf{Q}}^T \tilde{\mathbf{A}} \tilde{\mathbf{Q}}\})^{1/2} \propto \left( \prod_{i=1}^N \prod_{k=1}^4 \alpha_i^k \right)^{1/8}$ . This

assumption results in the improper prior in Eq. (4.3) and leads to tractable calculations.

The role of the parameters  $a_i^k$  is to capture the directional variation structure of the image. More specifically, a large variance (small  $a_i^k$ ) indicates the presence of a large variation along the direction of the difference, in other words an edge perpendicular to this direction. The introduction of the spatially varying  $a_i^k$  scales down the differences of adjacent pixels in regions of image discontinuities. As a result this prior maintains edges and suppresses noise in smooth areas of the image. This principle is identical to the one that motivated the use of the binary (0 or 1) line process idea; see for example [61], [20], [72], [74] and [92]. However, since the values of  $a_i^k$  are continuous our model can be considered as *generalization* of the MRF model with the on/off binary line process.

The drawback of the proposed prior is that it introduces  $4N$  parameters  $a_i^k$  that have to be estimated from  $N$  observations. This is clearly not a desirable situation from an estimation point of view. For this purpose we employ the Bayesian paradigm and consider  $a_i^k$  as random variables (instead of parameters) and introduce Gamma hyper-priors for them. In the case of a stationary model where all  $a_i^k$ 's are equal the over parameterization problem does not exist and it is rather straightforward to obtain good estimates for the unknown parameters using even maximum likelihood (ML).

The rationale for using a Gamma prior in the non stationary case is threefold. First, it is “conjugate” for the variance of a Gaussian and facilitates analysis of the Bayesian model [12]. Second, similar hierarchical models have been used successfully in Bayesian formulations of other statistical learning problems and produce *sparse* representations; see for example [128] and [132]. Sparse local differences encouraged by this model are a good model for image edges which are overall much less than the pixels in the image. Finally, as we shall see in what follows it produces update equations for  $a_i^k$ 's that were previously derived empirically.

We consider the following parameterization for the Gamma hyper-prior:

$$p(a_i^k; m_k, l_k) \propto (a_i^k)^{\frac{l_k-2}{2}} \exp\{-m_k (l_k - 2) a_i^k\}, \quad (4.4)$$

For such a representation the mean and variance of Gamma are given by  $E[a_i^k] = l_k (2m_k (l_k - 2))^{-1}$ , and  $Var[a_i^k] = l_k (2m_k^2 (l_k - 2)^2)^{-1}$  respectively, see [92] and [60]. This representation is used because the value of the parameter  $l_k$  can be also interpreted as the level of confidence to the prior knowledge provided by the Gamma hyper prior. More specifically, as  $l_k \rightarrow \infty$ ,  $E[a_i^k] \rightarrow (2m_k)^{-1}$  and  $Var[a_i^k] \rightarrow 0$ . In other words, the prior becomes very informative and restrictive resulting in  $a_i^k = (2m_k)^{-1} \forall i$ . This also implies that the image model becomes stationary. In contrast, when  $l_k \rightarrow 2$  then both  $E[a_i^k] \rightarrow \infty$  and  $Var[a_i^k] \rightarrow \infty$ , thus the prior becomes uninformative and does not influence at all the values of the  $a_i^k$ 's. In other words, the  $a_i^k$ 's are free from the moderating influence of the prior and are allowed to “vary wildly” following the data. In such case the image model becomes “highly non stationary”. As a result, the value of the parameter  $l_k$  can be also viewed as a way to adjust for the *degree of non stationarity* of the image model.

The graphical model corresponding to the observations generation mechanism is shown in Figure 4.1 below. Random variables are represented as ellipses and parameters of the distribution of the random variables as rectangles. The variables  $\tilde{\mathbf{a}}$  and  $\mathbf{f}$  are hidden, since they are not observed and relate the parameters with the observations. In the MAP approach, point estimates are obtained for  $\tilde{\mathbf{a}}$  and  $\mathbf{f}$ . In the Bayesian approach, instead of using point estimates for  $\tilde{\mathbf{a}}$ , their average influence is taken into consideration by integrating them out (marginalization), therefore point estimates are used only for  $\mathbf{f}$ .

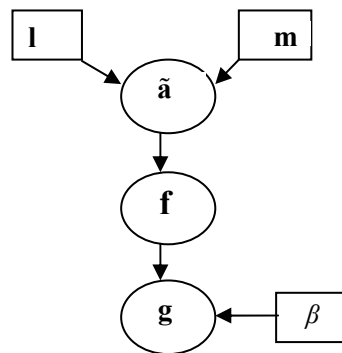


Figure 4.1: The graphical model of the observations.

### 4.3 Maximum a posteriori (MAP) estimation

At first we propose a MAP approach to infer  $\tilde{\mathbf{a}}$  and  $\mathbf{f}$ . This is based on maximization of the posterior probability. Thus we have:

$$p(\mathbf{f}, \tilde{\mathbf{a}} | \mathbf{g}; \beta, \mathbf{m}, \mathbf{l}) \propto p(\mathbf{g}, \mathbf{f}, \tilde{\mathbf{a}}; \beta, \mathbf{m}, \mathbf{l}) = p(\mathbf{g} | \mathbf{f}, \tilde{\mathbf{a}}; \beta) p(\mathbf{f} | \tilde{\mathbf{a}}) p(\tilde{\mathbf{a}}; \mathbf{m}, \mathbf{l}),$$

where:

$$\mathbf{m} = [m_1, m_2, m_3, m_4]^T, \mathbf{l} = [l_1, l_2, l_3, l_4]^T.$$

Maximizing the quantity  $p(\mathbf{g}, \mathbf{f}, \tilde{\mathbf{a}}; \beta, \mathbf{m}, \mathbf{l})$  with respect to  $\mathbf{f}$  and  $\tilde{\mathbf{a}}$  is equivalent to minimizing the negative logarithm

$$\begin{aligned} J_{MAP}(\mathbf{f}, \tilde{\mathbf{a}} | \mathbf{g}; \beta, \mathbf{m}, \mathbf{l}) &\propto -\log p(\mathbf{g}, \mathbf{f}, \tilde{\mathbf{a}}; \beta, \mathbf{m}, \mathbf{l}) = -\log p(\mathbf{g} | \mathbf{f}, \tilde{\mathbf{a}}; \beta) - \log p(\mathbf{f} | \tilde{\mathbf{a}}) - \\ &\log p(\tilde{\mathbf{a}}; \mathbf{m}, \mathbf{l}) \Rightarrow J(\mathbf{f}, \tilde{\mathbf{a}} | \mathbf{g}; \beta, \mathbf{m}, \mathbf{l}) = -\frac{N}{2} \log \beta + \frac{1}{2} \beta \|\mathbf{H}\mathbf{f} - \mathbf{g}\|^2 - \frac{1}{8} \sum_{k=1}^4 \sum_{i=1}^N \log a_i^k + \quad (4.5) \\ &+ \frac{1}{2} \sum_{k=1}^4 \sum_{i=1}^N (\mathbf{Q}^k \mathbf{f})^T \mathbf{A}^k \mathbf{Q}^k \mathbf{f} - \sum_{k=1}^4 \left( \frac{l_k - 2}{2} \sum_{i=1}^N \log a_i^k \right) + \sum_{k=1}^4 \left( m_k (l_k - 2) \sum_{i=1}^N a_i^k \right). \end{aligned}$$

To minimize the above function with respect to  $\mathbf{f}$  and  $\tilde{\mathbf{a}}$ , we adopt an iterative scheme that sets alternatively the gradient of  $\mathbf{f}$  and  $\tilde{\mathbf{a}}$  equal to zero. Setting  $\nabla_{\tilde{\mathbf{a}}} J_{MAP}(\mathbf{f}, \tilde{\mathbf{a}} | \mathbf{g}; \beta, \mathbf{m}, \mathbf{l}) = 0$  yields

$$(a_i^k)^* = \frac{\left( \frac{1}{8} + \frac{1}{2} (l_k - 2) \right)}{\left( \frac{1}{2} (\varepsilon_i^k)^2 + m_k (l_k - 2) \right)}. \quad (4.6)$$

Setting  $\nabla_{\mathbf{f}} J_{MAP}(\mathbf{f}, \tilde{\mathbf{a}} | \mathbf{g}; \beta, \mathbf{m}, \mathbf{l}) = 0$  yields:

$$\mathbf{f}^* = \left( \mathbf{H}^T \mathbf{H} + \beta^{-1} \sum_{k=1}^4 (\mathbf{Q}^k)^T \mathbf{A}^k \mathbf{Q}^k \right)^{-1} \mathbf{H}^T \mathbf{g}. \quad (4.7)$$

However, equation (4.7) cannot be solved in closed form since analytical inversion of  $\left[ \mathbf{H}^T \mathbf{H} + \beta^{-1} \sum_{k=1}^4 (\mathbf{Q}^k)^T \mathbf{A}^k \mathbf{Q}^k \right]$  is not possible due to the non-circulant nature of matrices  $\mathbf{A}^k$ .

Thus, we resort to a numerical solution using a conjugate gradient algorithm. The proposed MAP algorithm iterates between equations (4.6) and (4.7) till convergence, i.e. image estimate doesn't differ much from the estimate of the previous iteration.

It is interesting to point out that a formula similar to Eq. (4.6) was used in previous works to compute spatially varying regularization weights. Such a formula was derived based on heuristic arguments and empirical observations, see [76], [78], [80] and [84].

In addition, the observation of the previous section that the parameters  $l_k$  control the degree of non-stationarity of the model can be verified from Eq. (4.6), the MAP estimates of the  $(a_i^k)$ . More specifically, when  $l_k \rightarrow \infty$ ,  $(a_i^k)^* = (2m_k)^{-1} \forall i$ , and the image model becomes stationary. In contrast, when  $l_k \rightarrow 2$ ,  $(a_i^k)^* = ((\varepsilon_i^k)^2)^{-1} \forall i$ , thus the  $(a_i^k)^*$ 's are completely unaffected from the moderating effect of the Gamma hyper-prior and only follow the data. For example, in smooth areas of the image where the local residual in the denominator of Eq. (4.6) tend to zero, it holds that  $(a_i^k)^* \rightarrow \infty$ .

#### 4.4 Bayesian algorithm

In the Bayesian analysis of the proposed model, hidden variables are marginalized while parameters are estimated [12]. In our case, as explained in section 4.2,  $\mathbf{f}$  and  $\tilde{\mathbf{a}}$  are considered “hidden” (latent) variables, while  $\mathbf{m}$ ,  $\mathbf{l}$  and  $\beta$  are the unknown parameters. In the Bayesian inference paradigm, hidden variables are marginalized while parameters are estimated by maximizing the likelihood  $p(\mathbf{g}; \beta, m, l)$  of the observations  $\mathbf{g}$ :

$$p(\mathbf{g}; \beta, m, l) = \iint p(\mathbf{g}, \mathbf{f}, \tilde{\mathbf{a}}; \beta, m, l) d\mathbf{f} d\tilde{\mathbf{a}} = \iint p(\mathbf{g} | \mathbf{f}; \beta) p(\mathbf{f} | \tilde{\mathbf{a}}) p(\mathbf{a}; m, l) d\mathbf{f} d\tilde{\mathbf{a}}. \quad (4.8)$$

The exact evaluation of this Bayesian integral is not possible since we cannot integrate in closed form with respect to *both*  $\mathbf{f}$  and  $\tilde{\mathbf{a}}$ . Instead we marginalize in closed form only with respect to  $\tilde{\mathbf{a}}$ . We chose to marginalize  $\tilde{\mathbf{a}}$  for two reasons. First, because the maximization with respect to  $\mathbf{f}$  that follows is tractable. Second, because this approach requires explicit computation of  $\mathbf{f}$  as part of the Bayesian algorithm and we do not have to compute it separately. More specifically, we have:

$$p(\mathbf{g}, \mathbf{f}; \beta, \mathbf{m}, \mathbf{l}) = \int p(\mathbf{f} | \tilde{\mathbf{a}}) p(\tilde{\mathbf{a}}; \mathbf{m}, \mathbf{l}) d\tilde{\mathbf{a}}. \quad (4.9)$$

The calculation of the integral can be made as follows

$$\begin{aligned}
I &= \int p(\mathbf{f} | \tilde{\mathbf{a}}) p(\tilde{\mathbf{a}}; \mathbf{m}, \mathbf{l}) d\tilde{\mathbf{a}} = \prod_{k=1}^4 \left\{ \int_0^{+\infty} \prod_{i=1}^N p(a_i^k; m_k, l_k) \exp \left\{ -\frac{1}{2} \sum_{i=1}^N (\varepsilon_i^k)^2 a_i^k + \frac{1}{8} \sum_{i=1}^N \log a_i^k \right\} da_k \right\} = \\
&= \prod_{k=1}^4 \left[ \prod_{i=1}^N \Gamma \left( \frac{l_k}{2} \right)^{-1} (m_k (l_k - 2))^{\frac{l_k}{2}} \int_0^{+\infty} \prod_{i=1}^N (a_i^k)^{\frac{l_k}{2} + \frac{1}{8} - 1} \exp \left\{ - \left( m_k (l_k - 2) + \frac{1}{2} ([\mathbf{Q}^k \mathbf{f}](i))^2 \right) a_i^k \right\} da_k \right].
\end{aligned}$$

Observing the integrand in the above expression it is easy to notice it is of the form  $x^a e^{-bx}$ , i.e., it is similar to a Gamma PDF. Thus its integral is given by:

$$I = \prod_{k=1}^4 \left\{ \Gamma \left( \frac{l_k}{2} \right)^{-N} (m_k (l_k - 2))^{\frac{N l_k}{2}} \Gamma \left( \frac{l_k}{2} + \frac{1}{8} \right)^N \prod_{i=1}^N \left( m_k (l_k - 2) + \frac{1}{2} ([\mathbf{Q}^k \mathbf{f}](i))^2 \right)^{-\left( \frac{l_k}{2} + \frac{1}{8} \right)} \right\} \quad (4.10)$$

Replacing the integral of Eq. (4.10) in Eq. (4.9) gives:

$$\begin{aligned}
p(\mathbf{g}, \mathbf{f}; \beta, \mathbf{m}, \mathbf{l}) &= \left( \frac{\beta}{2\pi} \right)^{N/2} \prod_{k=1}^4 \left\{ \Gamma \left( \frac{l_k}{2} \right)^{-N} \Gamma \left( \frac{l_k}{2} + \frac{1}{8} \right)^N (m_k (l_k - 2))^{\frac{N l_k}{2}} \right\} \times \\
&\times \exp \left\{ -\frac{1}{2} \beta \|\mathbf{H}\mathbf{f} - \mathbf{g}\|^2 \right\} \prod_{k=1}^4 \left\{ \prod_{i=1}^N \left( m_k (l_k - 2) + \frac{1}{2} ([\mathbf{Q}^k \mathbf{f}](i))^2 \right)^{-\left( \frac{l_k}{2} + \frac{1}{8} \right)} \right\} \quad (4.11)
\end{aligned}$$

Thus, the image is estimated by its value at the mode of  $p(\mathbf{g}, \mathbf{f}; \beta, \mathbf{m}, \mathbf{l})$ . To compute the mode, we minimize  $J_B(\mathbf{g}, \mathbf{f}; \beta, m, l)$  with respect to  $\mathbf{f}$ ,

$$\mathbf{f}^* = \arg \min_{\mathbf{f}} J_B(\mathbf{g}, \mathbf{f}; \beta, \mathbf{m}, \mathbf{l}),$$

where:

$$\begin{aligned}
J_B(\mathbf{g}, \mathbf{f}^*; \beta, \mathbf{m}, \mathbf{l}) &= -\frac{1}{2} \beta \|\mathbf{H}\mathbf{f}^* - \mathbf{g}\|^2 - \frac{N}{2} \log \beta - \\
&- \sum_{k=1}^4 \left[ N \log \Gamma \left( \frac{l_k}{2} \right) - \frac{N l_k}{2} \log (m_k (l_k - 2)) - N \log \Gamma \left( \frac{l_k}{2} + \frac{1}{8} \right) \right] - \\
&- \sum_{k=1}^4 \left( \left( \frac{l_k}{2} + \frac{1}{8} \right) \sum_{i=1}^N \log \left( m_k (l_k - 2) + \frac{1}{2} ([\mathbf{Q}^k \mathbf{f}^*](i))^2 \right) \right).
\end{aligned}$$

The mode  $\mathbf{f}^*$  is found by an iterative modified Newton algorithm with the following update equation

$$\mathbf{f}^{k+1} = \mathbf{f}^k - \mu^k \nabla_{\mathbf{f}}^2 J_B(\mathbf{f}^k)^{-1} \nabla_{\mathbf{f}} J_B(\mathbf{f}^k), \quad (4.12)$$

where  $\nabla_{\mathbf{f}} J_B$  and  $\nabla_{\mathbf{f}}^2 J_B$  denote the Gradient and the Hessian matrix of the  $J_B$  function respectively (with respect to  $\mathbf{f}$ ). The gradient and the Hessian can be computed analytically, and are given by:

$$\nabla_{\mathbf{f}} J_B = \beta \mathbf{H}^T (\mathbf{H}\mathbf{f} - \mathbf{g}) + \sum_{k=1}^4 \left( \frac{l_k}{2} + \frac{1}{8} \right) (\mathbf{Q}^k)^T \mathbf{v}_k, \quad \nabla_{\mathbf{f}}^2 J_B = \beta \mathbf{H}^T \mathbf{H} + \sum_{k=1}^4 \left( \frac{l_k}{2} + \frac{1}{8} \right) (\mathbf{Q}^k)^T \Lambda_k \mathbf{Q}^k,$$

$$\text{where:} \quad \mathbf{v}_k(i) = \frac{[\mathbf{Q}^k \mathbf{f}](i)}{\left( m_k(l_k - 2) + \frac{1}{2}([\mathbf{Q}^k \mathbf{f}](i))^2 \right)}, \quad i = 1, 2, \dots, N,$$

and the matrix  $\Lambda_k$  is diagonal with diagonal elements  $\lambda_k(i)$  equal to:

$$\lambda_k(i) = \frac{1}{m(l-2) + \frac{1}{2}([\mathbf{Q}^k \mathbf{f}](i))^2} - [\mathbf{v}_k(i)]^2, \quad i = 1, \dots, N.$$

To find the step  $\mu^k$  in (4.12) we adopted a backtracking line search method [97].

#### 4.5 Numerical experiments

In this section we present numerical experiments to evaluate our algorithms. First, we compare the proposed methods with previous methods in terms of the quality of the images provided. Second, we compare the proposed MAP and Bayesian algorithms in terms of the bias and variance of the inferred restored images.

The metrics used to quantify the quality of the degraded images, the noise levels in our degraded images and the quality of our restoration results are the peak signal to noise ratio (*PSNR*), the signal to noise ratio (*SNR*), and the improvement in signal to noise ratio (*ISNR*), respectively. These metrics are defined as:

$$PSNR = 10 \log_{10} \frac{N(255)^2}{\|\mathbf{f} - \mathbf{g}\|^2}, \quad SNR = 10 \log_{10} \frac{\|\mathbf{f}\|^2}{N\beta^{-1}}, \quad \text{and} \quad ISNR = 10 \log_{10} \frac{\|\mathbf{f} - \mathbf{g}\|^2}{\|\mathbf{f} - \hat{\mathbf{f}}\|^2},$$

where,  $\beta^{-1}$ ,  $\mathbf{f}$ ,  $\hat{\mathbf{f}}$  and  $\mathbf{g}$ , are the noise variance, the original, restored and degraded images, respectively.

In our experiments we used the well-known  $256 \times 256$  “Lena” image, shown in Figure 4.2a. The image was blurred by a uniform  $7 \times 7$  PSF (normalized to sum to 1) and white Gaussian noise was added such that  $SNR=25,36dB$ . The degraded image is shown in Figure 4.2b with corresponding  $PSNR$  27.08 and 23.22dB, respectively.

In order to compare the proposed approaches with previous ones we implemented: (i) the classical Wiener filter in the DFT domain [70] using the degraded image to estimate the image power spectrum assuming that the additive noise variance is known. The resulting image is shown in Figure 4.2c. (ii) The classical Wiener filter in the DFT domain [70] using the original image to estimate the image power spectrum and assuming that the additive noise variance is known. Clearly this is not a realistic scenario; however, it compares our algorithm to the performance limit of the Wiener filter. The resulting image is shown in Figure 4.2d. (iii) The Bayesian approach using a stationary SAR prior [92]. The corresponding image and  $ISNR$  are shown in Figure 4.2e. (iv) The iterative constrained least squares (CLS) approach with spatially adaptive regularization [76], [95], [92], [80]. The optimal parameters for this model were found in a trial and error fashion. The resulting image is shown in Figure 4.2f. (v) The non stationary CGMRF based approach that uses a binary line process to model the image edges in [92]. The resulting image is shown in Figure 4.2g.

To facilitate learning the proposed image model we used the  $\beta^{-1}$  (additive noise variance) and equal  $m_k$  that was obtained by learning a stationary SAR model assuming a Laplacian operator  $\mathbf{Q}$  for the residuals [92]. The parameters  $m_k$  were obtained as  $m_k = 1 / (2a_{STAT})$  where  $a_{STAT}$  the image model parameter of the stationary SAR model. The parameters  $l_k$  were selected to be equal to a value denoted by  $l$ . Since, as explained previously they can be used to adjust, the degree of non-stationarity of the image model, values in the interval  $l = [2.1 - 2.5]$  were found using trial and error to provide the best restored images based on both visual criteria and the  $ISNR$  metric. Since both algorithms run very fast (1-3 minutes) and only one parameter is adjusted the trial and error procedure is feasible.

In order to test the performance limits of the proposed model we implemented the MAP approach estimating the model parameters from the original image. The resulting image is shown in Figure 4.2h. The resulting restored images using the proposed methods where all the unknowns are estimated from the observations are shown in Figure 4.2i for MAP and



Figure 4.2j for the Bayesian approach. From the restored images shown in these figures is clear that the proposed non stationary restoration algorithms provide *both* higher *ISNR* and visually more pleasing results than all previous stationary and non-stationary based methods. It is interesting to point out that even when the original image is used to estimate the image statistics, as in the case of the Wiener filter, both proposed approaches outperformed it.

We also tested the proposed algorithms with wavelet based approaches with respect to the *ISNR* metric using the three experiments described in [55]. Although the *ISNR* metric is not always an accurate measure of visual impression it is an objective metric of estimation performance. Our MAP algorithm for the first and third set of experiments in [55] gave better *ISNR*, as shown in Tables 4.1 and 4.2, respectively. In the first experiment the  $256 \times 256$  “Cameraman” image was degraded by additive noise with  $\beta^{-1}=0.308$  or  $SNR= 38.64$  dB, and uniform  $9 \times 9$  blur shown in Figure 4.3a. We also show and provide *ISNRs* for the following cases: i) the stationary restored image assuming an SAR prior [92], in Figure 4.3b. ii) The restored image obtained by the CLS spatially adaptive approach, [76] and [78], in Figure 4.3c and iii) the restored image by the proposed MAP algorithm in Figure 4.3d. In the third experiment described in [55] the  $512 \times 512$  “Lena” image was degraded  $\beta^{-1} = 49$  or  $SNR=16.62$  dB, and separable  $5 \times 5$  blur implemented by blurring with a PSF given by  $[1, 4, 6, 4, 1]/16$  in each direction. For the second experiment in [55] the *ISNR* obtained by the proposed here MAP algorithm was approximately equal to the best one obtained by the methods presented in [55]. In all experiments the same termination criterion was used as in [55]. The proposed Bayesian algorithm in this set of experiments was not as competitive and gave slightly lower *ISNR* than the best case of the results reported in [55].

Finally, in order to compare the properties of the proposed MAP and Bayesian algorithms we considered two metrics, the bias (*BIAS*) and the variance (*VAR*) of the restored images. These metrics were estimated by Monte-Carlo simulations using the following equations

$$BIAS = \|\mathbf{f} - \bar{\mathbf{f}}\|, \quad VAR = \frac{1}{M} \sum_{k=1}^M \|\bar{\mathbf{f}} - \hat{\mathbf{f}}_k\|^2, \quad \text{with } \bar{\mathbf{f}} = \frac{1}{M} \sum_{k=1}^M \hat{\mathbf{f}}_k,$$

where,  $\mathbf{f}$  is the original and  $\hat{\mathbf{f}}_k$  for  $k=1, \dots, M$ , the restored image, obtained from  $M = 10$  different restoration runs in which the degraded images were corrupted with different noise realizations. The results for three images (in addition to “Lena” and “Cameraman” a  $256 \times 256$  segment of the “Barbara” image was also used) at 3 different noise levels are

shown in Tables 4.3 and 4.4, respectively. The blur used here was circular Gaussian shaped with shape parameter  $\sigma^2 = 2$  (normalized to mean equal to 1).

The above experiments demonstrate that the Bayesian approach has a lower variance than the MAP approach, as expected since it marginalizes the directional variances, and does not use point estimates. However, in terms of bias, both MAP and Bayesian algorithms give comparable results.

In terms of computational cost both proposed algorithms were very fast. Typically, our algorithms required about 20 iterations to converge using as criterion the change of the likelihood between successive iterations to be less than 0.1%. Our algorithms were implemented in MATLAB and take about 1-4 minutes on a Pentium 4 at 2.8 GHz personal computer for  $256 \times 256$  images. In contrast, a C implementation of the deterministic relaxation MAP algorithm in [92] required 10-15 minutes on a Xeon 3.2 GHz machine. The constrained least squares method with spatially adaptive constraints [76], [95], [92], [80] was implemented using a conjugate gradient algorithm and is of the same computational complexity, given that the correct parameters have been found, to the proposed methods. The Wiener filter and the Bayesian approach with the stationary SAR model are much faster since all calculations are done in the DFT domain and require 5-10" using MATLAB on a Pentium 4 at 2.8 GHz personal computer.

#### 4.6 Conclusions and extensions

The power of the proposed image prior model was clearly demonstrated with the MAP approach when the original image was used to estimate the model parameters. Apart from the very high visual quality of the restored images, shown in Figure 4.2h, it outperforms in terms of *ISNR* by almost *5dB* the one obtained by the Wiener filter shown in Figure 4.2d, when the original image was also used to estimate the power spectrum. Furthermore, the proposed methods compared favorably with recent CGMRF and wavelet based methods [92] and [55], respectively.

Since the parameters  $l_k$  of the proposed hyperpriors can be viewed as quantifying the degree of non-stationarity of the image model, developing an algorithm to estimate the parameters in a rigorous manner seems a natural extension to this work. Thus, a more efficient algorithm than the Bayesian algorithm presented in this chapter can be derived, using for example the

variational methodology for approximate Bayesian inference and estimate the parameters by maximizing (approximately) the marginal likelihood, obtained by integrating with respect to both the image and variances. In contrast, the proposed Bayesian algorithm of this chapter maximizes a likelihood obtained by marginalizing the variances and not the image. In Chapter 5, we focus on estimating the model parameters using the VB methodology presented in Chapter 3.

Finally, the four directional difference operators may not be able to capture the salient features of the image, and hence, additional operators should be included to the model. To this end, filters from known transforms, such as the wavelet discrete transform and other filter banks can be incorporated as operators.

Table 4.1: *ISNR* comparisons with the experiments in Table I of [55].

<b>Method</b>	<b>ISNR</b>
Proposed MAP approach, $l=2.1$	8.78dB
Rule (22), UDWT [55]	7.47dB
Rule (22), Random shifts [55]	7.59dB
Modified Laplacian, UDWT [55]	7.26dB
Modified Laplacian, random shifts [55]	7.34dB
Shoft-threshold, UDWT [55]	7.26dB
Shoft-threshold, random shifts [55]	6.33dB
Result by Neelamani <i>et al</i> [100]	7.3dB
Result by Banham and Katsaggelos [9]	6.7dB

Table 4.2: *ISNR* comparison with experiments in Table III of [55].

Method	ISNR
Proposed MAP approach, $l=2.1$ ,	3.63dB
Rule (22), UDWT [55]	2.94dB
Rule (22), random shifts [55]	1.71dB
Modified Laplacian, UDWT [55]	2.75dB
Modified Laplacian, random shifts [55]	1.77dB
Shoft-threshold, UDWT [55]	2.75dB
Shoft-threshold, random shifts [55]	1.61dB
Best result by Liu and Moulin [88]	1.078dB

Table 4.3: *Bias* metric for the MAP and the Bayesian algorithms.

Noise Level	<u>SNR=65.34 dB</u>		<u>SNR =25.36 dB</u>		<u>SNR =14.91 dB</u>	
	Bayesian	MAP	Bayesian	MAP	Bayesian	MAP
Cameraman	9.36	9.15	12.23	11.57	15.22	14.08
Lena	5.31	5.31	6.92	7.44	9.29	10.71
Barbara	5.33	6.97	13.42	11.94	16.42	14.25

Table 4.4: Variance metric for the MAP and the Bayesian algorithms.

Noise Level	<u>SNR =65.34 dB</u>		<u>SNR =25.36 dB</u>		<u>SNR =14.91 dB</u>	
Algorithm	Bayesian	MAP	Bayesian	MAP	Bayesian	MAP
Cameraman	1.77e-004	1.77e-004	1.64e-004	3.60e-004	2.37e-004	6.49e-004
Lena	6.47e-005	9.16e-005	1.04e-004	1.38e-004	1.71e-004	5.25e-004
Barbara	1.69e-004	2.89e-004	1.62e-004	6.02e-004	1.88e-004	8.95e-004



*Figure 4.2a:* Original “Lena” image.



*Figure 4.2b:* Degraded “Lena” image with  $7 \times 7$  uniform blur and additive noise  $SNR=25.36$  dB.



*Figure 4.2c:* Wiener filter restoration,  $ISNR = 3.2$  dB.



*Figure 4.2d:* "Optimal" Wiener filter restoration,  $ISNR = 4.40$  dB.



*Figure 4.2e:* Stationary restoration,  $ISNR = 4.25$  dB.



*Figure 4.2f:* CLS method (adaptive smoothness constraint) restoration,  $\theta=1000$ ,  $a=0.01$ ,  $ISNR = 4.65$  dB.



*Figure 4.2g:* Restoration with GMRF algorithm [92],  $ISNR= 3.46$  dB.



*Figure 4.2h:* MAP “optimal” non stationary restoration,  $ISNR=10.43$  dB,  $l=2.01$ .



*Figure 4.2i:* MAP non stationary restoration,  $ISNR= 5.63$  dB,  $l=2.2$ .



*Figure 4.2j:* Bayesian non stationary restoration,  $ISNR=5.22$  dB,  $l=2.2$ .

Figure 4.2: Experiments with ‘Lena’ image, Gaussian blur and  $SNR=25,36$ dB.



*Figure 4.3a:* Degraded “cameraman” image with  $9 \times 9$  uniform blur and additive noise  $SNR=38.64$  dB, ( $\beta^{-1}=0.308$ ).



*Figure 4.3b:* Stationary restoration,  $ISNR = 6.44$  dB.



*Figure 4.3c:* CLS method (adaptive smoothness constraint) restoration,  $\theta=0.05$ ,  $\alpha=0.003$ ,  $ISNR = 7.22$  dB.



*Figure 4.3d:* MAP non stationary restoration,  $ISNR = 8.78$  dB,  $l=2.1$ .

Figure 4.3: Experiments with ‘Cameraman’ image, uniform blur and  $SNR = 38.64$  dB.



# **CHAPTER 5. BAYESIAN IMAGE RESTORATION WITH A PRIOR BASED ON A PRODUCT OF $t$ - DISTRIBUTIONS**

---

- 5.1. Introduction
  - 5.2. Imaging and image model
  - 5.3. Variational Inference
  - 5.4. Computational implementation
  - 5.5. Experiments
  - 5.6. Conclusions and suggestions for improvement
- 

In this chapter we extend the image prior introduced in Chapter 4 and propose a variational Bayesian restoration algorithm that estimates the parameters of the image model and bypasses some major difficulties of the algorithms proposed in Chapter 4. The prior is extended to involve an arbitrary number of convolutional operators, and not just the directional differences operators. Thus, the image prior is defined by assuming Student's- $t$  densities on the outputs of local convolutional filters. The resulting prior is in product form similar to the prior of Chapter 4. Priors based on products have been recognized to offer many advantages because they allow for simultaneous enforcement of multiple constraints. However, they are inconvenient for Bayesian inference because it is hard to find their normalization constant in closed form. In this chapter, a new Bayesian algorithm is proposed for the image restoration problem which bypasses this difficulty. Another difficulty is the maximization of the marginal likelihood which is intractable to obtain. This difficulty is bypassed by employing the variational methodology for approximate inference, with a constrained expectation step,

which is used to infer the restored image. Numerical experiments are shown that compare this methodology to previous ones and demonstrate its advantages. [26].

## 5.1 Introduction

Product-based image priors have been presented in section 4.3.6. Such priors combine in product form multiple probabilistic models. Each individual model gives high probability to data vectors that satisfy just one constraint. Vectors that satisfy only this constraint but violate others are ruled out by their low probability under the other terms of the product model. However, such priors are usually learned using a large training set of images and stochastic sampling methods and used in a number of image recovery problems based on “empirical” maximum a posteriori approaches and gradient descent minimization [114]. This differs from the herein proposed approach where the product prior is learnt only from the observations. The term “empirical” is used because the PoE priors used are not normalized; thus, the parameters of the recovery algorithm cannot be estimated or inferred rigorously but are adjusted rather empirically.

Extending the work in Chapter 4, we propose in this chapter a new Bayesian inference framework for image deconvolution using a prior in product form. This prior assumes that the outputs of local high-pass filters are Student’s-t distributed. The main contribution of this chapter is a Bayesian inference methodology that bypasses the difficulty of evaluating the normalization constant of product type priors. The methodology is based on a *constrained* variational approximation that uses the outputs of all the local high pass filters to produce an estimate of the original image. More specifically, a *constrained expectation step* is used to capture the relationship of the filter outputs of the prior to the original image. In this manner the use of improper priors is avoided and *all* the parameters of the prior model are estimated from the data. Thus, we avoid the “trial and error” parameter “tweaking” required in Chapter 4 and other state-of-the-art recently proposed restoration algorithms, which makes their use difficult for non-experts. Furthermore, the proposed restoration algorithm provides competitive performance compared to other methods.

In this chapter we also propose an efficient Lanczos-based computational framework tailored to the calculations required in our Bayesian algorithm. More specifically, a very large linear

system  $Ax = b$  is solved iteratively and the diagonal elements of a matrix  $Q^t A^{-1} Q$  are simultaneously estimated in an efficient manner.

The rest of this paper is organized as follows. In section 5.2 the imaging and image models are defined. In section 5.3 the variational restoration algorithm is derived. In section 5.4 we present the computational methodology used to implement our algorithm, while in section 5.5 numerical experiments are demonstrated. Finally, section 5.6 contains conclusions and directions for future work.

## 5.2 Imaging and Image Model

A linear imaging model is assumed. For convenience but without loss of generality, we use one-dimensional notation. The  $N \times 1$  vector  $\mathbf{g}$  represents the observed degraded image obtained by

$$\mathbf{g} = \mathbf{H}\mathbf{f} + \mathbf{n}, \quad (5.1)$$

where  $\mathbf{f}$  is the (unknown) original image,  $\mathbf{H}$  is an  $N \times N$  known convolution matrix and  $\mathbf{n}$  is additive white noise. We assume Gaussian statistics for the noise given by  $\mathbf{n} \sim N(\mathbf{0}, \beta^{-1}\mathbf{I})$  where  $\mathbf{0}$  is an  $N \times 1$  vector of zeros,  $\mathbf{I}$  is the  $N \times N$  identity matrix and  $\beta$  is the noise precision (inverse variance), which is assumed unknown.

Aiming at the definition of the image prior we first define  $P$  operators  $\mathbf{Q}_k$  for  $k = 1, \dots, P$  and use them to define  $P$  filter outputs:

$$\boldsymbol{\varepsilon}_k = \mathbf{Q}_k \mathbf{f}, \quad k = 1, \dots, P, \quad (5.2)$$

where  $\boldsymbol{\varepsilon}_k = [\varepsilon_k(1), \varepsilon_k(2), \dots, \varepsilon_k(N)]^T$ . The matrices  $\mathbf{Q}_k$  representing the operators are of size  $N \times N$  and the filter outputs  $\boldsymbol{\varepsilon}_k$  are of size  $N \times 1$ . These operators are zero mean convolutional high-pass filters and each one of them is used to impose a particular constraint on the restored image.

We assume that  $\varepsilon_k(i)$  for  $i = 1, \dots, N$  are i.i.d zero mean Student's-t distributed, with parameters  $\lambda_k$  and  $\nu_k$ :

$$\varepsilon_k(i) \sim St(\varepsilon_k(i); 0, \lambda_k, \nu_k), \quad \forall i, \forall k, \quad (5.3)$$

where

$$St(x; 0, \lambda_k, \nu_k) = \frac{\Gamma(\nu_k/2 + 1/2)}{\Gamma(\nu_k/2)} \left( \frac{\lambda_k}{\pi \nu_k} \right)^{1/2} \left( 1 + \frac{\lambda_k}{\nu_k} x^2 \right)^{-\frac{\nu_k+1}{2}}.$$

The Student's-t implies a two-level generation process [19]. More specifically,  $a_k(i)$  is first drawn from a Gamma distribution,  $p(a_k(i)) = \text{Gamma}\left(a_k(i); \frac{\nu_k}{2}, \frac{\nu_k}{2}\right)$ . Then, the  $\varepsilon_k(i)$  is generated from a zero-mean Normal distribution with precision  $\lambda_k a_k(i)$ , according to  $p(\varepsilon_k(i) | a_k(i)) = N\left(\varepsilon_k(i); 0, (\lambda_k a_k(i))^{-1}\right)$ . The probability density function of Eq. (5.3) can be written as an integral:

$$p(\varepsilon_k(i)) = St(\varepsilon_k(i); 0, \lambda_k, \nu_k) = \int_0^{+\infty} p(\varepsilon_k(i) | a_k(i)) p(a_k(i)) da_k(i).$$

The variables  $a_k(i)$  are called “hidden” (latent) because they are not apparent in Eq. (5.3), since they have been integrated out. There are two extremes in this generative model, depending on the value of the “degree of freedom” parameter  $\nu_k$ . As this parameter goes to infinity, the pdf from which the  $a_k(i)$ 's are drawn has its mass concentrated around 1. This in turn reduces the Student's-t to a Normal distribution, because all  $\varepsilon_k(i)$  are drawn from the same Normal with precision  $\lambda_k$ , since  $a_k(i) = 1$ . The other extreme is when  $\nu_k \rightarrow 0$  and the prior becomes uninformative. In general, for small values of  $\nu_k$  the probability mass of the Student's-t pdf is spread, rendering the Student's-t more “heavy-tailed”.

The use of heavy-tailed priors on high-pass filters of the image is a characteristic of most modern “edge preserving” image priors used for regularization in a stochastic setting; see for example [21], [117], [104], [8], [66], [114] and [130]. The main idea behind this assumption is that at the few edge areas of an image the filter outputs  $\varepsilon_k(i)$  will be large in absolute value. Thus, it is important to model them with a heavy-tailed pdf in order to allow the prior to encourage formation of edges. The downside of many such models is that most heavy-tailed pdfs are not amenable to Bayesian inference. For example the Generalized Gaussian and the Alpha Stable pdfs can be also heavy tailed. However, unlike the Student's-t where

Bayesian inference is possible [89], moment-based estimators have to be used for their parameters; see for example [101] and [42].

We now define the following notation for the variables  $a_k(i)$ . We denote by  $\tilde{\mathbf{a}} = [\mathbf{a}_1, \dots, \mathbf{a}_P]^\top$  a  $PN \times 1$  vector, where  $\mathbf{a}_k = [a_k(1), a_k(2), \dots, a_k(N)]$ . Also, for the filter outputs we use the notation  $\tilde{\boldsymbol{\varepsilon}} = [(\boldsymbol{\varepsilon}_1)^\top, (\boldsymbol{\varepsilon}_2)^\top, \dots, (\boldsymbol{\varepsilon}_P)^\top]^\top$ . We assume that the filter outputs are independent not only in each pixel location but also in each direction. This assumption makes subsequent calculations tractable. Thus, the cumulative density for the filter outputs conditioned on  $\tilde{\mathbf{a}}$  is

$$p(\tilde{\boldsymbol{\varepsilon}} | \tilde{\mathbf{a}}) = \prod_{k=1}^P p(\boldsymbol{\varepsilon}_k | \mathbf{a}_k), \quad (5.4)$$

where  $p(\boldsymbol{\varepsilon}_k | \mathbf{a}_k) = N(\mathbf{0}, (\lambda_k \mathbf{A}_k)^{-1})$  and  $\mathbf{A}_k$  is a diagonal matrix with elements the components of the vector  $\mathbf{a}_k$ .

At this point, the marginal distribution  $p(\mathbf{f})$  yearns for a closed form, using the relation between the image and the filter outputs, Eq. (5.2). However, this prior is analytically intractable because *one cannot find in closed form its normalization constant*. This problem stems from the fact that it is not possible to find the eigenvalues of the matrix  $\sum_{k=1}^P \mathbf{Q}_k^\top \mathbf{A}_k \mathbf{Q}_k$

because it is very large and the product  $\mathbf{Q}_k^\top \mathbf{A}_k \mathbf{Q}_k$  does not have a structure that is amenable to efficient eigenvalue computation. One contribution of this work is that we bypass this difficulty by exploiting the commuting property of convolutional operators and derive a constrained variational algorithm for approximate Bayesian inference. This algorithm is described in detail next.

### 5.3 Variational Inference

Since, as explained above, it is difficult to infer a solution for the image from the Bayesian model previously defined, a transformed imaging model is introduced in section 6.3.1.

#### 5.3.1 The Variational Algorithm for Equivalent Imaging Model

The imaging model of Eq. (2.1) can be written as

$$\mathbf{Q}_k \mathbf{g} = \mathbf{Q}_k \mathbf{H} \mathbf{f} + \mathbf{Q}_k \mathbf{n} \text{ for } k = 1, \dots, P. \quad (5.5)$$

Setting  $\mathbf{y}_k = \mathbf{Q}_k \mathbf{g}$  for  $k = 1, \dots, P$  and using Eq. (5.2), we can utilize the *commuting property* of the convolutional operators and write the imaging model as

$$\mathbf{y}_k = \mathbf{H} \boldsymbol{\varepsilon}_k + \mathbf{n}_k \text{ for } k = 1, \dots, P, \quad (5.6)$$

where  $\mathbf{y}_k$  are the observations of the newly defined model and the additive noise is

$$\mathbf{n}_k \sim N\left(0, \beta^{-1} \mathbf{Q}_k \mathbf{Q}_k^T\right).$$

In this model we assume that the filter outputs  $\boldsymbol{\varepsilon}_k$  of our filters  $\mathbf{Q}_k$  are the unknowns. Thus, the algorithm will infer  $\tilde{\boldsymbol{\varepsilon}}$  instead of  $\mathbf{f}$ . In this manner we bypass the need to define a prior for  $\mathbf{f}$ . For this reason, we must initially define the posterior of the observations  $\tilde{\mathbf{y}}$  given  $\tilde{\boldsymbol{\varepsilon}}$ . This is equal to the product of  $P$  Normal distributions, since the observations are assumed independent:

$$p(\tilde{\mathbf{y}} | \tilde{\boldsymbol{\varepsilon}}) = \prod_{k=1}^P p(\mathbf{y}_k | \boldsymbol{\varepsilon}_k), \text{ where } \tilde{\mathbf{y}} = \left[ (\mathbf{y}_1)^T, (\mathbf{y}_2)^T, \dots, (\mathbf{y}_P)^T \right]^T \text{ and}$$

$$p(\mathbf{y}_k | \boldsymbol{\varepsilon}_k) = N\left(\mathbf{H} \boldsymbol{\varepsilon}_k, \left(\beta \mathbf{Q}_k^T \mathbf{Q}_k\right)^{-1}\right) \text{ for } k = 1, \dots, P.$$

The prior for the residuals has been already defined in Eq. (5.3).

Working in the Bayesian framework, we define as latent (hidden) variables the residuals  $\tilde{\boldsymbol{\varepsilon}}$  and the inverse variances  $\tilde{\mathbf{a}}$ . Hence, the complete data likelihood is

$$p(\tilde{\mathbf{y}}, \tilde{\boldsymbol{\varepsilon}}, \tilde{\mathbf{a}}; \theta) = p(\tilde{\mathbf{y}} | \tilde{\boldsymbol{\varepsilon}}; \theta) p(\tilde{\boldsymbol{\varepsilon}} | \tilde{\mathbf{a}}; \theta) p(\tilde{\mathbf{a}}; \theta),$$

where  $\theta = [\beta, \nu_1, \dots, \nu_P, \lambda_1, \dots, \lambda_P]^T$ .

Estimation of the model parameters ideally could be obtained through maximization of the marginal distribution of the observations  $p(\tilde{\mathbf{y}}; \theta)$ :

$$\hat{\theta} = \arg \max_{\theta} \int \int p(\tilde{\mathbf{y}}, \tilde{\boldsymbol{\varepsilon}}, \tilde{\mathbf{a}}; \theta) d\tilde{\boldsymbol{\varepsilon}} d\tilde{\mathbf{a}}. \quad (5.7)$$

However, in the present case this marginalization is not possible. Furthermore, since the posterior of the hidden variables given the observations  $p(\tilde{\boldsymbol{\epsilon}}, \tilde{\mathbf{a}} | \tilde{\mathbf{y}})$  is not known explicitly, inference via the Expectation-Maximization (EM) algorithm is not possible [10].

For this reason, we resort to the variational methodology [19], [10] and [87]. According to this methodology, we introduce a lower bound on the logarithm of the marginal likelihood, which is actually the expectation of the logarithm of the complete data likelihood with respect to an auxiliary function of the hidden variables  $q(\tilde{\boldsymbol{\epsilon}}, \tilde{\mathbf{a}})$  minus the entropy of  $q(\tilde{\boldsymbol{\epsilon}}, \tilde{\mathbf{a}})$ :

$$\log p(\tilde{\mathbf{y}}; \theta) \geq L(q(\tilde{\boldsymbol{\epsilon}}, \tilde{\mathbf{a}}), \theta)$$

$$L(q(\tilde{\boldsymbol{\epsilon}}, \tilde{\mathbf{a}}), \theta) \equiv \int q(\tilde{\boldsymbol{\epsilon}}, \tilde{\mathbf{a}}) \log p(\tilde{\mathbf{y}}, \tilde{\boldsymbol{\epsilon}}, \tilde{\mathbf{a}}; \theta) d\tilde{\boldsymbol{\epsilon}} d\tilde{\mathbf{a}} - \int q(\tilde{\boldsymbol{\epsilon}}, \tilde{\mathbf{a}}) \log q(\tilde{\boldsymbol{\epsilon}}, \tilde{\mathbf{a}}) d\tilde{\boldsymbol{\epsilon}} d\tilde{\mathbf{a}}. \quad (5.8)$$

The inequality holds because the functional  $L$  is also equal to the logarithm of the marginal likelihood minus the always non-negative Kullback-Leibler divergence between the true posterior distribution  $p(\tilde{\boldsymbol{\epsilon}}, \tilde{\mathbf{a}} | \tilde{\mathbf{y}}; \theta)$  of the hidden variables and  $q(\tilde{\boldsymbol{\epsilon}}, \tilde{\mathbf{a}})$ ; see for example [19].

Equality holds in Eq. (5.8) when  $q(\tilde{\boldsymbol{\epsilon}}, \tilde{\mathbf{a}}) = p(\tilde{\boldsymbol{\epsilon}}, \tilde{\mathbf{a}} | \tilde{\mathbf{y}}; \theta)$ , or equivalently

$$q(\tilde{\boldsymbol{\epsilon}}, \tilde{\mathbf{a}}) = \prod_{k=1}^P q(\boldsymbol{\epsilon}_k, \mathbf{a}_k) = \prod_{k=1}^P p(\boldsymbol{\epsilon}_k, \mathbf{a}_k | \mathbf{y}_k; \theta_k), \quad (5.9)$$

because in this case the Kullback-Leibler divergence becomes zero.

In the variational Bayesian framework, instead of maximizing the unobtainable marginal likelihood, we maximize the bound  $L$ , Eq. (5.8), with respect to both  $q(\tilde{\boldsymbol{\epsilon}}, \tilde{\mathbf{a}})$  and  $\theta$  in the variational E and M steps, respectively. In other words, the unknown posterior  $p(\tilde{\boldsymbol{\epsilon}}, \tilde{\mathbf{a}} | \tilde{\mathbf{y}}; \theta)$  is approximated by  $q(\tilde{\boldsymbol{\epsilon}}, \tilde{\mathbf{a}})$ . One difficulty in this approach is that the maximization with respect to  $q(\tilde{\boldsymbol{\epsilon}}, \tilde{\mathbf{a}})$  is hard to obtain in closed form, although, we can bypass it by using the so-called Mean Field approximation [10]. According to this approximation, if we assume that

$$q(\boldsymbol{\epsilon}_k, \mathbf{a}_k) = q(\boldsymbol{\epsilon}_k) q(\mathbf{a}_k), \text{ for } k = 1, \dots, P, \quad (5.10)$$

then unconstrained optimization of the functional  $L(q(\tilde{\boldsymbol{\epsilon}}, \tilde{\mathbf{a}}), \theta)$  with respect to all  $q(\boldsymbol{\epsilon}_k)$  yields  $P$  Normal distributions:

$$q(\boldsymbol{\varepsilon}_k) = N(\mathbf{Q}_k \boldsymbol{\mu}_k, \mathbf{Q}_k \boldsymbol{\Sigma}_k \mathbf{Q}_k^T), \text{ for } k = 1, \dots, P, \quad (5.11)$$

with parameters  $\boldsymbol{\mu}_k = \beta \boldsymbol{\Sigma}_k \mathbf{H}^T \mathbf{g}$  and  $\boldsymbol{\Sigma}_k = (\beta \mathbf{H}^T \mathbf{H} + \lambda_k \mathbf{Q}_k^T \mathbf{A}_k \mathbf{Q}_k)^{-1}$ .

The difficulty that we encounter with the above posteriors, which were obtained by unconstrained optimization, is that they do not provide a method to infer  $\mathbf{f}$  from  $\boldsymbol{\varepsilon}_k$ , and they do not capture their common origin from  $\mathbf{f}$ , Eq. (5.2).

In order to bypass this difficulty we make the assumption that each of the posteriors  $q(\boldsymbol{\varepsilon}_k)$  is Normal; however, it is *constrained* so that it captures the common origin of all  $\boldsymbol{\varepsilon}_k$  from  $\mathbf{f}$ , as dictated by Eq. (5.2). In other words we assume that

$$q(\boldsymbol{\varepsilon}_k; \mathbf{m}, \mathbf{R}) = N(\mathbf{Q}_k \mathbf{m}, \mathbf{Q}_k \mathbf{R} \mathbf{Q}_k^T), \text{ for } k = 1, \dots, P, \quad (5.12)$$

where  $\mathbf{m}$  and  $\mathbf{R}$  are actually parameters representing the mean and covariance of the image  $\mathbf{f}$ , from which all  $\boldsymbol{\varepsilon}_k$  originate. In other words:

$$E[\boldsymbol{\varepsilon}_k] = \mathbf{Q}_k E[\mathbf{f}] = \mathbf{Q}_k \mathbf{m}, \quad Cov[\boldsymbol{\varepsilon}_k] = \mathbf{Q}_k Cov[\mathbf{f}] \mathbf{Q}_k^T = \mathbf{Q}_k \mathbf{R} \mathbf{Q}_k^T.$$

Thus,  $\mathbf{m}$  and  $\mathbf{R}$  are parameters that are used in our model and estimated during the restoration algorithm. Actually, the restored image is taken to be the estimate of  $\mathbf{m}$ .

### 5.3.2 The Variational Update Equations

The general variational algorithm using the Mean Field approximation [10] for approximate inference of a statistical model with  $\mathbf{y}$  as observation,  $n$  hidden variables  $\mathbf{x} = [\mathbf{x}_1, \dots, \mathbf{x}_n]$  and parameters denoted by  $\theta$ , aims to maximize the bound

$$L(q(\mathbf{x}), \theta) \equiv \int \prod_{i=1}^n q_i(\mathbf{x}) \log p(\mathbf{y}, \mathbf{x}; \theta) d\mathbf{x} - \int \prod_{i=1}^n q_i(\mathbf{x}_i) \log q(\mathbf{x}_i) d\mathbf{x}_i.$$

This is achieved by iterating between the two following steps, where  $(t)$  is the iteration index:

$$\text{VE-step: } q_i^{(t+1)}(\mathbf{x}) = \arg \max_{q_i(\mathbf{x})} L(q(\mathbf{x}), \theta^{(t)}), i = 1, \dots, n,$$



$$\text{VM-step: } \theta^{(t+1)}(\mathbf{x}) = \arg \max_{\theta} L(q^{(t+1)}(\mathbf{x}), \theta).$$

Thus, in the E-step of the variational algorithm, optimization of the functional is performed with respect to the auxiliary functions. However, in the present case, the functions  $q(\boldsymbol{\varepsilon}_k)$ ,  $k=1, \dots, P$ , are assumed to be Normal distributions with partially common mean and covariance; see Eq. (5.12); therefore, this bound is actually a function of the parameters  $\mathbf{R}$  and  $\mathbf{m}$  and a functional w.r.t. the auxiliary function  $q(\tilde{\mathbf{a}})$ . Using Eq. (5.10), the variational bound in our problem becomes

$$\begin{aligned} L(q(\tilde{\mathbf{a}}), \theta_1, \theta_2) &= \int \prod_{k=1}^P q(\boldsymbol{\varepsilon}_k; \theta_1) q(\mathbf{a}_k) \log p(\tilde{\mathbf{y}}, \tilde{\boldsymbol{\varepsilon}}, \tilde{\mathbf{a}}; \theta_2) d\tilde{\boldsymbol{\varepsilon}} d\tilde{\mathbf{a}} \\ &- \int \prod_{k=1}^P q(\boldsymbol{\varepsilon}_k; \theta_1) q(\mathbf{a}_k) \log \prod_{k=1}^P q(\boldsymbol{\varepsilon}_k; \theta_1) q(\mathbf{a}_k) d\tilde{\boldsymbol{\varepsilon}} d\tilde{\mathbf{a}}, \end{aligned} \quad (5.13)$$

where  $\theta_1 = [\mathbf{R}, \mathbf{m}]$  and  $\theta_2 = [\beta, \lambda_1, \dots, \lambda_P, \nu_1, \dots, \nu_P]^T$ . Thus, in the VE-step of our algorithm the bound must be optimized with respect to  $\mathbf{R}$ ,  $\mathbf{m}$  and  $q(\tilde{\mathbf{a}})$ :

$$\text{VE-step: } [q^{(t+1)}(\tilde{\mathbf{a}}), \theta_1^{(t+1)}] = \arg \max_{[q(\tilde{\mathbf{a}}), \theta_1]} L(q(\tilde{\mathbf{a}}), \theta_1, \theta_2^{(t)}).$$

Taking the derivative of  $L$  w.r.t to  $\mathbf{m}$ ,  $\mathbf{R}$  and  $q(\tilde{\mathbf{a}})$  (see Appendix), we find that the bound is maximized w.r.t. these parameters when

$$\mathbf{m}^{(t+1)} = \beta^{(t)} \mathbf{R}^{(t)} \mathbf{H}^T \mathbf{g}, \quad (5.14)$$

$$\mathbf{R}^{(t+1)} = \left( \beta^{(t)} \mathbf{H}^T \mathbf{H} + \frac{1}{P} \sum_{k=1}^P \lambda_k \mathbf{Q}_k^T \hat{\mathbf{A}}_k^{(t)} \mathbf{Q}_k \right)^{-1}, \quad (5.15)$$

$$q^{(t+1)}(a_k(i)) = \text{Gamma} \left( a_k(i); \frac{\nu_k}{2} + \frac{1}{2}, \frac{\nu_k}{2} + \frac{1}{2}, \lambda_k \left( (\mathbf{m}_k^{(t)}(i))^2 + \mathbf{C}_k^{(t)}(i, i) \right) \right), \forall k, \forall i, \quad (5.16)$$

where  $\mathbf{m}_k^{(t)} = \mathbf{Q}_k \mathbf{m}^{(t)}$  and  $\mathbf{C}_k^{(t)} = \mathbf{Q}_k \mathbf{R}^{(t)} \mathbf{Q}_k^T$ . Notice that since each  $q^{t+1}(a_k(i))$  is a Gamma pdf of the form  $q^{t+1}(a_k^{(t+1)}(i)) = \text{Gamma}(a_k^{(t+1)}(i); \alpha, \beta)$ , its expected value is

$$\langle a_k(i) \rangle_{q^{t+1}(a_k(i))} = \frac{\alpha}{\beta} = (\nu_k + 1) \left( \nu_k + \lambda_k \left( (\mathbf{m}_k^{(t)}(i))^2 + \mathbf{C}_k^{(t)}(i, i) \right) \right)^{-1}, \quad (5.17)$$

where  $\langle \cdot \rangle_{q(\cdot)}$  denotes the expectation w.r.t. an arbitrary distribution  $q(\cdot)$ . This is used in Eq.

(5.14) and (5.15), where  $\hat{\mathbf{A}}_k^{(t)}$  is a diagonal matrix with elements

$$\hat{\mathbf{A}}_k^{(t)}(i, i) = \langle a_k(i) \rangle_{q^{(t)}(a_k(i))}, i = 1, \dots, N.$$

At the variational M-step the bound is maximized with respect to the model parameters:

$$\text{VM-step: } \theta_2^{(t+1)} = \arg \max_{\theta_2} L(q^{(t+1)}(\tilde{\mathbf{a}}), \theta_1^{(t+1)}, \theta_2),$$

where  $L(q^{(t+1)}(\tilde{\mathbf{a}}), \theta_1^{(t+1)}, \theta_2) \propto \langle \log p(\tilde{\mathbf{y}}, \tilde{\mathbf{e}}, \tilde{\mathbf{a}}; \theta_2) \rangle_{q(\tilde{\mathbf{e}}; \theta_1^{(t+1)}), q^{(t+1)}(\tilde{\mathbf{a}})}$  is calculated using the results from Eq. (5.14)-(5.17).

The update for  $\beta$  is obtained after taking the derivative and equating to zero:

$$\beta^{(t+1)} = N \left( \left\| \mathbf{Hm}^{(t+1)} - \mathbf{g} \right\|_2^2 + \text{trace} \left\{ \mathbf{H}^T \mathbf{H} \mathbf{R}^{(t+1)} \right\} \right)^{-1}. \quad (5.18)$$

In the same way the maximum is attained for  $\lambda_k$ :

$$\lambda_k^{(t+1)} = N \left( \sum_{i=1}^N \left( \left( \mathbf{m}_k^{(t+1)}(i) \right)^2 + \mathbf{C}_k^{(t+1)}(i, i) \right) \langle a_k(i) \rangle_{q^{(t+1)}(a_k(i))} \right)^{-1}. \quad (5.19)$$

Finally, taking the derivative with respect to  $\nu_k$  and equating to zero, we find the ‘‘degrees of freedom’’ parameter of the Student’s-t by solving the equation

$$\begin{aligned} & \frac{1}{N} \left( \sum_{i=1}^N \log \langle a_k(i) \rangle_{q^{(t+1)}(\tilde{\mathbf{a}})} - \sum_{i=1}^N \langle a_k(i) \rangle_{q^{(t+1)}(\tilde{\mathbf{a}})} \right) + \psi \left( \nu_k^{(t)} \frac{1}{2} + \frac{1}{2} \right) - \\ & - \log \left( \nu_k^{(t)} \frac{1}{2} + \frac{1}{2} \right) - \psi \left( \frac{\nu_k}{2} \right) + \log \left( \frac{\nu_k}{2} \right) + 1 = 0 \end{aligned} \quad (5.20)$$

for  $\nu_k$ , where

$$\psi(x) = \frac{d}{dx} \log \Gamma(x) = \frac{\Gamma'(x)}{\Gamma(x)}$$

is the digamma function and  $\nu_k^{(t)}$  is the value of  $\nu_k$  at the previous iteration ( $t$ ) used to evaluate the expectations in Eq. (5.17) during the VE-step.

## 5.4 Computational Implementation

In our implementation, the variance of the additive noise is estimated in a preprocessing step and is kept fixed. The EM algorithm with a stationary Gaussian prior [94] and one output (the Laplacian operator) was used for this purpose. Furthermore, the EM-restored image was used to initialize our algorithm. For all experiments, four filter outputs  $P=4$  were used for the prior. We show the magnitude of the frequency responses of these filters in Fig. 5.2. The operators  $\mathbf{Q}_1$  and  $\mathbf{Q}_2$  correspond to the horizontal and vertical first order differences. Thus, these filters are used to model the vertical and horizontal image edge structure, respectively. The other two operators  $\mathbf{Q}_3$  and  $\mathbf{Q}_4$  are used to model the diagonal edge component contained in the vertical and horizontal edges, respectively. These filters are obtained by convolving the previous horizontal and vertical first order differences filters with fan filters with vertical and horizontal pass-bands, respectively. In our experiments the fan filters in [37] were used.

We solve equations (5.14) and (5.20) iteratively. For Eq. (5.20) we employ the bisection method, as also proposed in [89]. In the next few paragraphs we analyze how Eq. (5.14) is solved by a method based on the Lanczos process [107].

Omitting the subscripts  $k$  and superscripts  $(t)$  for convenience, we regard (5.9) as the linear system  $\mathbf{A}\mathbf{x} = \mathbf{b}$ , where  $\mathbf{A} = \mathbf{R}^{-1}$  is symmetric and positive definite,  $\mathbf{b} = \beta\mathbf{H}^T\mathbf{g}$ ,  $\mathbf{x} = \mathbf{m}$ , and products  $\mathbf{A}\mathbf{v}$  can be obtained efficiently for any given  $\mathbf{v}$ . In addition, we have the linear algebra problem of estimating the diagonals of matrix  $\mathbf{C} = \mathbf{Q}\mathbf{A}^{-1}\mathbf{Q}^T$  in Eq. (5.17). The matrix  $\mathbf{A} = \mathbf{R}^{-1}$  is very large; for example for  $256 \times 256$  images it is of dimension  $N \times N$  with  $N = 65,536$  and clearly an iterative method must be used.

The Lanczos process is an iterative procedure for transforming  $\mathbf{A}$  to tridiagonal form [63]. Given some starting vector  $\mathbf{b}$ , it generates vectors  $\{\mathbf{v}_n\}$  and scalars  $\{\alpha_n, \beta_n\}$  as follows:

1. Set  $\beta_1\mathbf{v}_1 = \mathbf{b}$  (meaning  $\beta_1 = \|\mathbf{b}\|_2$  and  $\mathbf{v}_1 = \mathbf{b} / \beta_1$  but exit if  $\beta_1 = 0$ ).
2. For  $n=1,2,\dots$ , set  $\mathbf{w} = \mathbf{A}\mathbf{v}_n$ ,  $\alpha_n = \mathbf{v}_n^T\mathbf{w}$ ,  $\beta_{n+1}\mathbf{v}_{n+1} = \mathbf{w} - \alpha_n\mathbf{v}_n - \beta_n\mathbf{v}_{n-1}$ .

After  $n$  steps, the situation can be summarized as

$$\mathbf{A}\mathbf{V}_n = \mathbf{V}_n\mathbf{T}_n + \beta_{n+1}\mathbf{v}_{n+1}\mathbf{e}_n^T, \quad (5.17)$$

$$\mathbf{V}_n = [\mathbf{v}_1 \quad \mathbf{v}_2 \quad \cdots \quad \mathbf{v}_n], \quad \mathbf{T}_n = \begin{pmatrix} \alpha_1 & \beta_2 & & \\ \beta_2 & \alpha_2 & \beta_3 & \\ & \ddots & \ddots & \ddots \\ & & & \alpha_n \end{pmatrix}, \quad (5.18)$$

where  $\mathbf{e}_n$  is the  $n^{\text{th}}$  unit vector,  $\mathbf{V}_n$  has theoretically orthonormal columns, and  $\mathbf{T}_n$  is tridiagonal and symmetric. In practice  $\mathbf{V}_n^T \mathbf{V}_n \neq \mathbf{I}$  unless  $\mathbf{v}_{n+1}$  is reorthogonalized with respect to previous vectors, but relation (5.14) remains accurate to machine precision. This permits  $\mathbf{V}_n$  and  $\mathbf{T}_n$  to be used to solve  $\mathbf{Ax} = \mathbf{b}$  accurately in a manner that is algebraically equivalent to the conjugate-gradients method, as described in [107]. (It also leads to reliable methods for solving  $\mathbf{Ax} = \mathbf{b}$  when  $\mathbf{A}$  is indefinite [107].) Note that  $\mathbf{v}_1$  must be proportional to  $\mathbf{b}$  as shown.

When  $\mathbf{A}$  is positive definite, each  $\mathbf{T}_n$  is also positive definite and we may form the Cholesky factorization  $\mathbf{T}_n = \mathbf{L}_n \mathbf{L}_n^T$  (with  $\mathbf{L}_n$  lower-triangular) by updating  $\mathbf{L}_{n-1}$ . The conjugate-gradient method computes a sequence of approximate solutions to  $\mathbf{Ax} = \mathbf{b}$  in the form  $\mathbf{x}_n = \mathbf{V}_n \mathbf{y}_n$ , where  $\mathbf{y}_n$  is defined by the equation  $\mathbf{T}_n \mathbf{y}_n = \beta_1 \mathbf{e}_1$ . Since  $\mathbf{V}_n (\beta_1 \mathbf{e}_1) = \mathbf{b}$  exactly for all  $n$ , we see from (6.17) that  $\mathbf{Ax}_n = \mathbf{b} + \mathbf{r}_n$ , where the residual vector  $\mathbf{r}_n = \beta_{n+1} \mathbf{v}_{n+1} (\mathbf{e}_n^T \mathbf{y}_n)$  becomes small if either  $\beta_{n+1}$  is small (unlikely in practice) or the last element of  $\mathbf{y}_n$  is small.

In practice, we do not compute  $\mathbf{y}_n$  itself because every element differs from  $\mathbf{y}_{n-1}$ . Instead, we compute two quantities  $\mathbf{z}_n$  and  $\mathbf{W}_n$  by applying forward substitution to the lower-triangular systems  $\mathbf{L}_n \mathbf{z}_n = \beta_1 \mathbf{e}_1$  and  $\mathbf{L}_n \mathbf{W}_n^T = \mathbf{V}_n^T$ , where

$$\mathbf{z}_n = \begin{bmatrix} \mathbf{z}_{n-1} \\ \zeta_n \end{bmatrix}, \quad \mathbf{W}_n = [\mathbf{W}_{n-1} \quad \mathbf{w}_n] \equiv \mathbf{V}_n \mathbf{L}_n^{-T}, \quad (5.19)$$

so that  $\mathbf{x}_n$  can be updated according to  $\mathbf{x}_n = \mathbf{V}_n \mathbf{y}_n = \mathbf{W}_n \mathbf{z}_n = \mathbf{x}_{n-1} + \zeta_n \mathbf{w}_n$ . Since  $\mathbf{L}_n$  is bidiagonal, only the most recent columns of  $\mathbf{V}_n$  need to be retained in memory. Thus, the previous equation is the update rule for the image estimate in the algorithm.

In order to estimate elements of  $\mathbf{A}^{-1}$ , we can make use of the same vectors  $\mathbf{w}_n$  in Eq. (5.19).

If we now assume that exact arithmetic holds, we see that

$$\mathbf{W}_n^T \mathbf{A} \mathbf{W}_n = \mathbf{L}_n^{-1} \mathbf{V}_n^T \mathbf{A} \mathbf{V}_n \mathbf{L}_n^{-T} = \mathbf{L}_n^{-1} \mathbf{T}_n \mathbf{L}_n^{-T} = \mathbf{L}_n^{-1} \mathbf{L}_n \mathbf{L}_n^T \mathbf{L}_n^{-T} = \mathbf{I}.$$

If we further assume that the Lanczos process continues for  $N$  iterations, we have  $\mathbf{W}_N^T \mathbf{A} \mathbf{W}_N = \mathbf{I}$ , so that  $\mathbf{W}_N \mathbf{W}_N^T = \mathbf{A}^{-1}$ . On this basis, if we define  $\mathbf{B}_n = \mathbf{W}_n \mathbf{W}_n^T$ , we have the sequence of estimates  $\mathbf{B}_n = \mathbf{B}_{n-1} + \mathbf{w}_n \mathbf{w}_n^T \approx \mathbf{A}^{-1}$ . To estimate its  $i$ th diagonal, we form the sum  $\mathbf{e}_i^T \mathbf{B}_n \mathbf{e}_i = \sum_n \mathbf{w}_{in}^2$ . Thus, we can obtain monotonically increasing estimates for all diagonals at very little cost<sup>3</sup>, in the manner of LSQR [106].

Similarly, for the matrix  $\mathbf{C}$ , whose diagonals we wish to estimate, we have

$$\mathbf{e}_i^T \mathbf{C} \mathbf{e}_i = \mathbf{e}_i^T \mathbf{Q} \mathbf{A}^{-1} \mathbf{Q}^T \mathbf{e}_i \approx \mathbf{e}_i^T \mathbf{Q} \sum_n (\mathbf{w}_n \mathbf{w}_n^T) \mathbf{Q}^T \mathbf{e}_i = \mathbf{e}_i^T \sum_n (\mathbf{q}_n \mathbf{q}_n^T) \mathbf{e}_i = \sum_n \mathbf{q}_{in}^2,$$

where  $\mathbf{q}_n = \mathbf{Q} \mathbf{w}_n$  can be formed at each Lanczos iteration and then discarded after use. This is how we evaluate  $\mathbf{C}_k^{(t)}(i, i)$  in (5.16).

Element estimation of inverses of large matrices is also required in many other recently developed Bayesian algorithms (see for example [8], [77], and [128]) and presently to the best of our knowledge are handled either by inaccurate circulant or diagonal approximations of the matrix  $\mathbf{A}$  or by very time-consuming Monte-Carlo approaches.

An iteration of the variational EM algorithm consists of the update steps given by Eq. (5.10-16). In our implementation, the parameter  $\beta$  is estimated in a preprocessing step, as described above. During the variational M-step the bisection method is used for the update of the parameters  $\nu_k$  with termination criterion  $|\nu_k^m - \nu_k^{m-1}| < 10^{-6}$ , where  $\nu_k^m$  the value of  $\nu_k$  at the  $m^{\text{th}}$  iteration of the bisection method. The linear system in Eq. (5.14) is solved by the iterative Lanczos procedure. The termination criterion for this algorithm is

$$\left\| \mathbf{r}_n^{(t)} \right\| = \left\| \mathbf{H}^T \mathbf{g} - \mathbf{R}^{(t-1)} \mathbf{m}_n^{(t)} \right\| < \left\| \left( \mathbf{R}^{(t-1)} \right)^{-1} \right\|_{fro} \left\| \mathbf{m}_n^{(t)} \right\| 10^{-9},$$

where  $n$  denotes the iteration index of the Lanczos process (hence  $n = 1, 2, \dots, M^{(t)}$ ). Thus,  $\mathbf{m}_n^{(t)}$  is the image estimate at the  $n$ -th Lanczos iteration and at the  $t$ -th iteration of the overall variational algorithm. Lastly,  $\left\| \cdot \right\|_{fro}$  denotes the *Frobenius* norm. As criterion for termination

<sup>3</sup> See <http://www.stanford.edu/group/sol/software/cglanczos.html> for Matlab code.

of the variational algorithm we used  $\|\mathbf{r}_{M^{(t+1)}}^{(t+1)}\| \geq \|\mathbf{r}_{M^{(t)}}^{(t)}\|$ . In other words, we terminate the overall algorithm when the residual of the Lanczos process at iteration  $t+1$  is larger than that of the iteration  $t$ .

The overall algorithm is described is summarized in the following three-step procedure:

1. Initialize  $\mathbf{m}^0$ ,  $\beta$  using a stationary model [94].

2. Repeat until convergence:

$t$ -th iteration:

- VE-step: Update  $\mathbf{m}^t$ ,  $\mathbf{R}^t$  and  $q^t(\mathbf{a}_k)$  using equations (5.14), (5.16) and (5.17) respectively. For the last equation,  $\mathbf{m}_k^{(t)}$  and  $\mathbf{C}_k^{(t)}$  are needed also to be calculated. Also, calculate the expected value of  $q^t(a_k(i))$  from (5.17), need for the VM-step and the next VE-step in the  $(t+1)$ th iteration.

- VM-step: Update  $\lambda_k^t$  using (5.19) and  $\nu_k^t$  by solving (5.20) for each  $k$ .

3. Use  $\mathbf{m}^t$  as the restored image estimate.

## 5.5 Numerical Experiments

We demonstrate the value of the proposed restoration approach by showing results from various experiments with three  $256 \times 256$  input images: “Lena”, “Cameraman” and “Shepp-Logan” phantom. Every image is blurred with two types of blur; the first has the shape of a Gaussian function with shape parameter 9, and the second is uniform with support a rectangular region of  $9 \times 9$  pixels. The blurred signal to noise ratio (BSNR) defined as follows was used to quantify the noise level:

$$BSNR = 10 \log_{10} \frac{\|\mathbf{H}\mathbf{f}\|^2}{\sigma^2 N},$$

where  $\sigma^2$  is the variance of the additive white Gaussian noise (AWGN). Three levels of AWGN were added to the blurred images with  $BSNR=40$ ,  $30$  and  $20$  dB. Thus, in total 18 image restoration experiments were performed to test the proposed algorithm.

As performance metric, the improvement in Signal to Noise Ratio (*ISNR*) was used:

$$ISNR = 20 \log_{10} \frac{\|\mathbf{f} - \mathbf{g}\|}{\|\mathbf{f} - \hat{\mathbf{f}}\|},$$

where  $\mathbf{f}$ ,  $\mathbf{g}$  and  $\hat{\mathbf{f}}$  are the original, observed degraded and restored images, respectively.

We present *ISNR* results comparing our algorithm with four total-variation (*TV*) based Bayesian algorithms in [17] abbreviated as *BFO1*, in [18] abbreviated as *BFO2*, and [8] abbreviated as *BMK1* and *BMK2*. For comparison purposes we also implemented a restoration algorithm based on *TV* regularization [23]. This algorithm minimizes the function  $J(\mathbf{f})$  with respect to the image:

$$J(\mathbf{f}) = \|\mathbf{g} - \mathbf{H}\mathbf{f}\|^2 + \lambda \sum_{k=1}^N \sqrt{(\mathbf{D}_h \mathbf{f})_k^2 + (\mathbf{D}_v \mathbf{f})_k^2} + c,$$

where  $\mathbf{D}_x \mathbf{f}$  and  $\mathbf{D}_y \mathbf{f}$  are the directional differences vectors of the image along the horizontal and vertical direction respectively. A conjugate gradient algorithm is used to minimize  $J(\mathbf{f})$  with a one-step-late quadratic approximation [23]. The parameters  $\lambda$  and  $c$  were kept fixed during the iterations of this algorithm and were selected by trial-and-error (*TE*) to optimize *ISNR* assuming knowledge of the original image. Since this algorithm assumes knowledge of the original it is not a realistic one. However, it provides the performance bound of the *TV* algorithm with fixed parameters. In Table 5.1 and 5.2 we present *ISNR* results comparing our algorithm with the above-mentioned methods in 18 experiments. The *ISNR* results for *BFO1*, *BFO2*, *BMK1* and *BMK2* were obtained from [18]. In these tables for reference purposes we also provide *ISNR* results for the stationary simultaneously autoregressive prior in [94].

In Fig. 5.1 restoration results are shown for the ‘‘Cameraman’’ image *BSNR*=40 *dB* noise and uniform blur. In this experiment the restored image by the proposed algorithm, is superior in *ISNR*, and is visually distinguishable from the *TV-TE* approach, which was optimized using the original image.

At this point we note that the proposed algorithm performed very well compared with the *TV*-based methods in [17],[18] [18] and [8]. More specifically, for the high *BSNR*=40 *dB* case it gave the best results from all methods (excluding *TV-TE* since it is unrealistic) in 5 out of 6 experiments. For the midlevel *BSNR*=30 *dB* case it gave the best performance in 5 out of 6

experiments. Finally, in the low  $BSNR=20$  dB case it gave the best result in 3 out of the 6 experiments. Overall the proposed algorithm gave the best  $ISNR$  results in 13 out of 18 experiments, compared to 3 out of 18 for  $BF01$  and 2 out of 18 for  $BF02$ .

We also compared our method with  $BF01$  [17], which based on the above experiments was the most competitive  $TV$  based method. We used the same three images and noise levels as above. We also used a  $5 \times 5$  pyramidal blur with impulse response given by  $[1 \ 4 \ 6 \ 4 \ 1]^T [1 \ 4 \ 6 \ 4 \ 1]/256$ . The  $ISNR$  results for this experiment are given in Table 5.3. For the implementation we used the code provided by the authors<sup>4</sup>. The  $ISNR$  results from this experiment are consistent with the previous ones.

## 5.6 Conclusions and suggestions for improvement

We presented a new Bayesian framework for image restoration that uses a product-based Student's-t type of priors. The main theoretical contribution is that by constraining the approximation of the posterior in the variational framework, we bypass the need for knowing the normalization constant of this prior. Thus, we avoid having to use improper priors, i.e. priors whose normalization constant is empirically selected, see for example [17], [18], [8], [26], [130] and [29]. Furthermore, the proposed methodology does not require empirical parameter selection as in the MAP methodology that uses a similar-in-spirit prior ([26] and [29]). We also presented a Lanczos-based computational scheme tailored to the computations required by our algorithm.

We demonstrated by the  $ISNR$  results in Tables 5.1 and 5.2 that the proposed method is competitive with the very recently proposed  $TV$ -based Bayesian algorithms ([17], [18] and [8]). More specifically, it appears that this approach is more competitive in the higher  $BSNR$  cases. Thus, it seems that in such cases the proposed Student's-t model has the ability to capture more accurately than  $TV$ -based priors the subtle features of the image present in the observations. However, in the presence of high levels of AWGN, this does not seem to be the case and the advantage of our proposed prior compared to  $TV$  priors seems to diminish. We believe that this is the case because high levels of noise “wipe out” the subtle features that our model can capture.

---

<sup>4</sup> <http://www.lx.it.pt/~jpaos>



We found empirically that modeling explicitly the diagonal edge structure contained in the vertical and horizontal edge (the use of operators  $\mathbf{Q}_3$  and  $\mathbf{Q}_4$ ) improved the performance of the proposed algorithm, for a wide range of images, blurs and  $SNRs$ . Selecting optimally such operators according to the image is a topic of further investigation.

Another topic of further investigation is the use of *generalized* Student's-t pdfs. These pdfs are produced by assuming spatial variation of the parameters of other densities than the Gaussian (spatial variation has been considered so far for the inverse variance of the Gaussian) and incorporating a Gamma hyperprior for each parameter. In Chapter 6 we make the TV prior spatially varying and in this way we obtain a *generalized* Student's-t prior.

Finally, the numerical method to estimate the diagonal elements of a very large matrix is amenable to further improvement with respect to speed and/or accuracy of the estimation.

Table 5.1: *ISNR* results comparing the proposed algorithm with the algorithms in [17], [18] and [8] using 3 images, 3 noise levels and Gaussian shaped blur. The *ISNR* results for the *BFO1*, *BFO2*, *BMK1* and *BMK2* algorithms are obtained from [8].

Gaussian shaped blur with $\sigma^2 = 9$		Lena	Cameraman	Shepp-Logan
<i>BSNR</i>	Method	<i>ISNR</i>	<i>ISNR</i>	<i>ISNR</i>
40dB	<i>Stationary</i>	3.33	2.44	3.56
	<i>Proposed</i>	<b>4.86</b>	3.45	<b>9.46</b>
	<i>TV-TE</i>	<u>4.87</u>	<u>3.34</u>	<u>8.30</u>
	<i>BFO1</i>	4.72	<b>3.51</b>	7.07
	<i>BFO2</i>	4.50	3.27	5.88
	<i>BMK1</i>	4.78	3.39	6.69
	<i>BMK2</i>	4.49	3.26	5.63
30dB	<i>Stationary</i>	2.54	1.89	2.80
	<i>Proposed</i>	<b>3.89</b>	2.74	<b>5.94</b>
	<i>TV-TE</i>	<u>3.82</u>	<u>2.82</u>	<u>5.50</u>
	<i>BFO1</i>	3.87	<b>2.89</b>	5.15
	<i>BFO2</i>	3.56	2.47	3.94
	<i>BMK1</i>	3.87	2.63	4.31
	<i>BMK2</i>	3.55	2.41	3.72
20dB	<i>Stationary</i>	2.23	1.43	2.14
	<i>Proposed</i>	2.76	1.86	<b>3.92</b>
	<i>TV-TE</i>	<u>3.20</u>	<u>2.27</u>	<u>3.75</u>
	<i>BFO1</i>	<b>3.02</b>	2.13	3.56
	<i>BFO2</i>	2.47	<b>2.23</b>	2.20
	<i>BMK1</i>	2.87	1.72	1.85
	<i>BMK2</i>	2.42	1.42	2.05

Table 5.2: *ISNR* results comparing the proposed algorithm with the algorithms in [17], [18] and [8] using 3 images, 3 noise levels and uniform blur. The *ISNR* results for the *BFO1*, *BFO2*, *BMK1* and *BMK2* algorithms are obtained from [8].

Uniform 9×9 blur		Lena	Cameraman	Shepp-Logan
<i>BSNR</i>	Method	<i>ISNR</i>	<i>ISNR</i>	<i>ISNR</i>
40dB	<i>Stationary</i>	4.72	4.57	5.31
	<i>Proposed</i>	<b>8.49</b>	<b>9.53</b>	<b>15.08</b>
	<i>TV-TE</i>	<u>8.43</u>	<u>9.07</u>	<u>16.63</u>
	<i>BFO1</i>	8.34	8.55	14.22
	<i>BFO2</i>	8.35	8.25	12.01
	<i>BMK1</i>	8.42	8.57	13.69
	<i>BMK2</i>	8.37	8.46	12.05
30dB	<i>Stationary</i>	4.06	3.24	3.56
	<i>Proposed</i>	<b>6.10</b>	<b>6.29</b>	<b>9.71</b>
	<i>TV-TE</i>	<u>5.93</u>	<u>6.26</u>	<u>10.66</u>
	<i>BFO1</i>	6.08	5.68	8.88
	<i>BFO2</i>	5.64	4.65	6.91
	<i>BMK1</i>	5.89	5.41	7.77
	<i>BMK2</i>	5.58	4.38	6.50
20dB	<i>Stationary</i>	2.68	2.19	2.49
	<i>Proposed</i>	3.98	<b>3.33</b>	<b>6.10</b>
	<i>TV-TE</i>	<u>3.90</u>	<u>3.33</u>	<u>6.26</u>
	<i>BFO1</i>	4.09	3.31	5.57
	<i>BFO2</i>	<b>4.14</b>	2.12	2.95
	<i>BMK1</i>	3.72	2.42	3.01
	<i>BMK2</i>	3.15	1.94	2.64

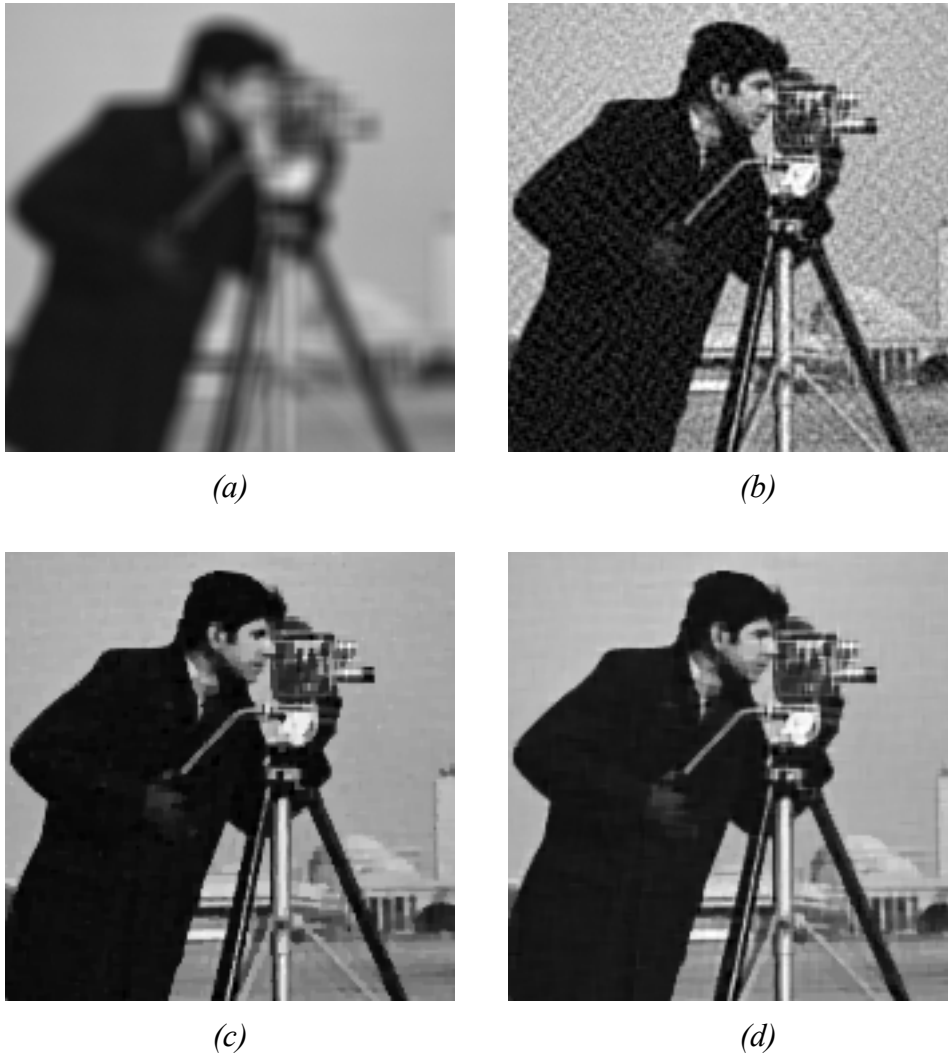


Figure 5.1: a) Degraded “Cameraman” image by uniform  $9 \times 9$  blur and noise with  $BSNR=40dB$ , b) restored image using a stationary Gaussian prior [94]  $ISNR=5.76 dB$ , c) restored image using  $TV-TE$   $ISNR = 9.07 dB$ , d) restored image using proposed algorithm  $ISNR = 9.53dB$ .

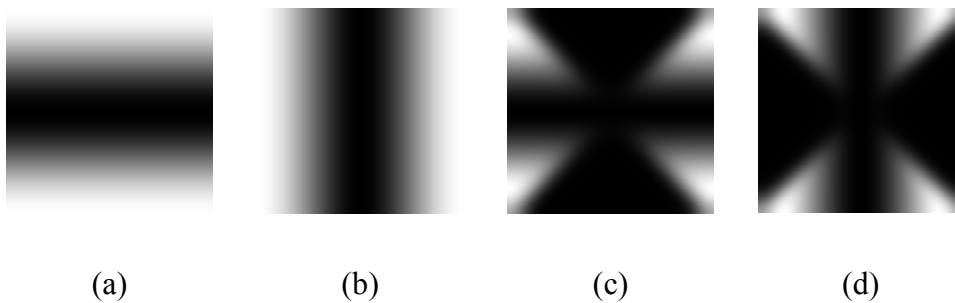


Figure 5.2: Magnitude of frequency responses of the filters used in the prior: (a) horizontal differences ( $Q_1$ ), (b) vertical differences ( $Q_2$ ), (c)  $Q_3$  and (d)  $Q_4$ .

Table 5.3: *ISNR* results comparing the proposed algorithm with the algorithms in [17] using 3 images, 3 noise levels and pyramidal blur.

Pyramidal blur		Lena	Cameraman	Shepp-Logan
<i>BSNR</i>	Method	<i>ISNR</i>	<i>ISNR</i>	<i>ISNR</i>
40dB	<i>Stationary</i>	4.82	3.82	2.68
	<b><i>Proposed</i></b>	<b>7.02</b>	<b>6.40</b>	<b>13.70</b>
	<i>BFOI</i>	5.56	6.07	10.87
30dB	<i>Stationary</i>	3.03	2.45	1.55
	<b><i>Proposed</i></b>	<b>4.81</b>	4.25	<b>8.51</b>
	<i>BFOI</i>	4.52	<b>4.35</b>	7.91
20dB	<i>Stationary</i>	1.57	1.26	1.01
	<b><i>Proposed</i></b>	<b>3.03</b>	<b>2.75</b>	<b>7.00</b>
	<i>BFOI</i>	3.01	2.60	5.91

## Appendix

In the VE-step the bound must be optimized with respect to  $\mathbf{R}$ ,  $\mathbf{m}$  and  $q(\tilde{\mathbf{a}})$ . With the mean field approximation (Eq. (5.10)) the bound becomes

$$L(q(\tilde{\mathbf{a}}), \theta_1, \theta_2) = \int \prod_{k=1}^P q(\boldsymbol{\varepsilon}_k; \theta_1) q(\mathbf{a}_k) \log p(\tilde{\mathbf{y}}, \tilde{\boldsymbol{\varepsilon}}, \tilde{\mathbf{a}}; \theta_2) d\tilde{\boldsymbol{\varepsilon}} d\tilde{\mathbf{a}} - \int \prod_{k=1}^P q(\boldsymbol{\varepsilon}_k; \theta_1) q(\mathbf{a}_k) \log \prod_{k=1}^P q(\boldsymbol{\varepsilon}_k; \theta_1) q(\mathbf{a}_k) d\tilde{\boldsymbol{\varepsilon}} d\tilde{\mathbf{a}},$$

where  $\theta_1 = [\mathbf{R}, \mathbf{m}]$  and  $\theta_2 = [\beta, \lambda_1, \dots, \lambda_p, \nu_1, \dots, \nu_p]^T$ .

Because at this point we aim to optimize with respect to  $\theta_1$ , we operate on the function  $L'$ , which includes only the terms that depend on the parameters  $\theta_1$ :

$$L \propto L'(\theta_1)$$

$$\begin{aligned} L'(\theta_1) &= \sum_{k=1}^P \int q(\boldsymbol{\varepsilon}_k; \theta_1) q(\mathbf{a}_k) \log p(\mathbf{y}_k | \boldsymbol{\varepsilon}_k; \theta_2) p(\boldsymbol{\varepsilon}_k | \mathbf{a}_k; \theta_2) d\boldsymbol{\varepsilon}_k d\mathbf{a}_k - \\ &- \sum_{k=1}^P \int q(\boldsymbol{\varepsilon}_k; \theta_1) \log q(\boldsymbol{\varepsilon}_k; \theta_1) d\boldsymbol{\varepsilon}_k. \end{aligned} \quad (\text{A.5.1})$$

The first sum is further analyzed:

$$\begin{aligned} &\sum_{k=1}^P \int q(\boldsymbol{\varepsilon}_k; \theta_1) q(\mathbf{a}_k) \log p(\mathbf{y}_k | \boldsymbol{\varepsilon}_k; \theta_2) p(\boldsymbol{\varepsilon}_k | \mathbf{a}_k; \theta_2) d\boldsymbol{\varepsilon}_k d\mathbf{a}_k \propto \\ &\propto \sum_{k=1}^P \left\langle -\beta (\mathbf{H}\boldsymbol{\varepsilon}_k - \mathbf{y}_k)^\top \mathbf{Q}_k^{-\top} \mathbf{Q}_k^{-1} (\mathbf{H}\boldsymbol{\varepsilon}_k - \mathbf{y}_k) - \lambda_k \boldsymbol{\varepsilon}_k^\top \mathbf{A}_k \boldsymbol{\varepsilon}_k \right\rangle_{q(\boldsymbol{\varepsilon}_k; \theta_1) q(\mathbf{a}_k)} = \\ &= -\beta P \|\mathbf{H}\mathbf{m} - \mathbf{g}\|_2^2 - \sum_{k=1}^P \lambda_k \mathbf{m}^\top \mathbf{Q}_k^\top \hat{\mathbf{A}}_k \mathbf{Q}_k \mathbf{m} - \text{trace} \left\{ \left( \beta \mathbf{P}\mathbf{H}^\top \mathbf{H} + \sum_{k=1}^P \lambda_k \mathbf{Q}_k^\top \hat{\mathbf{A}}_k \mathbf{Q}_k \right) \mathbf{R} \right\}, \end{aligned} \quad (\text{A.5.2})$$

where  $\hat{\mathbf{A}}_k$  is a diagonal matrix with elements

$$\hat{\mathbf{A}}_k(i, i) = \langle a_k(i) \rangle_{q(a_k(i))}, i = 1, \dots, N.$$

The second integral is the entropy of a Gaussian function, which is proportional to

$$\int q(\boldsymbol{\varepsilon}_k; \theta_1) \log q(\boldsymbol{\varepsilon}_k; \theta_1) d\boldsymbol{\varepsilon}_k \propto \frac{1}{2} \log \det |\mathbf{R}|. \quad (\text{A.5.3})$$

Setting the derivative of  $L'$  w.r.t  $\mathbf{R}$  equal to zero using Eq. (A.5.1-3) yields

$$\begin{aligned} \frac{\partial L'(\theta_1)}{\partial \mathbf{R}} = 0 &\Rightarrow \frac{\partial \text{trace} \left\{ \beta \mathbf{P}\mathbf{H}^\top \mathbf{H} \mathbf{R} + \sum_{k=1}^P \lambda_k \mathbf{Q}_k^\top \hat{\mathbf{A}}_k \mathbf{Q}_k \mathbf{R} \right\} - P \partial \log \det |\mathbf{R}|}{\partial \mathbf{R}} = 0 \\ &\Rightarrow \beta \mathbf{P}\mathbf{H}^\top \mathbf{H} + \sum_{k=1}^P \lambda_k \mathbf{Q}_k^\top \hat{\mathbf{A}}_k \mathbf{Q}_k - P \mathbf{R}^{-1} = 0 \Rightarrow \mathbf{R} = \left( \beta \mathbf{H}^\top \mathbf{H} + \frac{1}{P} \sum_{k=1}^P \lambda_k \mathbf{Q}_k^\top \hat{\mathbf{A}}_k \mathbf{Q}_k \right)^{-1}. \end{aligned}$$

Similarly, using Eq. (A.5.2), we find that the optimum for the mean:

$$\frac{\partial L'(\theta_1)}{\partial \mathbf{m}} = 0 \Rightarrow \mathbf{m} = \beta \mathbf{R}\mathbf{H}^\top \mathbf{g}.$$

The final part of the VE-step is the optimization w.r.t. the function  $q(\tilde{\mathbf{a}})$ . It is straightforward to verify that this is achieved when

$$q(\tilde{\mathbf{a}}) = \frac{\exp\left(\langle \log p(\tilde{\mathbf{y}}, \tilde{\boldsymbol{\varepsilon}}, \tilde{\mathbf{a}}) \rangle_{q(\tilde{\boldsymbol{\varepsilon}})}\right)}{\int \exp\left(\langle \log p(\tilde{\mathbf{y}}, \tilde{\boldsymbol{\varepsilon}}, \tilde{\mathbf{a}}) \rangle_{q(\tilde{\boldsymbol{\varepsilon}})}\right) d\tilde{\mathbf{a}}} = \prod_{k=1}^P \prod_{i=1}^N q(a_k(i)).$$

The product form is due to

$$\exp\langle \log p(\tilde{\mathbf{y}}, \tilde{\boldsymbol{\varepsilon}}, \tilde{\mathbf{a}}) \rangle_{q(\tilde{\boldsymbol{\varepsilon}})} \propto \prod_{k=1}^P \prod_{i=1}^N (a_k(i))^{\frac{\nu_k}{2} + \frac{1}{2} - 1} \exp\left\{-\frac{\nu_k}{2} a_k(i) - \frac{1}{2} \lambda_k \left( (\mathbf{m}_k(i))^2 + \mathbf{C}_k(i, i) \right) a_k(i)\right\}.$$

Hence, each  $q(a_k(i))$  is a Gamma distribution:

$$q(a_k(i)) = \text{Gamma}\left(a_k(i); \frac{\nu_k}{2} + \frac{1}{2}, \frac{\nu_k}{2} + \frac{1}{2} \lambda_k \left( (\mathbf{m}_k(i))^2 + \mathbf{C}_k(i, i) \right)\right),$$

where  $\mathbf{m}_k = \mathbf{Q}_k \mathbf{m}$  and  $\mathbf{C}_k = \mathbf{Q}_k \mathbf{R} \mathbf{Q}_k^T$ .

# CHAPTER 6. BAYESIAN IMAGE RESTORATION WITH A SPATIALLY ADAPTIVE TOTAL VARIATION PRIOR

- 
- 6.1. Introduction
  - 6.2. Imaging and image model
  - 6.3. Variational inference with the generalized Student's-t prior
  - 6.4. A Constrained Variational Inference Algorithm
  - 6.5. Computational Implementation
  - 6.6. Numerical Experiments
  - 6.7 Conclusions
- 

In this chapter a new image prior is introduced and is used for image restoration. This prior generalizes total-variation (TV) priors previously proposed, it is explicitly spatially adaptive, and it is based on a generalization of the Student's-t density function discussed in Chapter 5. Thus, similar to Chapter 5, Bayesian inference is used for image restoration with this new prior via the variational approximation. However, the variational approximation used herein is more complicated because of the non-quadratic terms involved in the TV prior. The proposed algorithm is fully automatic in the sense that all necessary parameters are estimated from the data. Numerical experiments are shown which demonstrate that image restoration based on this prior compares favorably with the algorithm of Chapter 5 and previous state-of-the-art restoration algorithms.

## 6.1 Introduction



As discussed in section 3.4, the TV prior has been used extensively and with great success for inverse problems due to the ability to smooth noise in flat areas of the image and at the same time preserve edges. However, TV-based image restoration has certain shortcomings. One of them is the selection of the regularization parameter which to a large extent till recently was made in an ad-hoc way. Recently, a Bayesian inference framework based on the variational approximation was proposed to handle this problem [8]. However, in this work the partition function of the image prior is only approximated and the quality of the approximation is not analyzed.

In this chapter we extend the work in [8] by introducing a new prior which has a number of novel features. The extension of the TV prior is performed in a way analogous with that used to derive the non-stationary prior of Chapters 4 and 5, where the Gaussian density is assumed. Precisely, we introduce a TV prior with spatially varying regularization parameters. In order to avoid the over parameterization due to the spatially varying nature of this prior, we introduce a Gamma hyperprior for the spatially adaptive regularization parameters of the local TV priors. Also, we use this prior in a product form, which means that we assume that the outputs of an arbitrary number of high-pass filters are distributed according to this prior. This gives two novel features to the new prior. First, it is explicitly spatially adaptive and thus it is better suited to capture the salient features of the image. Second, it is in product form and has the ability to enforce simultaneously a number of different properties to the image. With his prior arbitrary linear operators can be used, not just first order differences as TV. Thus, a prior similar to the one in [8] with an exactly calculated partition function is just a special case of it. In order to avoid the over parameterization due to the spatial adaptivity, a model with two layers of hidden variables is introduced, similar in spirit to the one used in Chapters 4 and 5. If the hidden variables of the second layer are marginalized, the resulting density function *has similar form* to a Student's-t pdf, thus we label it as *Generalized Student's-t*.

Due to the complexity of the prior, model we resort to the variational approximation for Bayesian inference [19]. However, we use two modifications. First, we determine a quadratic bound to the variational bound, in a manner similar to the methodology used in [8], to bypass the difficulties due to the non-quadratic nature of the new prior. Second, we use the *constrained variational* framework in a manner similar to that in Chapter 5, in order to bypass the problem of computing the partition function of the new prior.

In this chapter we also propose an iterative method to compute the diagonal elements of very large inverse matrices that are necessary for the herein proposed Bayesian algorithm. This method is similar in spirit to the one employed in Chapter 5. However, it is based on conjugate-gradients and not on the Lanczos and converges faster.

The rest of this chapter is organized as follows. In section 6.2 we present the imaging and image models. In section 6.3 we present the variational algorithm for Bayesian inference and a brief analysis of the mechanisms that are introduced for spatial adaptivity by using TV and Student's-t based priors. In section 6.4 we present the details of the computational implementation of our algorithm. In section 6.5 numerical experiments are provided and in section 6.6 conclusions and future work.

## 6.2 Imaging and image model

In what follows we use one dimensional notation for simplicity. A linear imaging model is assumed. Let  $\mathbf{f}$  be the original image represented as a  $N \times 1$  vector, blurred by a convolutional operator  $\mathbf{H}$ , of size  $N \times N$ . The degraded observation is given by

$$\mathbf{g} = \mathbf{H}\mathbf{f} + \mathbf{n} \quad (6.1)$$

where  $\mathbf{n}$  is the noise  $N \times 1$  vector modeled as white Gaussian, i.e.  $\mathbf{n} \sim N(\mathbf{0}, \beta^{-1}\mathbf{I})$ , where  $\mathbf{0}$  and  $\mathbf{I}$  the  $N \times N$  zero and identity matrices, respectively.

### 6.2.1 Generalized Student's-t image prior

Image priors in product form are very attractive since they have the ability to enforce simultaneously many properties to an image. For this purpose we propose herein a prior in product form for the image. To define such a prior we introduce  $P$  pairs of linear convolutional operators (filters)  $(\mathbf{Q}_1, \mathbf{Q}_2), (\mathbf{Q}_3, \mathbf{Q}_4), \dots, (\mathbf{Q}_{2P-1}, \mathbf{Q}_{2P})$  of size  $N \times N$  and assume that the filter outputs  $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_{2P})$  are produced according to

$$\boldsymbol{\varepsilon}_l = \mathbf{Q}_l \mathbf{f}, l = 1, \dots, 2P. \quad (6.2)$$

Then, for each pixel location  $i$ , it is assumed that each pair  $\varepsilon_{2k}(i)$  and  $\varepsilon_{2k-1}(i)$  is jointly distributed with probability density function

$$p(\varepsilon_{2k}(i), \varepsilon_{2k-1}(i) | a_k(i)) = \frac{\lambda_k^2 a_k(i)^2}{2\pi} \exp\left(-\lambda_k a_k(i) \sqrt{\varepsilon_{2k}(i)^2 + \varepsilon_{2k-1}(i)^2}\right). \quad (6.3)$$

with  $k = 1, \dots, P$ .

Notice that for  $P=1$  and  $\mathbf{Q}_1$  and  $\mathbf{Q}_2$  the first order difference operators in the vertical and horizontal direction and  $a_k(i) = a$ , the prior becomes identical in form to the total-variation (TV) based prior proposed in [17] and [8]. However, the prior proposed herein is more general because it can use any linear operator not just first order differences. Furthermore,  $a_k(i)$  varies for every pixel  $i$  which makes it explicitly *spatially adaptive*.

In order to avoid the over-parameterization problem of this prior in a similar manner to the priors used in Chapter 4 and 5, we assume  $a_k(i)$  to be Gamma distributed hidden random variable which can be marginalized according to:

$$p(a_k(i)) = \text{Gamma}(a_k(i); \nu_k / 2, \nu_k / 2).$$

The marginal distribution of  $\varepsilon_{2k}(i)$  and  $\varepsilon_{2k-1}(i)$  can be computed in *closed form* and is given by

$$\begin{aligned} p(\varepsilon_{2k}(i), \varepsilon_{2k-1}(i)) &= \int_{a_k(i)} p(\varepsilon_{2k}(i), \varepsilon_{2k-1}(i) | a_k(i)) p(a_k(i)) da_k(i) \quad (6.4) \\ &= \frac{\Gamma(\nu_k / 2 + 1/2)}{\Gamma(\nu_k / 2)} \left(\frac{\lambda_k}{\pi \nu_k}\right)^{1/2} \left(1 + \frac{\lambda_k \sqrt{\varepsilon_{2k}(i)^2 + \varepsilon_{2k-1}(i)^2}}{\nu_k}\right)^{-\nu_k/2-1/2}, \end{aligned}$$

for  $k = 1, 2, \dots, P$ .

This density function is very similar in form to the Student's-t pdf. Thus, in the rest of this paper we label it as *generalized Student's-t*. This prior combines the advantages of both TV-based and Student's-t based priors. The former being the ability to suppress noise and maintain edges in an image and the latter explicit spatial adaptivity through the hidden random variables  $a_k(i)$ .

At this point we note that we have not provided a prior for the image  $p(\mathbf{f})$ . This was intentional, because we cannot compute it in closed form. More specifically, it is difficult to define a prior for the image  $\mathbf{f}$  based on the prior in Eq. (6.3) because we cannot compute the

partition function for such prior. First, the non-quadratic exponent in the pdf in Eq. (6.3) makes this calculation intractable even if our prior was not in product form. Furthermore, since we want to use a prior in product form even with a quadratic exponent it is not possible to compute the partition function. For this purpose, in the next section we bypass this difficulty by working in the domain of the filter outputs  $\boldsymbol{\varepsilon}$ , in a similar fashion as in Chapter 5 where the prior in (6.3) can be used directly and there is no need to define a prior for  $\mathbf{f}$ . The downside of this choice is that it is not obvious how to merge the estimates of all the  $\boldsymbol{\varepsilon}_l$ ,  $l=1, \dots, 2P$  to generate one estimate for  $\mathbf{f}$ . For this purpose we propose the *constrained variational* approach which we also present in the following section.

### 6.3 Variational inference with the generalized Student's-t prior

We thus introduce an alternative imaging model, which is derived by applying the operators  $\mathbf{Q}_l$  to the original imaging model in (6.1). This yields:

$$\mathbf{y}_l = \mathbf{H}\boldsymbol{\varepsilon}_l + \mathbf{n}_l, l = 1, \dots, 2P, \quad (6.5)$$

where  $\mathbf{y}_l = \mathbf{Q}_l \mathbf{g}$ ,  $\mathbf{n}_l = \mathbf{Q}_l \mathbf{n}$  and thus  $\mathbf{n}_l \sim N(0, (\beta \mathbf{Q}_l \mathbf{Q}_l^T)^{-1})$ .

This imaging model will be used in what follows and we treat  $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_{2P})$ , where  $\boldsymbol{\varepsilon}_l = (\varepsilon_l(1), \dots, \varepsilon_l(N))$   $l = 1, \dots, 2P$  and  $\mathbf{a} = (\mathbf{a}_1, \dots, \mathbf{a}_P)$ , where  $\mathbf{a}_k = (a_k(1), \dots, a_k(N))$   $k = 1, \dots, P$ , as *hidden variables*.

Then, according to Bayesian inference we find the posterior distributions for the hidden variables and estimate the parameters  $\theta = (\lambda_1, \dots, \lambda_p, \nu_1, \dots, \nu_p)$ . The marginal of the observations  $p(\mathbf{y}; \theta)$ , where  $\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_{2P})$ , which is required to find the posteriors of the hidden variables is hard to compute [19]. More specifically, the integral

$$p(\mathbf{y}) = \int_{\boldsymbol{\varepsilon}, \mathbf{a}} p(\mathbf{y}, \boldsymbol{\varepsilon}, \mathbf{a}) d\boldsymbol{\varepsilon} d\mathbf{a} \quad (6.6)$$

where

$$p(\mathbf{y}, \boldsymbol{\varepsilon}, \mathbf{a}) = p(\mathbf{y} | \boldsymbol{\varepsilon}) p(\boldsymbol{\varepsilon} | \mathbf{a}) p(\mathbf{a}) \quad (6.7)$$

with

$$p(\mathbf{y} | \boldsymbol{\varepsilon}) = \prod_{l=1}^{2P} p(\mathbf{y}_l | \boldsymbol{\varepsilon}_l), \quad p(\mathbf{y}_l | \boldsymbol{\varepsilon}_l) = N(\boldsymbol{\varepsilon}_l, (\beta \mathbf{Q}_l \mathbf{Q}_l^T)^{-1}) \quad (6.8)$$

$$p(\boldsymbol{\varepsilon} | \mathbf{a}) = \prod_{k=1}^P \prod_{i=1}^N p(\varepsilon_{2k}(i), \varepsilon_{2k-1}(i) | a_k(i)) \quad (6.9)$$

and

$$p(\mathbf{a}) = \prod_{k=1}^P \prod_{i=1}^N p(a_k(i)) \quad (6.10)$$

is intractable.

The variational algorithm that we describe in what follows, bypasses this difficulty and maximizes a *lower bound* that can be found instead of the log-likelihood of the observations  $\log p(\mathbf{y}; \theta)$  [19]. This bound is obtained by subtracting from  $\log p(\mathbf{y}; \theta)$  the Kullback-Leibler divergence, which is always positive, between an arbitrary  $q(\boldsymbol{\varepsilon}, \mathbf{a})$  and  $p(\boldsymbol{\varepsilon}, \mathbf{a} | \mathbf{y}; \theta)$ :

$$L(q(\boldsymbol{\varepsilon}, \mathbf{a}), \theta) = \log p(\mathbf{y}; \theta) - KL(q(\boldsymbol{\varepsilon}, \mathbf{a}) \| p(\boldsymbol{\varepsilon}, \mathbf{a} | \mathbf{y}; \theta)), \quad (6.11)$$

and is equal to

$$L(q(\boldsymbol{\varepsilon}, \mathbf{a}); \theta) = \int_{\boldsymbol{\varepsilon}, \mathbf{a}} q(\boldsymbol{\varepsilon}, \mathbf{a}) \log p(\boldsymbol{\varepsilon}, \mathbf{a}, \mathbf{y}; \theta) d\boldsymbol{\varepsilon} d\mathbf{a} - \int_{\boldsymbol{\varepsilon}, \mathbf{a}} q(\boldsymbol{\varepsilon}, \mathbf{a}) \log q(\boldsymbol{\varepsilon}, \mathbf{a}) d\boldsymbol{\varepsilon} d\mathbf{a}, \quad (6.12)$$

where we use the notation  $p(\cdot; \theta)$  to denote that  $\theta$  is a set of hyperparameters which are not treated as random variables. We could also have used  $p_\theta(\cdot)$ .

When  $q(\boldsymbol{\varepsilon}, \mathbf{a}) = p(\boldsymbol{\varepsilon}, \mathbf{a} | \mathbf{y}; \theta)$ , this bound is maximized and  $L(q(\boldsymbol{\varepsilon}, \mathbf{a}); \theta) = \log p(\mathbf{y}; \theta)$ . Because the exact posterior  $p(\boldsymbol{\varepsilon}, \mathbf{a} | \mathbf{y}; \theta) = p(\boldsymbol{\varepsilon}, \mathbf{a}, \mathbf{y}; \theta) p(\mathbf{y}; \theta)$  cannot be found we use an approximation of the posterior. The mean-field approximation is a commonly used approach to maximize the variational bound w.r.t.  $q(\boldsymbol{\varepsilon}, \mathbf{a}), \theta$  [19]. According to this approach the hidden variables are assumed to be independent i.e.  $q(\boldsymbol{\varepsilon}, \mathbf{a}) = q(\boldsymbol{\varepsilon}) q(\mathbf{a})$ . However, for the herein model this is still not sufficient to obtain in closed form  $q(\boldsymbol{\varepsilon})$  which is necessary for inference using this approach. More specifically, the square root in the joint  $p(\boldsymbol{\varepsilon}, \mathbf{a}, \mathbf{y}; \theta)$  which originates from the prior  $p(\boldsymbol{\varepsilon} | \mathbf{a})$  makes the definition of  $q(\boldsymbol{\varepsilon})$  intractable.

### 6.3.1 A Lower Bound for $L(q(\boldsymbol{\varepsilon}, \mathbf{a}), \theta)$

For this purpose we use the methodology proposed in [10] and introduce a *lower bound* also on  $L$  also. More specifically, we use the inequality

$$\sqrt{w} \leq \frac{w+u}{2\sqrt{u}}, \quad (6.13)$$

which holds for  $w \geq 0$  and  $u > 0$ . Notice that equality holds when  $w = u$ . This inequality is used for every pixel  $i$  by setting  $w_k(i) = \varepsilon_{2k}(i)^2 + \varepsilon_{2k-1}(i)^2$ , for  $k = 1, 2, \dots, P$ , where  $u_k(i)$  are auxiliary variables used for this approximation. Using this and the prior in Eq. (7.3) we have

$$p(\varepsilon_{2k}(i), \varepsilon_{2k-1}(i) | a_k(i)) \geq M(\varepsilon_{2k}(i), \varepsilon_{2k-1}(i), u_k(i), a_k(i)) = \quad (6.14)$$

$$\frac{\lambda^2 a_k(i)^2}{2\pi} \exp\left(-\frac{a_k(i)}{2} \frac{\varepsilon_{2k}(i)^2 + \varepsilon_{2k-1}(i)^2 + u_k(i)}{\sqrt{u_k(i)}}\right),$$

for  $k = 1, \dots, P$ .

We also define  $\mathbf{u}_k = (u_k(1), \dots, u_k(N))$  and  $\mathbf{u} = (\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_P)$ . Let us now define

$$L^b(q(\boldsymbol{\varepsilon}), q(\mathbf{a}), \mathbf{u}, \theta) = \int_{\boldsymbol{\varepsilon}, \mathbf{a}} q(\boldsymbol{\varepsilon})q(\mathbf{a}) \log \frac{F(\mathbf{y}, \boldsymbol{\varepsilon}, \mathbf{a}; \mathbf{u}, \theta)}{q(\boldsymbol{\varepsilon})q(\mathbf{a})} d\boldsymbol{\varepsilon}d\mathbf{a}, \quad (6.15)$$

where

$$F(\mathbf{y}, \boldsymbol{\varepsilon}, \mathbf{a}; \mathbf{u}, \theta) = p(\mathbf{y} | \boldsymbol{\varepsilon}) \left[ \prod_{k=1}^P \prod_{i=1}^N M(\varepsilon_{2k}(i), \varepsilon_{2k-1}(i), u_k(i), a_k(i)) \right] p(\mathbf{a}; \theta). \quad (6.16)$$

then, since  $F(\mathbf{y}, \boldsymbol{\varepsilon}, \mathbf{a}; \mathbf{u}, \theta) \leq p(\mathbf{y}, \boldsymbol{\varepsilon}, \mathbf{a})$  we have

$$L^b(q(\boldsymbol{\varepsilon}), q(\mathbf{a}), \mathbf{u}, \theta) \leq L(q(\boldsymbol{\varepsilon}, \mathbf{a}), \theta). \quad (6.17)$$

and consequently the bound becomes tight when

$$\max_{\mathbf{u}} L^b(q(\boldsymbol{\varepsilon}), q(\mathbf{a}), \mathbf{u}, \theta) \leq L(q(\boldsymbol{\varepsilon}, \mathbf{a}), \theta). \quad (17.8)$$

Notice that the new lower bound  $L^b$  is quadratic in the hidden variables  $\boldsymbol{\varepsilon}$ , thus it is possible to find  $q(\boldsymbol{\varepsilon})$  that maximize it. In contrast the original bound  $L$  was not quadratic in  $\boldsymbol{\varepsilon}$ .

## 6.4 A Constrained Variational Inference Algorithm

As we have already explained,  $\boldsymbol{\varepsilon}_l$ , for  $l=1, \dots, 2P$ , are used instead of  $\mathbf{f}$  to avoid the computation of the normalization constant of the prior on  $\mathbf{f}$ . Thus, a question that needs to be addressed is how one finds  $\mathbf{f}$  given the different  $q(\boldsymbol{\varepsilon}_l)$ .

Unconstrained maximization of the bound  $L^b(q(\boldsymbol{\varepsilon}), q(\mathbf{a}), \mathbf{u}, \theta)$  using the mean field approximation results in  $q(\boldsymbol{\varepsilon}_l) = N(\boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l)$ , where  $\boldsymbol{\mu}_l = \beta \boldsymbol{\Sigma}_l \mathbf{H}^T \mathbf{g}$  and  $\boldsymbol{\Sigma}_l = (\beta \mathbf{H}^T \mathbf{H} + \mathbf{Q}_l^T \langle \mathbf{A} \rangle_l \mathbf{U}_l^{-1/2} \mathbf{Q}_l)^{-1}$ ,  $l=1, \dots, 2P$  and  $\mathbf{U}_l = \text{diag}\{u_l(1), \dots, u_l(N)\}$  and  $\langle \mathbf{A} \rangle_l = \text{diag}\{\langle a_l(1) \rangle_a, \dots, \langle a_l(N) \rangle_a\}$ , where  $\langle \cdot \rangle_a$  denotes the expectation w.r.t the distribution of  $\mathbf{a}$ .

Clearly each  $\boldsymbol{\mu}_l$  suggests a different estimate for  $\mathbf{f}$  given by  $\hat{\mathbf{f}} = \mathbf{Q}_l^{-1} \boldsymbol{\mu}_l$ . Thus, one needs to find a methodology to *merge* the information from all  $q(\boldsymbol{\varepsilon}_l)$  into one estimate of  $\mathbf{f}$ .

For this purpose the *constrained variational* approximation first proposed in [12] is applied. According to this approach, each  $q(\boldsymbol{\varepsilon}_k)$  is constrained to have the form:

$$q(\boldsymbol{\varepsilon}_l) = N(\mathbf{Q}_l \mathbf{m}, \mathbf{Q}_l \mathbf{R} \mathbf{Q}_l^T), \quad (6.19)$$

where  $\mathbf{m}$  is a  $N \times 1$  vector, taken as mean of the image, and  $\mathbf{R}$  the  $N \times N$  image covariance matrix. This form is consistent with the equation  $\boldsymbol{\varepsilon}_l = \mathbf{Q}_l \mathbf{f}$  for which

$$\bar{\boldsymbol{\varepsilon}}_l = E[\boldsymbol{\varepsilon}_l] = \mathbf{Q}_l E[\mathbf{f}] = \mathbf{Q}_l \mathbf{m}$$

and

$$E[(\boldsymbol{\varepsilon}_l - \bar{\boldsymbol{\varepsilon}}_l)(\boldsymbol{\varepsilon}_l - \bar{\boldsymbol{\varepsilon}}_l)^T] = \mathbf{Q}_l E[(\mathbf{f} - \mathbf{m})(\mathbf{f} - \mathbf{m})^T] \mathbf{Q}_l^T = \mathbf{Q}_l \mathbf{R} \mathbf{Q}_l^T$$

with  $\mathbf{R} = E[(\mathbf{f} - \mathbf{m})(\mathbf{f} - \mathbf{m})^T]$ . Using this approximation the parameters  $\mathbf{m}$  and  $\mathbf{R}$  are learned instead of  $q(\boldsymbol{\varepsilon}_l)$  according to the herein constraint variational methodology.

We now present the maximization method by giving the updates for the variables of the bound  $L^b$  in the  $j$ -th iteration. In the VE-step, maximization is performed with respect to  $q(\mathbf{a})$ ,  $\mathbf{m}$  and  $\mathbf{R}$ .

$$\text{VE-step:} \quad [\mathbf{m}^j, \mathbf{R}^j, q^j(\mathbf{a})] = \arg \max_{\mathbf{m}, \mathbf{R}, q(\mathbf{a})} L^b(q(\boldsymbol{\varepsilon}), q(\mathbf{a}), \mathbf{u}^{j-1}, \theta^{j-1}) \quad (6.20)$$

$$\text{VM-step:} \quad [\mathbf{u}^j, \theta^j] = \underset{\mathbf{u}, \theta}{\operatorname{argmax}} L^b(q^j(\boldsymbol{\varepsilon}), q^j(\mathbf{a}), \mathbf{u}, \theta) \quad (6.21)$$

The updates for the VE-Step are derived in the Appendix. These are

$$q^j(\boldsymbol{\varepsilon}_i) = N(\mathbf{Q}_i \mathbf{m}^j, \mathbf{Q}_i \mathbf{R}^j \mathbf{Q}_i^T), \quad (6.22)$$

where

$$\mathbf{m}^j = \beta \mathbf{R}^j \mathbf{H}^T \mathbf{g}, \quad (6.23)$$

$$\mathbf{R}^j = (\beta \mathbf{H}^T \mathbf{H} + \frac{1}{2P} \sum_{k=1}^P \lambda_k^{j-1} (\mathbf{Q}_{2k}^T \langle \mathbf{A}_k \rangle^{j-1} (\mathbf{U}_k^{-1/2})^{j-1} \mathbf{Q}_{2k} + \mathbf{Q}_{2k-1}^T \langle \mathbf{A}_k \rangle^{j-1} (\mathbf{U}_k^{-1/2})^{j-1} \mathbf{Q}_{2k-1}))^{-1}.$$

From the above equations it is clear that  $\mathbf{m}$  merges information from all filters  $\mathbf{Q}_l$  to produce the estimate of  $\mathbf{m}$  which is used as the estimate of  $\mathbf{f}$ .

Finally, the approximate posterior of  $\mathbf{a}$  in the VE-step is given by

$$q^j(a_k(i)) = \text{Gamma}\left(a_k(i); \frac{V_k^{j-1}}{2} + 2, \frac{V_k^{j-1}}{2} + \lambda_k^{j-1} \sqrt{u_k^{j-1}(i)}\right) \quad (6.24)$$

for  $i = 1, \dots, N$  and  $k = 1, 2, \dots, P$ . Thus, the expectation of  $a_k(i)$  w.r.t  $q^j(a_k(i))$  is

$$\langle a_k(i) \rangle_{q^j(\mathbf{a})} = \frac{V_k^{j-1} + 4}{V_k^{j-1} + 2\lambda_k^{j-1} \sqrt{u_k^{j-1}(i)}} \quad (6.25)$$

In the VM-step, the bound is maximized w.r.t to the parameters. To find  $\mathbf{u}^j$  we have to solve

$$u_k^j(i) = \underset{u_k(i)}{\operatorname{argmin}} \frac{\langle \varepsilon_{2k}(i)^2 + \varepsilon_{2k-1}(i)^2 \rangle_{q^j(\boldsymbol{\varepsilon})} + u_k(i)}{\sqrt{u_k(i)}} \quad (6.26)$$

where  $\langle \cdot \rangle_{q^j(\boldsymbol{\varepsilon})}$  the expectation w.r.t.  $q^j(\boldsymbol{\varepsilon})$ , which produces

$$u_k^j(i) = \langle \varepsilon_{2k}(i)^2 + \varepsilon_{2k-1}(i)^2 \rangle_{q^j(\boldsymbol{\varepsilon})} = \sum_{r=0}^1 ((\mathbf{m}_{2k-r}^j(i))^2 + \mathbf{C}_{2k-r}^j(i, i)) \quad (6.27)$$

for  $i = 1, \dots, N$  and  $k = 1, 2, \dots, P$ , where



$$\mathbf{m}_{2k-r}^j = \mathbf{Q}_{2k-r} \mathbf{m}^j, \mathbf{C}_{2k-r}^j = \mathbf{Q}_{2k-r} \mathbf{R}^j \mathbf{Q}_{2k-r}^T.$$

For  $\lambda_k$  we have that

$$L^b(q^j(\boldsymbol{\varepsilon}), q^j(\mathbf{a}), \mathbf{u}, \theta) = 2N \sum_{k=1}^P \log \lambda_k - \sum_{k=1}^P \sum_{i=1}^N \lambda_k \langle a_k(i) \rangle_{q^j(\mathbf{a})} \sqrt{u_k^j(i)} + \text{constant}$$

when this function is considered as a function of  $\lambda_k$  only. Thus, the update formula is

$$\lambda_k^j = \frac{2N}{\sum_{i=1}^N \langle a_k(i) \rangle_{q^j(\mathbf{a})} \sqrt{u_k^j(i)}}. \quad (6.28)$$

Similarly, for  $v_k$ ,  $k=1, 2, \dots, P$ , we have that

$$\begin{aligned} L^b(q^j(\boldsymbol{\varepsilon}), q^j(\mathbf{a}), \mathbf{u}, \theta) &= \frac{v_k}{2} \sum_{i=1}^N \langle \log a_k(i) \rangle_{q^j(\mathbf{a})} - \frac{v_k}{2} \sum_{i=1}^N \langle a_k(i) \rangle_{q^j(\mathbf{a})} - N\Gamma\left(\frac{v_k}{2}\right) \\ &+ N \frac{v_k}{2} \log\left(\frac{v_k}{2}\right) + \text{constant} \end{aligned} \quad (6.29)$$

when this function is considered as a function of  $v_k$  only. Then  $v_k^j$  is the root of the function  $\phi$  which is proportional to derivative of  $L^b(q^j(\boldsymbol{\varepsilon}), q^j(\mathbf{a}), \mathbf{u}, \theta)$  with respect to  $v_k$

$$\begin{aligned} \phi(v_k) &= \frac{1}{N} \sum_{i=1}^N \log \langle a_k(i) \rangle_{q^j(\mathbf{a})} - \frac{1}{N} \sum_{i=1}^N \langle a_k(i) \rangle_{q^j(\mathbf{a})} + \psi\left(\frac{v_k^{j-1}}{2} + 2\right) \\ &- \log\left(\frac{v_k^{j-1}}{2} + 2\right) - \psi\left(\frac{v_k}{2}\right) + \log\left(\frac{v_k}{2}\right) + 1, \end{aligned} \quad (6.30)$$

where  $\psi$  is the digamma function. We find  $\phi(v_k^j) = 0$  numerically using the bisection method.

At this point it is worth noting that when a TV prior is used in a Bayesian framework [10] the prior introduces a mechanism for spatially adaptive regularization in the restoration filter.

This is manifested by the diagonal matrices  $[\mathbf{W}(\mathbf{u}^k)]_{ii} = \frac{1}{\sqrt{u_i^k}}$  Eq. (6.34) used in the

restoration filters Eq.s (33) and (48) in [8]. These  $u_i^k$  appear also herein as the  $u_k(i)$  in Eq. (6.27). When a Student's-t prior is used in a similar framework [12] spatial adaptivity is

introduced to the restoration filter via the diagonal matrices  $\hat{\mathbf{A}}_k$  in Eq. (6.11) with elements  $a_k(i)$  given in Eq. (5.14) of Chapter 5. Similar  $a_k(i)$  appear herein also in Eq. (6.25).

Since the herein used prior integrates both the TV and Student's-t priors in [8] and in Chapter 5, respectively, it contains simultaneously both previous mechanisms for spatial adaptive regularization as manifested by the product  $\langle \mathbf{A}_k \rangle^{j-1} (\mathbf{U}_k^{-1/2})^{j-1}$  in Eq. (6.23).

## 6.5 Computational Implementation

In our implementation  $P = 4$  was used. In other words, four filter outputs were used for the prior and it is a product with two terms. We show the magnitude of the frequency responses of these filters in Fig. 6.3. The operators  $\mathbf{Q}_1$  and  $\mathbf{Q}_2$  correspond to the horizontal and vertical first order differences. Thus, these filters are used to model the vertical and horizontal image edge structure, respectively. The other two operators  $\mathbf{Q}_3$  and  $\mathbf{Q}_4$  are used to model the diagonal edge component contained in the vertical and horizontal edges, respectively. These filters are obtained by convolving the previous horizontal and vertical first order differences filters with fan filters with vertical and horizontal pass-bands, respectively. In our experiments the fan filters in [37] were used. In our prior  $\varepsilon_l$  with  $l = 1, 2$  were used together in one term of the prior and  $l = 3, 4$  in the other.

The fan filters combined with the difference filters were found empirically to provide better results than the use of the horizontal and vertical difference filters only. To explain the choice of the fan filters we note that ideally we expect from a filter when applied to an image to produce outputs as close to zero as possible. The first order differences filters have to some extent this property, but at the edges of the image this property is canceled. Thus, more filters are needed that produce outputs closer to zero. The motive to incorporate the fan filters to our algorithm is the use of them in the contourlet transform [37], which is shown to have more close to zero coefficients than the classical wavelet transform. Their ability to provide closer to zero outputs is interpreted as the ability to capture the correlations of the image edges. Hence, this renders the model more accurate. We must also note a key difference in our model with respect to [37]; in the contourlet transform the Laplacian pyramid is used as a first filter and the fan filters are applied on its output. Here, we have first order differences in the horizontal and vertical direction, something that has been found empirically to provide better

results in image resoration than the Laplacian operator. For this reason, the filters  $\mathbf{Q}_3$  and  $\mathbf{Q}_4$  are the result of the vertical and horizontal fan filter applied to  $\mathbf{Q}_1$  (horizontal) and  $\mathbf{Q}_2$  (vertical), respectively. The magnitude of the filters frequency response is shown in Fig. 5.3.

One iteration of the proposed algorithm consists of the equations (6.23)-( 6.30). The image is taken to be equal to  $\mathbf{m}$  which is obtained by solving the linear system in Eq. (6.23). The dimensions of the matrices involved in Eq. (6.23) are  $N \times N$  where  $N$  the number of pixels in the image. We solve this system iteratively by using the conjugate-gradient algorithm [116]. We also utilized this method to evaluate the diagonal elements of matrix  $\mathbf{C}_k$  in Eq. (6.24). More specifically, we utilized the  $\mathbf{R}^{-1}$ -conjugate vectors  $\mathbf{p}_i, i=1, \dots, K, K < N$ , for which  $(\mathbf{p}_i^T \mathbf{R}^{-1} \mathbf{p}_i = \delta_{ij})$ . Then according to the conjugate-gradient algorithm the image estimate is updated at every iteration as:

$$\mathbf{m}_i = \mathbf{m}_{i-1} + a\mathbf{p}_i$$

where  $a$  is a scalar [116]. If the method is allowed to iterate  $N$  times we have  $\mathbf{P}^T \mathbf{R}^{-1} \mathbf{P} = \mathbf{I}$ , where  $\mathbf{P} = [\mathbf{p}_1 \dots \mathbf{p}_N]$  with  $\mathbf{p}_i, i=1, \dots, N$ , all the  $\mathbf{R}^{-1}$ -conjugate vectors. Then,  $\mathbf{R} = \mathbf{P}^T \mathbf{P}$  and the diagonal elements of  $\mathbf{C}_k = \mathbf{Q}_k \mathbf{R} \mathbf{Q}_k^T$  can be computed by the formula

$$(\mathbf{Q}_k \mathbf{R} \mathbf{Q}_k^T)(i, i) = \sum_{j=1}^N \mathbf{p}'_j(i)^2 \approx \sum_{j=1}^K \mathbf{p}'_j(i)^2 \quad (6.31)$$

where  $\mathbf{p}'_j = \mathbf{Q}_k \mathbf{p}_j$ . In practice the number of iterations  $K$  required for convergence of the conjugate-gradient method is much smaller than  $N$  ( $K \ll N$ ). We found for  $256 \times 256$  images where  $N = 65,536$  that for  $K \approx 200$  the conjugate-gradient algorithm converges. The obtained results using this approach are significantly better than using a circulant approximation for  $\mathbf{R}$  or omitting the elements  $\mathbf{C}_k(i, i)$  from Eq. (6.27). At this point it is worth noting that the similar in spirit Lanczos-based algorithm which was proposed in Chapter 5 required for similar size images  $K \approx 1000 - 2000$  to converge. Thus, the herein proposed algorithm is faster than the algorithm in Chapter 5.

As termination criterion we chose

$$|(\mathbf{R}^j)^{-1} \mathbf{m}^j - \mathbf{H}^T \mathbf{g}| > |(\mathbf{R}^{j-1})^{-1} \mathbf{m}^{j-1} - \mathbf{H}^T \mathbf{g}| \quad (6.32)$$

where  $\mathbf{m}$  is the image estimate at the  $j$ -th iteration and it is the solution of the linear system  $(\mathbf{R}^j)^{-1}\mathbf{m} = \mathbf{H}^T \mathbf{g}$  that the conjugate-gradients algorithm solves.

The algorithm is initialized by the resulting image estimate of a Bayesian algorithm that uses a spatially invariant simultaneously autoregressive image prior [94]. In other words, we set the initial image estimate  $\mathbf{m}^0$  equal to the restored image by this algorithm. The noise precision  $\beta$  is also estimated by the algorithm in [94] and we fix it to this value for the rest of our algorithm. Thus, the overall algorithm can be summarized in the following steps:

- Initialize  $\mathbf{m}^0$  and  $\beta$  with the algorithm using a stationary prior
- Until convergence do
  1. Update the parameters  $\mathbf{u}$ ,  $\lambda$  and  $\nu$  from equations (6.27), (6.28) and (6.30)
  2. Update the image estimate  $\mathbf{m}^j$  from equation (6.23) along with the diagonal elements of  $\mathbf{R}^j$
  3. Check for convergence using (6.32)

For  $256 \times 256$  images this algorithm implemented in Matlab requires 3-5 minutes on a Pentium 4 3.40GHz personal computer. This is 2-5 times faster than the algorithm in [12]. This difference in speed is attributed to the smaller number of iterations required by the conjugate-gradient based implementation used herein.

## 6.6 Numerical Experiments

We demonstrate the value of the proposed restoration approach by testing it in experiments with two well known  $256 \times 256$  input images: *Lena* and *Cameraman*. Every image is blurred with three types of blur; the first blur has the shape of a Gaussian function with shape parameter 9, the second is uniform with support a rectangular region of dimension  $9 \times 9$  and the third is pyramidal blur with PSF  $[1\ 4\ 6\ 4\ 1]^T [1\ 4\ 6\ 4\ 1] / 256$ . The blurred signal to noise ratio (*BSNR*) defined as follows was used to quantify the noise level:

$$BSNR = 10 \log_{10} \frac{\|\mathbf{H}\mathbf{f}\|_2^2}{N\sigma^2},$$

where  $\sigma^2$  is the variance of the additive white Gaussian noise (AWGN). Three levels of AWGN were added to the blurred images with  $BSNR = 40, 30$  and  $20$  dB, respectively. Thus, in total 18 image restoration experiments were conducted.

As performance metric, the improvement in Signal to Noise Ratio ( $ISNR$ ) was used:

$$ISNR = 10 \log_{10} \frac{\|\mathbf{f} - \mathbf{g}\|_2^2}{\|\mathbf{f} - \hat{\mathbf{f}}\|_2^2}$$

where  $\mathbf{f}$ ,  $\mathbf{g}$  and  $\hat{\mathbf{f}}$  are the original, observed degraded and restored images, respectively.

We compare our restoration method with four recent TV-based algorithms: the algorithms in [17] and [18] abbreviated by BFO1 and BFO2, respectively and the algorithms in [8] abbreviated as BMK1 and BMK2. We also used compared it with the variational Bayesian algorithm in Chapter 5 which is abbreviated as CGLS.

The  $ISNR$  results of this comparison are shown in Tables 6.1, 6.2 and 6.3 for the experiments with uniform, Gaussian and pyramidal blurs, respectively. The  $ISNR$  results with algorithms abbreviated as BMK1, BMK2, BFO1 and BFO2 in Tables 6.1 and 6.2 are taken from [8]. We also show the restored images for 2 experiments in Figures 6.1 and 6.2. Looking carefully at the restored images by the herein proposed algorithm and comparing them to those in Chapter 5 we observe that they seem less "cartoon-like". In other words, that textured areas seem to have been better preserved. We also compared the images in the Fig. 6.1 and 6.2 in terms of their perceived visual quality. For this purpose we used the Visual Information Fidelity ( $VIF$ ) metric in [120]. The comparisons based on the  $VIF$  metric are also in favor of the herein proposed algorithm.

We note that the proposed algorithm produces the highest  $ISNR$  values for 6 out of 6 experiments with the  $9 \times 9$  uniform blur. For Gaussian blur,  $Gen-t$  provides the best result in 2 out of 6 experiments. For the pyramidal blur the proposed method gave the best results in 4 out of 6 experiments. Overall the proposed algorithm gave better  $ISNR$  results in 12 out of the 18 experiments we performed. Additionally, in the experiments where  $Gen-t$  does not produce the best  $ISNR$  results, the difference with the best  $ISNR$  is small. It is also worth noting that unlike the CGLS algorithm in Chapter 5 which "showed a preference" (produced better  $ISNR$  results) for the experiments with higher  $BSNRs$  the herein proposed algorithm does not seem to have any such "preference".

## 6.7 Conclusions

We presented a new promising image prior that is based on the generalized Student's-t pdf and a variational algorithm that estimates all the parameters of this model automatically and finds the restored image. We compared this restoration approach with that of Chapter 5 and previous state-of-the-art methods and found that it appears to be superior.

This prior can be extended in a number of ways. For example, more operators can be used in order to capture more directional dependencies of the image edges. Another option is to investigate how to relax the independence assumption between the different filter outputs and adjacent pixels in our image model.

## Appendix

In the VE-step the bound must be optimized with respect to  $\mathbf{R}$ ,  $\mathbf{m}$  and  $q(\mathbf{a})$ . The mean field approximation and Eq. (6.15) yield

$$L^b(q(\mathbf{a}), \theta_1, \theta_2) = \int \prod_{k=1}^{2P} q(\boldsymbol{\varepsilon}_k; \theta_1) \prod_{k=1}^P q(\mathbf{a}_k) \log F(\mathbf{y}, \boldsymbol{\varepsilon}, \mathbf{a}; \theta_2) d\boldsymbol{\varepsilon} d\mathbf{a} - \\ - \int \prod_{k=1}^{2P} q(\boldsymbol{\varepsilon}_k; \theta_1) \prod_{k=1}^P q(\mathbf{a}_k) \log \left( \prod_{k=1}^{2P} q(\boldsymbol{\varepsilon}_k; \theta_1) \prod_{k=1}^P q(\mathbf{a}_k) \right) d\boldsymbol{\varepsilon} d\mathbf{a},$$

where  $\theta_1 = [\mathbf{R}, \mathbf{m}]^T$  and  $\theta_2 = [\mathbf{u}, \lambda_1, \dots, \lambda_P, \nu_1, \dots, \nu_P]^T$ .

Because at this point we aim to optimize with respect to  $\theta_1$ , we operate on the function  $L'$ , which includes only the terms that depend on the parameters  $\theta_1$ :

$$L^b \propto L'(\theta_1)$$

$$L'(\theta_1) = \sum_{k=1}^P \int q(\boldsymbol{\varepsilon}_{2k-1}; \theta_1) q(\boldsymbol{\varepsilon}_{2k}; \theta_1) q(\mathbf{a}_k) \log \prod_{l=2k-1, 2k} (p(\mathbf{y}_l | \boldsymbol{\varepsilon}_l; \theta_2) d\boldsymbol{\varepsilon}_l)$$

$$\prod_{i=1}^N M(\boldsymbol{\varepsilon}_{2k}(i), \boldsymbol{\varepsilon}_{2k-1}(i), a_k(i); \boldsymbol{\theta}_2) d\mathbf{a}_k - \sum_{k=1}^{2P} \int q(\mathbf{e}_k; \boldsymbol{\theta}_1) \log q(\mathbf{e}_k; \boldsymbol{\theta}_1) d\mathbf{e}_k. \quad (\text{A.6.1})$$

The first sum is proportional to

$$\begin{aligned} & \propto \sum_{k=1}^{2P} \left\langle -\boldsymbol{\beta}(\mathbf{H}\boldsymbol{\varepsilon}_k - \mathbf{y}_k)^T \mathbf{Q}_k^{-T} \mathbf{Q}_k^{-1} (\mathbf{H}\boldsymbol{\varepsilon}_k - \mathbf{y}_k) - \lambda_{\lceil \frac{k}{2} \rceil} \boldsymbol{\varepsilon}_k^T \boldsymbol{\Lambda}_{\lceil \frac{k}{2} \rceil} \boldsymbol{\varepsilon}_k \right\rangle_{q(\mathbf{e}_k; \boldsymbol{\theta}_1) q(\mathbf{a}_k)} = \\ & = -\beta P \|\mathbf{H}\mathbf{m} - \mathbf{g}\|_2^2 - \sum_{k=1}^{2P} \lambda_{\lceil \frac{k}{2} \rceil} \mathbf{m}^T \mathbf{Q}_k^T \hat{\boldsymbol{\Lambda}}_{\lceil \frac{k}{2} \rceil} \mathbf{Q}_k \mathbf{m} - \text{trace} \left\{ \left( \beta P \mathbf{H}^T \mathbf{H} + \sum_{k=1}^{2P} \lambda_{\lceil \frac{k}{2} \rceil} \mathbf{Q}_k^T \hat{\boldsymbol{\Lambda}}_{\lceil \frac{k}{2} \rceil} \mathbf{Q}_k \right) \mathbf{R} \right\}, \quad (\text{A.6.2}) \end{aligned}$$

where  $\lceil \cdot \rceil$  denotes the 'ceiling' of a real number and  $\boldsymbol{\Lambda}_{\lceil \frac{k}{2} \rceil}$ ,  $\hat{\boldsymbol{\Lambda}}_{\lceil \frac{k}{2} \rceil}$  are diagonal matrix with elements

$$\boldsymbol{\Lambda}_{\lceil \frac{k}{2} \rceil}(i, i) = \frac{a_k(i)}{\sqrt{u_k(i)}}, \quad \hat{\boldsymbol{\Lambda}}_{\lceil \frac{k}{2} \rceil}(i, i) = \frac{\langle a_k(i) \rangle_{q(a_k(i))}}{\sqrt{u_k(i)}}, \quad i = 1, \dots, N.$$

The second integral is the entropy of a Gaussian function, which is proportional to

$$\int q(\boldsymbol{\varepsilon}_k; \boldsymbol{\theta}_1) \log q(\boldsymbol{\varepsilon}_k; \boldsymbol{\theta}_1) d\boldsymbol{\varepsilon}_k \propto \frac{1}{2} \log \det |\mathbf{R}|. \quad (\text{A.6.3})$$

Setting the derivative of  $L'$  w.r.t  $\mathbf{R}$  equal to zero using Eq. (A.6.1)-(A.6.3) yields

$$\begin{aligned} \frac{\partial L'(\boldsymbol{\theta}_1)}{\partial \mathbf{R}} = 0 & \Rightarrow \frac{\partial \text{trace} \left\{ \beta P \mathbf{H}^T \mathbf{H} \mathbf{R} + \sum_{k=1}^{2P} \lambda_{\lceil \frac{k}{2} \rceil} \mathbf{Q}_k^T \hat{\boldsymbol{\Lambda}}_{\lceil \frac{k}{2} \rceil} \mathbf{Q}_k \mathbf{R} \right\} - P \partial \log \det |\mathbf{R}|}{\partial \mathbf{R}} = 0 \\ & \Rightarrow \beta P \mathbf{H}^T \mathbf{H} + \sum_{k=1}^{2P} \lambda_{\lceil \frac{k}{2} \rceil} \mathbf{Q}_k^T \hat{\boldsymbol{\Lambda}}_{\lceil \frac{k}{2} \rceil} \mathbf{Q}_k - P \mathbf{R}^{-1} = 0 \Rightarrow \mathbf{R} = \left( \beta \mathbf{H}^T \mathbf{H} + \frac{1}{P} \sum_{k=1}^{2P} \lambda_{\lceil \frac{k}{2} \rceil} \mathbf{Q}_k^T \hat{\boldsymbol{\Lambda}}_{\lceil \frac{k}{2} \rceil} \mathbf{Q}_k \right)^{-1}. \end{aligned}$$

Similarly, using Eq. (A.6.2), we find that the optimum for the mean:

$$\frac{\partial L'(\boldsymbol{\theta}_1)}{\partial \mathbf{m}} = 0 \Rightarrow \mathbf{m} = \beta \mathbf{R} \mathbf{H}^T \mathbf{g}.$$

The final part of the VE-step is the optimization w.r.t. the function  $q(\mathbf{a})$ . It is straightforward to verify that this is achieved when

$$q(\mathbf{a}) = \frac{\exp\left(\langle \log F(\mathbf{y}, \boldsymbol{\varepsilon}, \mathbf{a}) \rangle_{q(\boldsymbol{\varepsilon})}\right)}{\int \exp\left(\langle \log F(\mathbf{y}, \boldsymbol{\varepsilon}, \mathbf{a}) \rangle_{q(\boldsymbol{\varepsilon})}\right) d\mathbf{a}} = \prod_{\kappa=1}^P \prod_{i=1}^N q(\mathbf{a}_\kappa(i)).$$

The product form is due to

$$\begin{aligned} \exp\langle \log F(\mathbf{y}, \boldsymbol{\varepsilon}, \mathbf{a}) \rangle_{q(\boldsymbol{\varepsilon})} &\propto \\ \prod_{\kappa=1}^P \prod_{i=1}^N (a_\kappa(i))^{\frac{V_\kappa}{2}+2-1} \exp\left\{-\frac{V_\kappa}{2}a_\kappa(i) - \lambda_\kappa \sqrt{u_\kappa(i)}a_\kappa(i)\right\}. \end{aligned}$$

Hence, each  $q(a_\kappa(i))$  is a Gamma distribution:

$$q(a_\kappa(i)) = \text{Gamma}\left(a_\kappa(i); \frac{V_\kappa}{2} + 2, \frac{V_\kappa}{2} + \lambda_\kappa \sqrt{u_\kappa(i)}\right).$$



Table 6.1: *ISNR* 's for the experiments using uniform blur  $9 \times 9$ .

Uniform blur $9 \times 9$		Lena	Cameraman
<i>BSNR</i> (dB)	Method	<i>ISNR</i> (dB)	
<i>BSNR</i> = 40	<i>Gen-t</i>	<b>8.52</b>	<b>9.61</b>
	<i>CGLS</i>	8.49	9.53
	<i>BMK1</i>	8.34	8.55
	<i>BMK2</i>	8.35	8.25
	<i>BFO1</i>	8.42	8.57
	<i>BFO2</i>	8.37	8.46
<i>BSNR</i> = 30	<i>Gen-t</i>	<b>6.25</b>	<b>6.55</b>
	<i>CGLS</i>	6.10	6.29
	<i>BMK1</i>	6.08	5.68
	<i>BMK2</i>	5.64	4.65
	<i>BFO1</i>	5.89	5.41
	<i>BFO2</i>	5.58	4.38
<i>BSNR</i> = 20	<i>Gen-t</i>	<b>4.24</b>	<b>3.55</b>
	<i>CGLS</i>	3.98	3.33
	<i>BMK1</i>	4.09	3.31
	<i>BMK2</i>	4.14	2.12
	<i>BFO1</i>	3.72	2.42
	<i>BFO2</i>	3.15	1.94

Table 6.2: *ISNR* 's for the experiments using Gaussian blur with variance 9.

Gaussian blur variance 9		Lena	Cameraman
<i>BSNR</i> (dB)	Method	<i>ISNR</i> (dB)	
<i>BSNR</i> = 40	<i>Gen-t</i>	4.64	3.49
	<i>CGLS</i>	<b>4.86</b>	<b>3.45</b>
	<i>BMK1</i>	4.72	<b>3.51</b>
	<i>BMK2</i>	4.50	3.27
	<i>BFO1</i>	4.78	3.39
	<i>BFO2</i>	4.49	3.26
<i>BSNR</i> = 30	<i>Gen-t</i>	<b>4.08</b>	2.81
	<i>CGLS</i>	3.89	2.74
	<i>BMK1</i>	3.87	<b>2.89</b>
	<i>BMK2</i>	3.56	2.47
	<i>BFO1</i>	3.87	2.63
	<i>BFO2</i>	3.55	2.41
<i>BSNR</i> = 20	<i>Gen-t</i>	<b>3.09</b>	2.07
	<i>CGLS</i>	2.76	1.86
	<i>BMK1</i>	3.02	2.13
	<i>BMK2</i>	2.47	<b>2.23</b>
	<i>BFO1</i>	2.87	1.72
	<i>BFO2</i>	2.42	1.42

Table 6.3: *ISNR* 's for the experiments using pyramidal blur.

Pyramidal blur		Lena	Cameraman
<i>BSNR</i> (dB)	Method	<i>ISNR</i> (dB)	
<i>BSNR</i> = 40	<i>Gen-t</i>	6.81	<b>6.80</b>
	<i>CGLS</i>	<b>7.02</b>	6.40
	<i>BFO1</i>	5.56	6.07
<i>BSNR</i> = 30	<i>Gen-t</i>	4.67	<b>4.60</b>
	<i>CGLS</i>	<b>4.81</b>	4.25
	<i>BFO1</i>	4.52	4.35
<i>BSNR</i> = 20	<i>Gen-t</i>	<b>3.12</b>	<b>2.97</b>
	<i>CGLS</i>	3.03	2.75
	<i>BFO1</i>	3.01	2.60



(a)



(b)

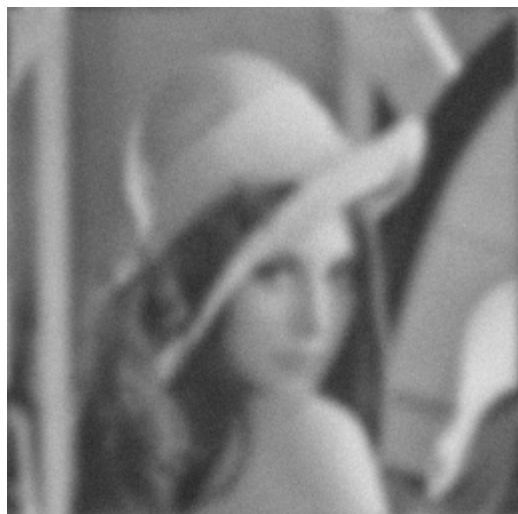


(c)



(d)

Figure 6.1: Experiment on *Cameraman* image with uniform  $9 \times 9$  blur and  $BSNR = 30$ ;  $ISNR$  and  $VIF$  [120] comparisons: (a) Degrated image,  $VIF = 2.01$  (b) restored with spatially invariant prior, [94],  $ISNR = 3.29$  and  $VIF = 2.22$ , (c) restored image with method in Chapter 5,  $ISNR = 5.88$  and  $VIF = 4.52$ , (d) restored image with the proposed algorithm,  $ISNR = 6.55$  and  $VIF = 4.55$ .



(a)



(b)



(c)



(d)

Figure 6.2: Experiment on *Lena* image with Gaussian blur (variance 9) and  $BSNR = 20$  ;

$ISNR$  and  $VIF$  [120] comparisons: (a) Degraded image,  $VIF = 0.04$  , (b) Restored with spatially invariant prior, [94],  $ISNR = 2.32$  and  $VIF = 0.31$  , (c) restored image with method in Chapter 5,  $ISNR = 2.76$  and  $VIF = 0.38$  , (d) restored image with the proposed algorithm,  $ISNR = 3.14$  and  $VIF = 0.39$  .

# CHAPTER 7. SUPER-RESOLUTION BASED ON FAST REGISTRATION AND MAXIMUM A POSTERIORI RECONSTRUCTION

- 
- 7.1. Introduction
  - 7.2. Imaging model
  - 7.3. Image prior model
  - 7.4. Pre-processing step of the super-resolution algorithm
  - 7.5. Maximum a posteriori (MAP) reconstruction
  - 7.6. Experiments
  - 7.7. Conclusions and future work
- 

In this chapter we propose a maximum a posteriori (MAP) framework for the super-resolution problem. The main contributions of this method are two; first, the use of the locally adaptive edge preserving prior of Chapters 4 and 5 to the super resolution problem. Second, an efficient two-step reconstruction methodology is proposed that begins with an initial registration using only the low resolution degraded observations. This is followed by a fast iterative algorithm implemented in the discrete Fourier transform domain in which the restoration, interpolation and the registration subtasks of this problem are preformed simultaneously. We present examples with both synthetic and real data that demonstrate the advantages of the proposed framework. [29].

## 7.1 Introduction

The first contribution presented in this chapter is that we utilize for the super resolution problem the prior presented in Chapters 4 and 5. With this prior the super-resolution problem is regularized in a spatially adaptive manner. The first level of this model captures the

correlations, while the second level provides a description of the local image edge structure in different directions. Thus, using this prior model it is possible to reconstruct the images without smoothed edges or ringing artifacts in the vicinity of edges, which is very usual in reconstruction of high-resolution image.

The second contribution in this method is a new two-step reconstruction algorithm. In the work in [133] the imaging model assumes only shifts and does not incorporate rotations. Apart from this, the registration task was extremely slow, because registration was performed using the high-resolution image as it was gradually reconstructed. Furthermore, it is based on a method that uses only first order derivatives. The first stage of the herein proposed methodology is a preprocessing step that approximately registers the degraded low-resolution observations. These “almost-registered” low-resolution observations are used subsequently by an iterative algorithm which simultaneously reconstructs the high-resolution images and finds their registration parameters. We propose this sub-optimal two-stage approach in order to speed up the super-resolution algorithm. Thus, the MAP functional is maximized based on coarse estimation of rotation and translation between image pairs. We have found that such coarse estimation provides sufficient accuracy to effectively remove the rotational and coarse (super-pixel) translational motion between image pairs. This algorithm is implemented entirely in the discrete Fourier transform (DFT) domain. Furthermore, the registration sub-task is based on the Newton-Raphson (NR) algorithm that utilizes analytically calculated first and second order derivatives and converges rapidly since NR algorithms exhibit quadratic convergence [97]. The purpose of the preprocessing step is to ameliorate one of the main difficulties of NR methods which are known to be effective only when they are initialized close to the solution.

The rest of this paper is organized as follows. In sections 7.2 and 7.3 we present the imaging model and the proposed image prior models, respectively. In section 7.4 we describe the pre-registration step and in section 7.5 the MAP based restoration algorithm is presented. In section 7.6 we provide experimental results with synthetic and real data that demonstrate the properties of our algorithm. Finally, in section 7.7 we provide conclusions and thoughts for future research.

## **7.2 Imaging model**

A linear imaging model is assumed. We denote as  $d$  the integer decimation factor. In other words, the imaging model assumes a high resolution image of size  $N_H \times 1$ , where  $N_H = dN$ . This model also assumes as observations  $P$  low resolution images of size  $N \times 1$  by applying the  $PN \times N_H$  degradation operator  $\mathbf{B}$  to the high resolution image. Then, white noise is added at each observation. Let  $\mathbf{y}$  be a  $PN \times 1$  vector, containing the  $P$  low resolution images  $\mathbf{y}_i$ :

$$\mathbf{y} = [\mathbf{y}_1^T \quad \mathbf{y}_2^T \quad \cdots \quad \mathbf{y}_P^T]^T,$$

where  $\mathbf{y}_i$  is a  $N \times 1$  vector, representing a low resolution image. Using this notation, the observations are given by:

$$\mathbf{y} = \mathbf{B}\mathbf{x} + \mathbf{n}, \quad (7.1)$$

where  $\mathbf{x}$  the (unknown) original  $N_H \times 1$  high-resolution image to be estimated,  $\mathbf{B}$  is a  $PN \times N_H$  degradation matrix and  $\mathbf{n} = [\mathbf{n}_1^T \quad \mathbf{n}_2^T \quad \cdots \quad \mathbf{n}_P^T]^T$  a  $PN \times 1$  vector consisting of  $P$   $N \times 1$  additive white noise vectors. We assume Gaussian statistics for the noise given by  $\mathbf{n}_i \sim N(\mathbf{0}, \beta_i^{-1} \mathbf{I})$ ,  $i=1, \dots, P$ , where  $\mathbf{0}$  is a  $N \times 1$  vector with zeros,  $\mathbf{I}$  the  $N \times N$  identity matrix respectively, and  $\beta_i^{-1}$ ,  $i=1, \dots, P$ , are the noise variances of the observations that are assumed unknown and statistically independent with each other. The degradation operator  $\mathbf{B}$  is given by:

$$\mathbf{B} = [\mathbf{B}_1^T \quad \cdots \quad \mathbf{B}_P^T]^T,$$

where  $\mathbf{B}_i = \mathbf{D}\mathbf{H}_i\mathbf{S}(\delta_i)\mathbf{R}(\theta_i)$  for  $i=1, \dots, P$ . The matrix  $\mathbf{D}$  is the known  $N \times N_H$  decimation matrix.  $\mathbf{H}_i$ ,  $i=1 \dots P$ , are the shift-invariant  $N_H \times N_H$  blurring convolutional operators, and  $\mathbf{S}(\delta_i)$ , for  $i=1, \dots, P$ , are the  $N_H \times N_H$  shift-invariant shifting operators. Each  $\delta_i$  is a scalar which represents translation (with respect to the first image) and is assumed unknown. The shift operator,  $\mathbf{S}(\delta_i)$ , is the Shannon interpolation operator which is shift-invariant [32]. The impulse response of the shift operator is given by:

$$S_{shift}(m; \delta_i) = \frac{\sin(\pi(m - \delta_i))}{\pi(m - \delta_i)}, \quad m = 1, \dots, N.$$



The shift-invariant operators are assumed circulant. This is very useful for computational purposes because such matrices can be easily diagonalized in the DFT domain. One difficulty that arises in the super resolution problem is the decimation operator which is not square and thus not circulant. In this work we take advantage of the simple form of this matrix, and, despite its non-circulant nature, we obtain tractable calculations in the DFT domain.

Lastly, the  $N_H \times N_H$  matrix  $\mathbf{R}(\theta_i)$  represents the rotation of each observation relative to the unknown ideal image  $\mathbf{x}$ . The imaging model assumes that image  $i$  is a rotated (as well as shifted) version of the first image, with angle  $\theta_i$ . Using all the above definitions, Eq. (7.1) can be rewritten as such:

$$\mathbf{y}_i = \mathbf{B}_i \mathbf{x} + \mathbf{n}_i = \mathbf{D} \mathbf{H}_i \mathbf{S}(\delta_i) \mathbf{R}(\theta_i) \mathbf{x} + \mathbf{n}_i, \text{ for } i = 1, \dots, P. \quad (7.2)$$

### 7.3 Image prior model

Since we utilize a MAP algorithm, a prior for the image is necessary. The prior used here is non-stationary and has been used with success in other image processing problems [27] and [26]. This image prior model assumes that the first order differences of the image  $\mathbf{x}$  in four directions,  $0^0$ , and  $90^0$  respectively, are given by:

$$\varepsilon^1(i, j) = \mathbf{x}(i, j) - \mathbf{x}(i, j+1), \quad \varepsilon^2(i, j) = \mathbf{x}(i, j) - \mathbf{x}(i+1, j) \quad (7.3)$$

with  $\varepsilon^k(i, j)$   $k=1,2$ , the difference residuals for the image location  $(i, j)$ . The above equations can be also written in matrix vector form for the entire image as  $\mathbf{Q}^k \mathbf{x} = \boldsymbol{\varepsilon}^k$ ,  $k=1,2$  where  $\mathbf{Q}^k$  are the  $N_H \times N_H$  directional difference operators for  $N_H \times 1$  images. Without loss of generality, in what follows, for convenience, we will use one dimensional notation; in other words, we assume  $\boldsymbol{\varepsilon}^k = [\varepsilon_1^k \quad \varepsilon_2^k \quad \dots \quad \varepsilon_{N_H}^k]^T$ . We also assume that the residuals have Gaussian statistics according to  $\varepsilon_i^k \sim N(0, (a_i^k)^{-1})$ , for  $i=1, \dots, N_H$  and  $k=1,2$  where  $a_i^k$  is the inverse variance of  $\varepsilon_i^k$ .

For the inverse variances (i.e. the  $a_i^k$ 's) we introduce the notation  $\mathbf{A}^k = \text{diag} \{a_1^k, a_2^k, \dots, a_N^k\}$  a  $N_H \times N_H$  diagonal matrix,  $\tilde{\mathbf{A}} = \text{diag} \{\mathbf{A}^1, \mathbf{A}^2\}$  a  $2N_H \times 2N_H$  diagonal matrix and

$\tilde{\mathbf{a}} = \left[ (\mathbf{a}^1)^\top, (\mathbf{a}^2)^\top \right]^\top$  a  $2N_H \times 1$  vector, consisting of four vectors  $\mathbf{a}^k = [a_1^k, a_2^k, \dots, a_{N_H}^k]^\top$ . Also

for the errors we use the notation  $\tilde{\boldsymbol{\varepsilon}} = \left[ (\boldsymbol{\varepsilon}^1)^\top, (\boldsymbol{\varepsilon}^2)^\top \right]^\top$ . We assume that the errors in each direction and at each pixel location are independent. This is based on the assumption that at each pixel location an edge can occur at any direction independently of what happens in adjacent pixels. This assumption makes subsequent calculations tractable. Thus, the joint density for the errors is Gaussian and is given as

$$p(\tilde{\boldsymbol{\varepsilon}}; \tilde{\mathbf{a}}) \propto \prod_{k=1}^2 \prod_{i=1}^{N_H} (a_i^k)^{1/2} \exp\left(-0.5 \left( (\boldsymbol{\varepsilon}^k)^\top \mathbf{A}^k \boldsymbol{\varepsilon}^k \right)\right) = \prod_{k=1}^2 \prod_{i=1}^{N_H} (a_i^k)^{1/2} \exp\left(-0.5 (\tilde{\boldsymbol{\varepsilon}}^\top \tilde{\mathbf{A}} \tilde{\boldsymbol{\varepsilon}})\right).$$

To relate  $\tilde{\boldsymbol{\varepsilon}}$  with the image  $\mathbf{x}$  we define the  $2N_H \times N$  operator  $\tilde{\mathbf{Q}} = \left[ (\mathbf{Q}^1)^\top, (\mathbf{Q}^2)^\top \right]^\top$ . Then, the relation between the image and the residuals is  $\tilde{\boldsymbol{\varepsilon}} = \tilde{\mathbf{Q}}\mathbf{x}$ . Based on this relation and  $p(\tilde{\boldsymbol{\varepsilon}}; \tilde{\mathbf{a}})$  we can define an *improper prior* (one that does not integrate to 1) for the image  $\mathbf{x}$  [26]. This prior is given by:

$$p(\mathbf{x}; \tilde{\mathbf{a}}) \propto \prod_{k=1}^2 \prod_{i=1}^{N_H} (a_i^k)^{1/4} \exp\left(-0.5 \left( (\tilde{\mathbf{Q}}\mathbf{x})^\top \tilde{\mathbf{A}} \tilde{\mathbf{Q}}\mathbf{x} \right)\right) = \prod_{k=1}^2 \prod_{i=1}^{N_H} (a_i^k)^{1/4} \exp\left(-0.5 \left( (\mathbf{Q}^k\mathbf{x})^\top \mathbf{A}^k \mathbf{Q}^k\mathbf{x} \right)\right). \quad (7.4)$$

The role of the parameters  $a_i^k$  is to capture the directional variation structure of the image. More specifically, a large variance (small  $a_i^k$ ) indicates the presence of a large variation along the direction of the difference, in other words an edge perpendicular to this direction. The introduction of the spatially varying  $a_i^k$  scales down the differences of adjacent pixels in regions of image discontinuities. As a result this prior maintains edges and suppresses noise in smooth areas of the image.

The drawback of this prior as described thus far is that it introduces  $2N_H$  parameters  $a_i^k$  that have to be estimated from  $PN$  observations. This is clearly not a desirable situation from an estimation point of view. To address this, we employ the Bayesian paradigm and consider  $a_i^k$  as random variables (instead of parameters) and introduce Gamma hyper-priors for them. In the case of a stationary model where all  $a_i^k$  are equal, the over-parameterization problem does

not exist, and it is rather straightforward to obtain good estimates for the unknown parameters using even maximum likelihood (ML).

We consider the following parameterization for the Gamma hyper-prior:

$$p(a_i^k; m_k, l_k) \propto (a_i^k)^{\frac{l_k-2}{2}} \exp\{-m_k(l_k-2)a_i^k\}. \quad (7.5)$$

For such a representation the mean and variance of the Gamma pdf are given by  $E[a_i^k] = l_k(2m_k(l_k-2))^{-1}$ , and  $Var[a_i^k] = l_k(2m_k^2(l_k-2)^2)^{-1}$  respectively. This representation is used because the value of the parameter  $l_k$  can be also interpreted as the level of confidence to the prior knowledge provided by the Gamma hyper prior. More specifically, as  $l_k \rightarrow \infty$ ,  $E[a_i^k] \rightarrow (2m_k)^{-1}$  and  $Var[a_i^k] \rightarrow 0$ . In other words, the prior becomes very informative and restrictive, resulting in  $a_i^k = (2m_k)^{-1} \forall i$ . In contrast, when  $l_k \rightarrow 2$  then both  $E[a_i^k] \rightarrow \infty$  and  $Var[a_i^k] \rightarrow \infty$ , thus, in this case, the prior becomes uninformative and does not influence the values of the  $a_i^k$ 's.

#### 7.4 Pre-processing step of the super-resolution algorithm

For this imaging model, the non-circulant nature of the rotation matrix  $\mathbf{R}$  renders computationally impractical simultaneous registration and restoration for large images. In contrast, all other matrices used in both the imaging and image prior model have characteristics that can be exploited in the DFT domain to render both tasks computationally very efficient. Particularly, the blurring  $\mathbf{H}$  and shift matrices  $\mathbf{S}$  are circulant, hence diagonal in the DFT domain. As mentioned before, the decimation matrix  $\mathbf{D}$ , which is not circulant, has a convenient structure in the DFT domain that helps bypass computational difficulties. Finally, matrices  $\mathbf{Q}^k$  and  $\mathbf{A}^k$  of the image prior are circulant and diagonal, respectively. For this combination one can exploit the diagonal structure by *alternating* calculations in the DFT and spatial domain.

To bypass the problems with the rotation, a pre-processing step is performed before the super-resolution algorithm. In this step we estimate the registration parameters between the low resolution observations. At this point it is important to notice that as far as the rotation is

concerned the rotations between the degraded low resolution and the high resolution images of the imaging model in Eq. (5.2) are the same. However, as far as the shifts are concerned they are *not* since the shifts of the low resolution images must be also multiplied by the decimation factor. Thus, these parameters in the preprocessing step will be called  $\delta'_i$  and  $\theta_i$ ,  $i = 2, \dots, P$ , for translation and rotation, respectively. Using this notation, we assume that image  $\mathbf{y}_i$  resulted by applying both translation and rotation with respect to the first image  $\mathbf{y}_1$  (or the reverse). In other words, we have

$$\mathbf{y}_i = \mathbf{S}'(\delta'_i) \mathbf{R}'(\theta_i) \mathbf{y}_1 \quad \text{or} \quad \mathbf{y}_1 = \mathbf{S}'(-\delta'_i) \mathbf{R}'(-\theta_i) \mathbf{y}_i,$$

where  $\mathbf{R}'$  and  $\mathbf{S}'$  are the  $N \times N$  rotation and shift matrices respectively, smaller than their respective  $N_H \times N_H$  matrices  $\mathbf{R}$  and  $\mathbf{S}$ . Thus, image  $\mathbf{y}_1$  is the reference image.

We define the vector that represents the difference between the registered image  $i$  and the reference image to be

$$\mathbf{L}_i = \mathbf{S}'(-\delta'_i) \mathbf{R}'(-\theta_i) \mathbf{y}_i - \mathbf{y}_1, \quad \text{for } i = 2, \dots, P.$$

Mathematically speaking, in this registration pre-processing step, we aim to estimate the registration parameters by minimizing the quantity in the following equation:

$$\left[ \hat{\delta}'_i, \hat{\theta}_i \right] = \arg \min_{[\delta'_i, \theta_i]} \|\mathbf{L}_i\|_2^2, \quad \text{for } i = 2, \dots, P.$$

The minimization is achieved using the simplex search method [83]. Having computed the registration parameters  $\delta'_i$  and  $\theta_i$ , at the end of the pre-processing step the low resolution observations  $\mathbf{y}_i$  are replaced by the “almost-registered” low resolution images given by

$$\mathbf{z}_i = \mathbf{S}'(\text{int}[\hat{\delta}'_i + 0.5]) \mathbf{R}'(\hat{\theta}_i) \mathbf{y}_i, \quad i = 2, \dots, P,$$

where  $\text{int}[\cdot]$  denotes the integer part of the real number. This is intentional because low resolution images that are shifted by a fraction of a pixel are required in order to achieve super-resolution reconstruction [32]. For the rest of the paper, we assume as observations the registered versions  $\mathbf{z}_i$  of the initially observed images  $\mathbf{y}_i$ . We also define as  $\mathbf{z}$  the vector that contains all the  $\mathbf{z}_i$  as

$$\mathbf{z} = \left[ \mathbf{z}_1^T \quad \mathbf{z}_2^T \quad \cdots \quad \mathbf{z}_P^T \right]^T.$$

In this way, the rotation is removed from the observations. Thus, the rotation matrices can be omitted from the imaging model used for super-resolution reconstruction, described in the next section.

### 7.5 Maximum a posteriori (MAP) reconstruction

The super-resolution image  $\mathbf{x}$  is estimated from the observations  $\mathbf{z}$ , (after the preprocessing step), utilizing a MAP approach in which we infer simultaneously  $\tilde{\mathbf{a}}$ ,  $\mathbf{x}$  and  $\boldsymbol{\zeta} = [\zeta_2 \ \cdots \ \zeta_p]^T$  where the registration parameters have changed to according to  $\delta_i = \delta'_i + \zeta_i$ . At this point we must note that even in the absence of noise

$$\mathbf{z}_i = \mathbf{B}_i(\zeta_i)\mathbf{x} + \mathbf{n}_i = \mathbf{D}\mathbf{H}_i\mathbf{S}(\zeta_i)\mathbf{x} + \mathbf{n}_i, \quad \text{for } i = 1, \dots, P,$$

To correct this, we make the assumption that the coarsely registered  $\mathbf{z}_i$ , using rotation and translation, satisfies the equation  $\mathbf{z}_i = \mathbf{D}\mathbf{H}_i\mathbf{S}(\delta_i)\mathbf{x} + \mathbf{n}'_i$  where  $\mathbf{n}'_i$  is an error term. Thus, the imaging model that is finally solved by the MAP algorithm is

$$\mathbf{z}_i = \mathbf{D}\mathbf{H}_i\mathbf{S}(\zeta_i)\mathbf{x} + \mathbf{w}_i, \quad \text{for } i = 1, \dots, P,$$

where  $\mathbf{w}_i = \mathbf{n}'_i + \mathbf{n}_i$  the new error term which is assumed WGN with precision  $b_i$ .

MAP estimation is based on maximization of the posterior probability. Thus, based on Bayes' theorem we have:

$$p(\mathbf{x}, \tilde{\mathbf{a}} | \mathbf{z}; \mathbf{b}, \mathbf{m}, \mathbf{l}, \boldsymbol{\zeta}) \propto p(\mathbf{z}, \mathbf{x}, \tilde{\mathbf{a}}; \mathbf{b}, \mathbf{m}, \mathbf{l}, \boldsymbol{\zeta}) = p(\mathbf{z} | \mathbf{x}, \tilde{\mathbf{a}}; \mathbf{b}, \boldsymbol{\zeta}) p(\mathbf{x} | \tilde{\mathbf{a}}) p(\tilde{\mathbf{a}}; \mathbf{m}, \mathbf{l}),$$

where:

$$\mathbf{m} = [m_1, m_2, m_3, m_4]^T, \mathbf{l} = [l_1, l_2, l_3, l_4]^T, \mathbf{b} = [b_1 \cdots b_p].$$

Maximizing the quantity  $p(\mathbf{x}, \tilde{\mathbf{a}} | \mathbf{z}; \mathbf{b}, \mathbf{m}, \mathbf{l}, \boldsymbol{\zeta})$  with respect to  $\mathbf{x}$ ,  $\tilde{\mathbf{a}}$  and  $\boldsymbol{\zeta}$  is equivalent to minimizing the negative logarithm:

$$\begin{aligned}
J_{MAP}(\mathbf{x}, \tilde{\mathbf{a}}, \zeta) &= -\log p(\mathbf{z}, \mathbf{x}, \tilde{\mathbf{a}}; \mathbf{b}, \mathbf{m}, \mathbf{l}, \zeta) = -\log p(\mathbf{z} | \mathbf{x}, \tilde{\mathbf{a}}; \mathbf{b}, \zeta) - \log p(\mathbf{x} | \tilde{\mathbf{a}}) - \log p(\tilde{\mathbf{a}}; \mathbf{m}, \mathbf{l}) = \\
&= -\frac{N}{2} \sum_{i=1}^P \log b_i + \frac{1}{2} \sum_{i=1}^P b_i \|\mathbf{B}_i(\zeta_i) \mathbf{x} - \mathbf{z}_i\|_2^2 - \frac{1}{4} \sum_{k=1}^2 \sum_{i=1}^N \log a_i^k + \frac{1}{2} \sum_{k=1}^2 \sum_{i=1}^N (\mathbf{Q}^k \mathbf{x})^T \mathbf{A}^k \mathbf{Q}^k \mathbf{x} \\
&\quad - \sum_{k=1}^2 \left( \frac{l_k - 2}{2} \sum_{i=1}^N \log a_i^k \right) + \sum_{k=1}^2 \left( m_k (l_k - 2) \sum_{i=1}^N a_i^k \right). \tag{7.6}
\end{aligned}$$

To minimize the above function with respect to  $\mathbf{x}$ ,  $\tilde{\mathbf{a}}$  and  $\zeta$  we adopt an iterative scheme that sets alternatively the gradient with respect to  $\mathbf{x}$ ,  $\tilde{\mathbf{a}}$  and  $\zeta$  equal to zero.

Setting  $\nabla_{\tilde{\mathbf{a}}} J_{MAP}(\mathbf{x}, \tilde{\mathbf{a}}, \zeta) = 0$  yields:

$$\left( a_i^k \right)^* = \frac{\left( \frac{1}{4} + \frac{1}{2} (l_k - 2) \right)}{\left( \frac{1}{2} (\varepsilon_i^k)^2 + m_k (l_k - 2) \right)}. \tag{7.7}$$

The observation of the previous section that the parameters  $l_k$  express the degree of confidence to the prior can be viewed from another point when looking Eq. (7.7), the MAP estimates of the  $(a_i^k)$ . More specifically, when  $l_k \rightarrow \infty$ ,  $(a_i^k)^* = (2m_k)^{-1} \forall i$ ; thus the  $(a_i^k)^* \forall i$  are equal, and the image model becomes *stationary*. In contrast, when  $l_k \rightarrow 2$ ,  $(a_i^k)^* = \left( (\varepsilon_i^k)^2 \right)^{-1} \forall i$ ; thus the  $(a_i^k)^*$ 's are completely unaffected from the moderating effect of the Gamma hyper-prior and only follow the data. In this case the image model can be viewed as the ‘‘most non-stationary’’.

Setting  $\nabla_{\mathbf{x}} J_{MAP}(\mathbf{x}, \tilde{\mathbf{a}}, \zeta) = 0$  yields:

$$\mathbf{x}^* = \left( \sum_{i=1}^P b_i \mathbf{B}_i^T(\zeta_i) \mathbf{B}_i(\zeta_i) + \sum_{k=1}^2 (\mathbf{Q}^k)^T \mathbf{A}^k \mathbf{Q}^k \right)^{-1} \sum_{i=1}^P b_i \mathbf{B}_i^T(\zeta_i) \mathbf{z}_i. \tag{7.8}$$

Eq. (5.8) cannot be solved in closed form since analytical inversion of  $\sum_{i=1}^P b_i \mathbf{B}_i^T(\zeta_i) \mathbf{B}_i(\zeta_i) + \sum_{k=1}^2 (\mathbf{Q}^k)^T \mathbf{A}^k \mathbf{Q}^k$  is not possible due to the non-circulant nature of the matrices  $\mathbf{B}_i$  and  $\mathbf{A}^k$ . Thus, we resort to a numerical solution using a conjugate gradient algorithm [97]. In this algorithm, the space and DFT domains are alternated when expressions with circulant and diagonal matrices are computed. More specifically,

multiplications with circulant matrices (convolutions) are performed in the DFT domain while multiplications with diagonal matrices are performed in the space domain.

In the case of the registration parameters, it is not possible to find in closed form the  $\zeta$  that make the gradient  $\nabla_{\zeta} J_{MAP}$  equal to zero, or equivalently to minimize the quantity  $J_{MAP}$  with respect to  $\zeta$ :

$$\zeta^* = \arg \min_{\zeta} J_{MAP}(\mathbf{x}, \tilde{\mathbf{a}}, \zeta),$$

which can also be written as:

$$\zeta_i^* = \arg \min_{\delta_i} J_{MAP}(\zeta_i) = \arg \min_{\delta_i} \|\mathbf{B}_i(\zeta_i) \mathbf{x} - \mathbf{z}_i\|_2^2, \text{ for } i = 2, \dots, P, \quad (7.9)$$

where  $J_{MAP}(\zeta_i)$  denotes the part of  $J_{MAP}(\mathbf{x}, \tilde{\mathbf{a}}, \zeta)$  that depends on  $\zeta_i$ . Since  $\zeta_i^*$  cannot be found in closed form we resort to the Newton-Raphson algorithm. This method is chosen due to its convergence speed [97]. Registration is equivalent to the minimization task in Eq. (7.9).

By the definition of the matrix  $\mathbf{B}_i$  with  $\mathbf{R}(\theta_i) = \mathbf{I}$ , in Eq. (7.9) is

$$J_{MAP}(\zeta_i) = 2\mathbf{z}_i^T \mathbf{D}\mathbf{S}(\zeta_i) \mathbf{H}_i \mathbf{x} + \mathbf{x}^T \mathbf{H}_i^T \mathbf{S}^T(\zeta_i) \mathbf{D}^T \mathbf{D}\mathbf{S}(\zeta_i) \mathbf{H}_i \mathbf{x} + \mathbf{z}_i^T \mathbf{z}_i. \quad (7.10)$$

The DFT domain is used to evaluate (7.10), since it allows easy analytic calculations of the first and second derivatives of the objective function. Since the shift parameters are independent with each other, it is sufficient to demonstrate the derivatives for one  $\zeta_i$ . The details of the derivative calculation of  $J_{MAP}(\zeta_i)$  are given in the Appendix. With the derivatives calculated, the update scheme of the Newton-Raphson algorithm is

$$\zeta_i^{n+1} = \zeta_i^n - \frac{\partial J_{MAP}(\zeta_i^n)}{\partial \zeta_i} \left( \frac{\partial^2 J_{MAP}(\zeta_i^n)}{\partial \zeta_i^2} \right)^{-1}. \quad (7.11)$$

The shift parameters are initialized as  $\zeta_{init} = \hat{\boldsymbol{\delta}}' \cdot d$ , where  $\hat{\boldsymbol{\delta}}' = [\hat{\delta}'_2, \dots, \hat{\delta}'_p]^T$  are the shift parameters estimated in the preprocessing step (section 5.4) and  $d$  is the decimation factor. This initialization provides starting values close to the solution, which is essential for the convergence of the Newton-Raphson algorithm to the correct solution [97].

## 7.6 Experiments

In order to test the proposed methodology, we used both artificially generated and real data. We compared the new MAP super-resolution algorithm with the non-stationary prior with the E-M super-resolution algorithm in [133] that uses a stationary prior. We also compared our super-resolution algorithm with one that uses total variation (TV) regularization [48]. For this comparison a gradient descent algorithm was used given by

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \beta \left\{ \mathbf{B}^T (\mathbf{B}\mathbf{x}^{(k)} - \mathbf{z}) + \lambda \left( (\mathbf{Q}^1)^T \mathbf{v}_1^{(k)} + (\mathbf{Q}^2)^T \mathbf{v}_2^{(k)} \right) \right\} \quad (7.12)$$

where the superscript  $(k)$  denotes the iteration number,  $\mathbf{v}_m(i) = \text{sign}(\boldsymbol{\varepsilon}^m(i))$ , with  $\boldsymbol{\varepsilon}^1 = \mathbf{Q}^1 \mathbf{x}$  and  $\boldsymbol{\varepsilon}^2 = \mathbf{Q}^2 \mathbf{x}$  the first order horizontal and vertical differences of the image,  $\lambda$  the regularization parameter and  $\beta$  the step of the algorithm. In the following experiments the  $\lambda$  and  $\beta$  parameters were selected by trial and error to provide the best possible results. This is a difficult task. However, in general as  $\lambda$  increases the image becomes blurrier and the algorithm converges for smaller step  $\beta$ . For all methods we used the same registration algorithm. The results generated by Eq. (7.12) are not a comparison with the methodology presented in [48] since although similar priors are used the other aspects of the super-resolution algorithm (registration, chosen PSF) are different. Nevertheless, the authors of [48] have published results with the herein used data sets in [51] and [47] where the interested reader can resort. In the preprocessing step of the herein proposed algorithm the interpolation algorithm in [135] was used for rotation in order to handle boundary artifacts.

In the first experiment, eight  $128 \times 128$  low resolution images were generated by performing translation and rotation to the well-known ‘‘Cameraman’’ image of size  $256 \times 256$ , before blurring and then down-sampling by a factor of 2. The PSF of the blur was uniform  $5 \times 5$ . Lastly, noise was added, corresponding to  $SNR=20dB$  (the same for all images). This metric

is defined as  $SNR = 10 \log \frac{\|\mathbf{x}\|_2^2}{N_H \sigma^2}$  where  $\sigma^2$  is the variance of the additive noise and  $N_H$  is

the size of the image  $\mathbf{x}$ .



The Mean Square Error metric ( $MSE$ ) between the restored image and the original was used to evaluate the performance of the algorithm. The  $MSE$  is defined as  $MSE = \frac{\|\mathbf{x} - \hat{\mathbf{x}}\|_2^2}{N_H}$ , where  $\mathbf{x}$  and  $\hat{\mathbf{x}}$  are the original and estimated images, respectively.

Fig. 7.1.1.(a) shows one of the observed low resolution degraded image. In Fig. 7.1.2.(a), 7.1.2.(b) and 7.1.(c) we show the super-resolved images and the corresponding  $MSE$ 's, using the stationary prior in [133], TV regularization as implemented in Eq. (7.12), and the new algorithm based on the non stationary prior, respectively. Also, to demonstrate the robustness of the proposed registration methodology, we show the true and the estimated registration parameters in *Table 7.1*. We observed in all the experiments we performed with simulated data that the proposed pre-processing step estimated the rotation parameters with an accuracy of almost four decimal digits in degrees. The reconstructed super-resolved images assuming knowledge of the registration parameters are almost identical to their reconstructed counterparts using the estimated parameters. From these experiments we can draw two conclusions. First, the proposed non-stationary prior improves the reconstruction of the high-resolution images. Indeed, the  $MSE$  using the non-stationary model is significantly lower apart from the difference in the visual quality of the images. Second, the proposed two step registration methodology seems very accurate (when the image formation model is correct).

We also used the proposed super-resolution algorithm on two real data sets. The first contains 20 low resolution degraded images. In Fig. 7.2.1.(a) one of these images is shown. Their original size was slightly smaller than  $64 \times 64$ , so they were padded with zeros, extending their size exactly to  $64 \times 64$  pixels. In this data set the low resolution images were only translated and did not contain any rotations. Super-resolved images of double size ( $2x$ ) are shown in Fig. 7.2.2.(a), 7.2.2.(b) and 7.2.2.(c) using the stationary, TV regularization and the non-stationary priors, respectively.

The second set includes four low resolution degraded images that contain both translations and rotations and one of them is shown in Fig. 7.3.1.(a). Each low resolution image is of size  $128 \times 128$ . In order to test the ability of the proposed priors to reconstruct beyond the resolution of the available data, we quadrupled ( $4x$ ) the size of the reconstructed super-resolved images. The  $4x$ , super-resolved images with the stationary, TV regularization, and non-stationary prior are shown in Fig. 7.3.2.(a), 7.3.2.(b), and 7.3.2.(c), respectively.

In estimating the shape of the blur for the real data sets, a Gaussian-shaped blur was assumed. This choice was motivated by the observation that Gaussian shaped functions are smooth and have good approximation properties. The width of each blur was experimentally estimated using trial and error experiments. The width is captured by the variance of Gaussian PSF. For the first set, the values of the variances of the Gaussian shaped PSFs were found in the range [2.5-4] pixels and for the second the variance was set equal to 4.

To facilitate learning the proposed image model, we used equal  $b_i^{-1}$  for all  $i$  (additive noise variances) and equal  $m_k$  for all  $k$  obtained by learning a stationary SAR model [133]. The parameters  $m_k$  were obtained as  $m_k = 1/(2a_{STAT})$  where  $a_{STAT}$  the image model parameter of the stationary SAR model. The parameters  $l_k$  were selected to be equal to  $l = 2.1$  for the reconstruction of both the real data and synthetic data. This value was found by trial and error experiments. We observed that as  $l \rightarrow 2$  the reconstructed images assume a “cartoon” like appearance where large edges are preserved and areas with small variations are flattened out. When  $l \rightarrow \infty$ , as also explained previously, the reconstructed images assume the appearance of images that were reconstructed by a stationary prior model. In other words, at the expense of ringing in edges and noise amplification in smooth areas, textured areas can be better reconstructed. The selection of  $l = 2.1$  reflects our subjective choice between these two opposing trends. For the case of the TV regularization, the algorithm’s parameters were also found by trial and error. We set for the first experiment  $\beta = 0.05$ ,  $\lambda = 1$ , for the second  $\beta = 0.01$ ,  $\lambda = 1$ , and for the last one  $\beta = 0.1$ ,  $\lambda = 2$ .

The super-resolution estimates of  $\mathbf{x}$ ,  $a_i^k$  and  $\zeta_i$  were found by iterating between Eq. (7.7), (7.8) and (7.11) till convergence. In the presented experiments, the convergence criterion was

$$\frac{\|\mathbf{x}^t - \mathbf{x}^{t+1}\|_2}{\|\mathbf{x}^t\|_2} < \frac{10^{-3}}{\bar{b}},$$

where  $t$  denotes the iteration number and  $\bar{b}$  is the average of the inverse noise variances

$$\bar{b} = \frac{1}{P} \sum_{i=1}^P b_i.$$

Finally, we would like to note that the MAP function in Eq. (7.6), although derived using a completely different principle, can be viewed as a half-quadratic function that is generated

using a  $\varphi_{HL}$ -like potential function (Table II, pp. 302 in [31]) with appropriate choice of parameters, for details see [26]. The convergence of alternating direction minimization of half quadratic functions has been rigorously shown in [31]. It has been shown that if the generating potential function is strictly convex, and the null spaces of matrices  $\mathbf{B}$  and  $\mathbf{Q}^k$  do not intersect, the MAP function is convex. However, the  $\varphi_{HL}$ -like potential function is not convex thus the proposed alternating direction minimization converges to a local minimum. For this reason, good initialization of the algorithm is important.

## 7.7 Conclusions and future work

Inspection of the super-resolved images in Fig. 7.1.2.(a)-(c), 7.2.2.(a)-(c), and 7.3.2.(a)-(c) reveals that the resolution in every case has significantly been improved. The letters in the super resolved images (Fig. 7.2.2.(a)-(c) and 7.3.2.(a)-(c)) are now easily legible.

Furthermore, the images reconstructed using the proposed non-stationary prior, Fig. 7.1.2.(c), 7.2.2.(c), and 7.3.2.(c), are visually more pleasant and display less ringing at the edges as compared to both stationary and TV based super-resolution reconstruction. The *MSE* for the reconstructed images using non-stationary prior is also smaller than both the stationary and the TV based models. It is worth noticing that for the first experiment, the *MSE* results when using the real registration parameters are almost identical to that when the registration parameters are estimated. This demonstrates the robustness of the proposed algorithm regarding the registration parameters.

In what follows we report implementation times for the ‘‘Cameraman’’ experiment. Registration in the pre-processing step requires 26 minutes. One iteration of the stationary model based algorithm requires 4-5 seconds with almost 4 seconds the time for fast sub-pixel registration. One iteration of the non-stationary MAP algorithm requires about 38 seconds, out of which 4 seconds are required for fast sub-pixel registration, and the rest for 40 iterations of the conjugate gradient algorithm in Eq. (7.8). The TV algorithm requires about 1 second per iteration of the gradient algorithm in Eq. (7.12). These times we obtained using a Pentium 4 3.4GHz PC and a Matlab implementation.

In the future we plan to include a PSF estimation step in the formulation of this problem. This can be achieved by blind-deconvolutions methodologies. Faster rotation estimation in the pre-

processing step would be also desirable. In addition, it would be interesting to include the rotation matrix in the imaging model and estimate in parallel all the registration parameters. Furthermore, methodologies to better model the statistical errors if the imaging model is not accurate will be considered.

## Appendix

Assume the  $N_H \times N_H$  DFT matrix  $\mathbf{W}_1$  and the  $N \times N$  DFT matrix  $\mathbf{W}_2$ . Then  $\mathbf{X} = \mathbf{W}_1 \mathbf{x}$  and  $\mathbf{Z}_i = \mathbf{W}_2 \mathbf{z}_i$  are the DFTs of the vectors  $\mathbf{x}$  and  $\mathbf{z}_i$ , respectively. The matrices  $\Lambda_S = \mathbf{W}_1 \mathbf{S}(\delta_i) \mathbf{W}_1^{-1}$  and  $\Lambda_H = \mathbf{W}_1 \mathbf{H}_i \mathbf{W}_1^{-1}$  are diagonal due to the circulant nature of the matrices  $\mathbf{S}(\delta_i)$  and  $\mathbf{H}_i$ . It can also be shown that

$$\Lambda_D = \mathbf{W}_2 \mathbf{D} \mathbf{W}_1^{-1} = [\mathbf{I}_1, \mathbf{I}_2, \dots, \mathbf{I}_d] / d \quad (\text{A.7.1})$$

is a  $N \times N_H$  block matrix that contains  $d$  identity matrices of size  $N \times N$ . Then, we can write:

$$J_{MAP}(\zeta_i) = 2 \text{real} \left\{ \mathbf{Z}_i^H \Lambda_D \Lambda_S \Lambda_H \mathbf{X} \right\} + \mathbf{X}^H \Lambda_H^* \Lambda_S^* \Lambda_D^T \Lambda_D \Lambda_S \Lambda_H \mathbf{X} + \mathbf{Z}_i^H \mathbf{Z}_i. \quad (\text{A.7.2})$$

where the symbols ‘H’ denote the Hermitian and ‘\*’ the conjugate. For simplicity, the diagonal element of a matrix is denoted as  $[m]$ . Then we can write

$$J_{MAP}(\delta_i) \propto 2 \text{real} \left\{ \sum_{m=1}^N \mathbf{Z}_i^* [m] \mathbf{T}_i [m] \right\} + \sum_{m=1}^N \mathbf{T}_i^* [m] \mathbf{T}_i [m], \quad (\text{A.7.3})$$

where  $\mathbf{T}_i [m]$  are the elements of the  $N \times 1$  vector  $\mathbf{T}_i$  and they are

$$\mathbf{T}_i [m] \equiv \frac{\sum_{n=0}^{d-1} \left( \Lambda_S \left[ m + \frac{nN}{d} \right] \Lambda_H \left[ m + \frac{nN}{d} \right] \mathbf{X} \left[ m + \frac{nN}{d} \right] \right)}{d}. \quad (\text{A.7.4})$$

The evaluation of the first and second derivatives of Eq. (A.7.2) is very convenient in the DFT domain since the parameter  $\zeta_i$  is only in the diagonal elements of the matrix  $\Lambda_S$ . These elements, see for example [32], are equal to:

$$\Lambda_s[m] = \exp\{-2j\pi\zeta_i(m-1)/N\}, \text{ for } m = 1, \dots, N/2,$$

where  $j^2 = -1$ . The remaining elements are a “mirrored” version of the previous ones; in other words:

$$\Lambda_s[m] = \exp\{-2j\pi\zeta_i(N-m+1)/N\}, \text{ for } m = \frac{N}{2} + 1, \dots, N.$$

The first and second derivatives for the first half are respectively:

$$\frac{\partial \Lambda_s[m]}{\partial \zeta_i} = \frac{-2j\pi(m-1)}{N} \exp\{-2j\pi\zeta_i(m-1)/N\}, \text{ for } m = 1, \dots, N/2, \quad (\text{A.7.5})$$

$$\frac{\partial^2 \Lambda_s[m]}{(\partial \zeta_i)^2} = \frac{-4\pi^2}{N^2} (m-1)^2 \exp\{-2j\pi\zeta_i(m-1)/N\}, \text{ for } m = 1, \dots, N/2, \quad (\text{A.7.6})$$

and for the second half:

$$\frac{\partial \Lambda_s[m]}{\partial \zeta_i} = \frac{-2j\pi(N-m+1)}{N} \exp\{-2j\pi\zeta_i(N-m+1)/N\}, \text{ for } m = \frac{N}{2} + 1, \dots, N \quad (\text{A.7.7})$$

$$\frac{\partial^2 \Lambda_s[m]}{(\partial \zeta_i)^2} = \frac{-4\pi^2}{N^2} (N-m+1)^2 \exp\{-2j\pi\zeta_i(N-m+1)/N\}, \text{ for } m = \frac{N}{2} + 1, \dots, N. \quad (\text{A.5. 8})$$

The derivative of the terms in Eq. (A.7.2) is given by applying Eq. (A.7.5) - (A.7.8):

$$\frac{\partial J_{MAP}(\zeta_i)}{\partial \zeta_i} = \sum_{m=1}^N \mathbf{Z}_i^*[m] \frac{\partial \mathbf{T}_i[m]}{\partial \zeta_i} + \sum_{m=1}^N \mathbf{Z}_i[m] \frac{\partial \mathbf{T}_i^*[m]}{\partial \zeta_i} + \sum_{m=1}^N \mathbf{T}_i^*[m] \frac{\partial \mathbf{T}_i[m]}{\partial \zeta_i} + \sum_{m=1}^N \frac{\partial \mathbf{T}_i^*[m]}{\partial \zeta_i} \mathbf{T}_i[m].$$

From the definition of  $\mathbf{T}_i$ , it is:

$$\frac{\partial \mathbf{T}_i[m]}{\partial \zeta_i} = \sum_{n=0}^{d-1} \frac{\partial \Lambda_s\left[m + \frac{nN}{d}\right]}{\partial \zeta_i} \Lambda_H\left[m + \frac{nN}{d}\right] \mathbf{X}\left[m + \frac{nN}{d}\right],$$

and:

$$\frac{\partial \mathbf{T}_i^*[m]}{\partial \zeta_i} = \sum_{n=0}^{d-1} \frac{\partial \Lambda_s^* \left[ m + \frac{n}{d} N \right]}{\partial \zeta_i} \Lambda_H^* \left[ m + \frac{n}{d} N \right] \mathbf{X}^* \left[ m + \frac{n}{d} N \right].$$

Similarly, the second derivative is:

$$\begin{aligned} \frac{\partial^2 J_{MAP}(\zeta_i)}{\partial \zeta_i^2} &= \sum_{m=1}^N \mathbf{Z}_i^*[m] \frac{\partial^2 \mathbf{T}_i[m]}{\partial \zeta_i^2} + \sum_{m=1}^N \mathbf{Z}_i[m] \frac{\partial^2 \mathbf{T}_i^*[m]}{\partial \zeta_i^2} + \\ &+ \sum_{m=1}^N \mathbf{R}_i^*[m] \frac{\partial^2 \mathbf{T}_i[m]}{\partial \zeta_i^2} + 2 \sum_{m=1}^N \left| \frac{\partial \mathbf{T}_i[m]}{\partial \zeta_i} \right|^2 + \sum_{m=1}^N \frac{\partial^2 \mathbf{T}_i^*[m]}{\partial \zeta_i^2} \mathbf{T}_i[m]. \end{aligned}$$

To be precise, in our application we deal with 2-D signals where here are two translations parameters per image  $\vec{\zeta}_i = (\zeta_i^x, \zeta_i^y)$ . Thus, in Newton-Raphson update equation, Eq. (7.10)

$\frac{\partial J_{MAP}(\vec{\zeta}_i^n)}{\partial \vec{\zeta}_i}$  is a  $2 \times 1$  gradient vector and  $\left( \frac{\partial^2 J_{MAP}(\vec{\zeta}_i^n)}{\partial \vec{\zeta}_i^2} \right)$  a  $2 \times 2$  Hessian matrix involved.

However, the inversion of a  $2 \times 2$  matrix is easily found in closed form; hence the 2-D version of the registration algorithm is also very fast.



Figure 7.1.1: Low resolution degraded observation



(a)



(b)



(c)

Figure 7.1.2: (a) Stationary 2x [133], ( $MSE=195.2$ ), (b) Total Variation, Eq. 7.12 ( $MSE=182.1$ ), (c) Non-stationary MAP 2x super-resolved image ( $MSE=162.4$ ).

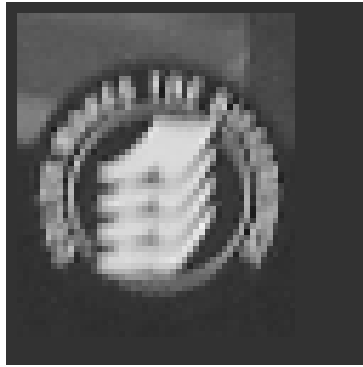


Figure 7.2.1: Low resolution observation.



(a)



(b)



(c)

Figure 7.2.2: Super-resolved images; (a) 2x stationary [133], (b) 2x Total Variation, Eq. (7.12), (c) 2x MAP non-stationary.

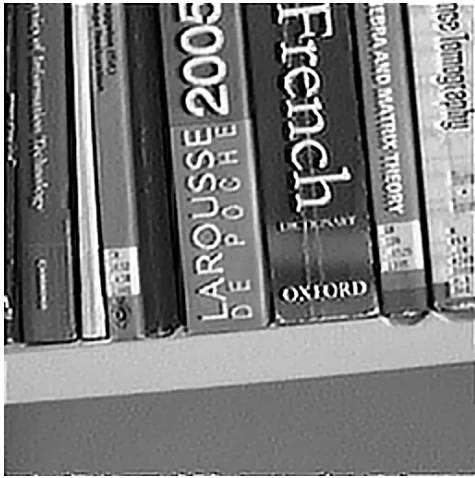




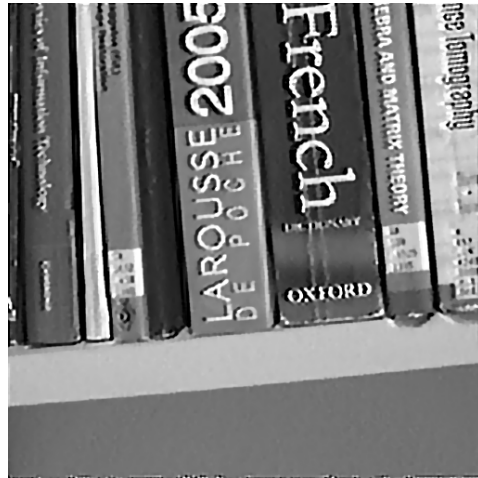
Figure 7.3.1: Low resolution observation.



(a)



(b)



[1] (c)

Figure 7.3.2: Super-Resolution experiment with real data; super-resolved images; (a) 4x stationary [133], (b) 4x Total Variation, Eq. (7.12), (c) 4x MAP non-stationary.

Table 7. 1 Original and estimated parameters for the artificially generated images.

	$\theta$	$\hat{\theta}$	$(\delta^x, \delta^y)$	$(\hat{\delta}^x, \hat{\delta}^y)$
Im 2	1.00	0.99	(0.30, -0.20)	(0.31, -0.21)
Im 3	2.00	2.00	(0.10, -0.30)	(0.14, -0.32)
Im 4	3.00	3.00	(-0.20, 0.10)	(-0.15, 0.06)
Im 5	1.50	1.50	(-0.15, 0.25)	(-0.13, 0.23)
Im 6	0.50	0.48	(0.00, 0.10)	(0.00, 0.09)
Im 7	-1.00	-1.00	(0.05, 0.12)	(0.04, 0.14)
Im 8	-2.00	-2.02	(0.14, 0.32)	(0.13, 0.35)

## CHAPTER 8. CONCLUSIONS AND FUTURE RESEARCH

---

In this dissertation we presented novel Bayesian algorithms for image restoration and super-resolution based on new edge-preserving image priors. More specifically, the Bayesian formulation of these problems enables the incorporation of hierarchical image priors with two levels, with the second level manifesting the spatial adaptive regularization mechanism. This mechanism leads to restored/reconstructed images for restoration and super-resolution, respectively, with preserved edges. Furthermore, working in a Bayesian framework, we have developed methodologies to infer the image and to estimate the model parameters, i.e. the noise variance and the prior parameters, in an iterative manner.

In Chapter 4 we introduced a spatially adaptive image prior by extending the SAR model used in previous restoration algorithms. The extension was performed by making the precisions of a Gaussian prior, imposed on the local directional differences, spatially varying. Then, a conjugate Gamma hyperprior was imposed on the precision parameters. Integration with respect to the precisions results in a Student's-t density. This density was also used as a prior in Chapter 5. However, the prior was imposed in the outputs of an arbitrary number of high-pass filters. Thus, this prior is in product form and has the ability to enforce simultaneously multiple constraints to the solution of the model. Lastly, in Chapter 6, the prior was generalized by extending this time the TV prior instead of the Gaussian. The extension was made again by assuming spatially varying regularization parameters and then a Gamma hyperprior was imposed on them.

Iterative algorithms were developed to learn the proposed models and to restore/reconstruct the image. First, in Chapter 4, the image is treated as a parameter and the precisions of the Gaussian prior as hidden variables. Then, both a MAP and a Bayesian algorithm were proposed to estimate the image and the precisions. Numerical experiments showed superiority of the MAP algorithm with respect to other state-of-the-art methodologies.

In Chapter 5, the image is treated as a hidden variable and a variational Bayesian algorithm is employed for inference, which means a posterior distribution is obtained for the image. In addition, the model parameters are estimated in a rigorous manner, in contrast to many popular methodologies for image restoration, in which they are selected empirically. Furthermore, apart from the first order differences, more high-pass operators are used. A Lanczos-based numerical method has also been developed to estimate the diagonal elements of the covariance matrix of the posterior distribution obtained for the image. Numerical experiments demonstrate that the proposed algorithm is superior to that of Chapter 4.

In most of the experiments, the proposed algorithms produced the best results in terms of *ISNR*, compared to state-of-the-art methods. However, are not superior in high levels of noise, but they exhibit the same performance. A possible explanation of this is that high levels of noise diminish the subtle features of the images and the proposed model cannot capture these features. In this situation, all methods become equal in performance.

In Chapter 7, the proposed image model was successfully incorporated in the super-resolution problem. The MAP framework algorithm was found to be superior of other state-of-the-art methodologies in terms both of MSE and visual quality. The high-resolution images have their edges-preserved, something which was the initial motivation for the incorporation of the proposed prior. As a result, in an illustrative experiment, letters that are illegible in the low-resolution images become legible in the high-resolution image.

Numerical methods to estimate the diagonal elements of an inverse matrix are presented in Chapters 5 and 6. These methods are faster than other methodologies used to accomplish this type of calculation. Also, it would be interesting to conduct detailed experiments in order to test the performance of this approximate estimation when applied to a known matrix. The method in Chapter 6 was proven to be faster than that of Chapter 5.

An interesting issue for future research is the relaxation of the independence assumption of the outputs of the filters. In most images the edges are formed in a continuous manner and this is an indication that the filter outputs are correlated. A possible methodology to introduce correlations to the image model is to assume, for example, a tridiagonal covariance matrix for the Gaussian prior and impose a Wishart distribution on it, instead of the Gamma. This type of hyperprior has already been used successfully in Machine Learning problems [36].

The incorporation of alternative and/or additional filters in the image prior is a topic for further investigation. The operators used in the image prior in Chapters 5 and 6 can be replaced by filters successfully employed in several problems, such as wavelets, filter banks, curvelets, or even filters learned from real images.

The super-resolution algorithm could also be extended so as to avoid the pre-processing step and in this way to avoid the sub-optimal pre-registration procedure. The registration should be ideally performed as a step in the MAP algorithm. Also a Bayesian algorithm can be employed for the registration that treats the registration parameters as random variables and impose a hyperprior for them.

An open issue for the super-resolution problem is the estimation of the blurring operator that is applied to the image before downsampling. Estimation of the point-spread-function (PSF) of the blurring operation can be achieved with methodologies borrowed from the blind image deconvolution research.

## BIBLIOGRAPHY

---

- [1] T. Akgun, Y. Altunbasak, R. M. Mersereau, “Super-resolution reconstruction of hyperspectral images,” *IEEE Trans. on Image Processing*, Vol. 14, No. 11, pp. 1860 – 1875, November 2005.
- [2] M. Allain, J. Idier and Y. Goussard, “On local and global convergence of half-quadratic algorithms,” *IEEE Transactions on Image Processing*, vol. 15, no. 5, pp. 1130-1142, May 2006.
- [3] S. Alliney, and S. A. Ruzinsky, “An algorithm for the minimization of mixed  $l_1$  and  $l_2$  norms with application to Bayesian estimation,” *IEEE Transactions on Signal Processing*, vol. 42, no. 3, pp. 618-627, 1994.
- [4] G. L. Anderson and A. N. Netravali, “Image restoration based on a subjective criterion”, *IEEE Trans. Syst., Man, Cybern*, vol. SMC-6, pp. 845-853, Dec. 1976.
- [5] H.C. Andrews and B.R. Hunt, *Digital Image Restoration*, Englewood Cliffs, NJ: Prentice Hall, 1997.
- [6] D. Andrews and C. Mallows. “Scale mixtures of normal distributions,” *Journal of the Royal Statistical Society (B)*, vol. 36, pp. 99-102, 1974.
- [7] G. Aubert, M. Barlaud, L. Blanc-Feraud, and P. Charbonnier, “Deterministic edge-preserving regularization in computed imaging,” Tech. Rep. 94-01, 13S, Univ. Nice, Sophia Antipolis, France, 1994.
- [8] S. D. Babacan, R. Molina, A.K Katsaggelos, “Parameter Estimation in TV Image Restoration Using Variational Distribution Approximation,” *IEEE Transactions on Image Processing*, vol. 17, no3, pp. 326-339, 2008.
- [9] M. Banham and A. Katsaggelos, “Spatially adaptive wavelet-based multiscale image restoration,” *IEEE Transactions on Image Processing*, vol. 5, pp. 619–634, 1996.
- [10] Beal, M.J., “Variational Algorithms for Approximate Bayesian Inference”, PhD. Thesis, Gatsby Computational Neuroscience Unit, University College London, 2003.

- [11] M. Belge, M. E. Kilmer, and E. L. Miller, “Wavelet domain image restoration with adaptive edge-preserving regularity,” *IEEE Transactions on Image Processing*, vol. 9, pp. 597–608, 2000.
- [12] J. Berger, *Statistical Decision Theory and Bayesian Analysis*, Springer 1993.
- [13] J. M. Bernardo, A. F. M. Smith, *Bayesian Theory*, John Wiley 2000.
- [14] M. Bertalmio, V. Caselles, B. Rouge, and A. Sole, “TV based image restoration with local constraints,” *J. Sci. Comput.*, vol. 19, no. 1-3, pp. 95-122, Dec. 2003.
- [15] J. Besag, “Spatial interaction and the statistical analysis of lattice systems (with discussion)”, *J. Royal Stat. Soc.*, series B, vol. 34, pp. 75-83, 1972.
- [16] J. Bioucas-Dias, “Bayesian wavelet-based image deconvolution: a GEM algorithm exploiting a class of heavy-tailed priors,” *IEEE Transactions on Signal Processing*, vol. 15, pp. 937–951, 2006.
- [17] J. Bioucas-Dias, M. Figueiredo, and J. Oliveira, “Adaptive Bayesian/total-variation image deconvolution: A majorization-minimization approach,” in *EUSIPCO’ 2006*, Florence, Italy, 2006.
- [18] J. Bioucas-Dias, M. Figueiredo, and J. Oliveira, Total-variation image deconvolution: A majorization-minimization approach”, in *ICASSP’2006*, 2006.
- [19] C. Bishop, *Pattern Recognition and Machine Learning*, Springer Verlag, 2006.
- [20] C. A. Bouman, “Markov Random Fields and stochastic image models”, *IEEE Transactions on Image Processing*, Tutorial, Washington, October 1995.
- [21] C. Bouman and K. Sauer, “A Generalized Gaussian Image Model for Edge Preserving MAP Estimation”, *IEEE Trans. on Image Processing*, Vol. 2, No. 3, pp. 296-310, July 1993.
- [22] F. Champagnat and J. Idier, “A connection between half-quadratic criteria and EM algorithms”, *IEEE Transactions on Image Processing Letters*, vol. 11, no. 9, pp. 709-712, September 2004.
- [23] T. Chan, S. Esedoglu, F. Park, and A. Yip, *Recent developments in total variation image restoration*, in the *Handbook of Mathematical Models in Computer Vision*, (N. Paragios, Y. Chen, and O. Faugeras, eds.), Springer Verlag, 2005.

- [24] G. Chantas, N. P. Galatsanos, A. Likas, and M. Saunders, "Variational Bayesian Image Restoration Based on a Product of  $t$ -Distributions Image Prior", *IEEE Transactions on Image Processing*, vol. 17, no. 10, pp. 1795-1805, October 2008.
- [25] G. Chantas, N. P. Galatsanos, and A. Likas, "Bayesian Image Restoration Based on Variational Inference and a Product of Student-t Priors", *Proceedings of the IEEE International MLSP Workshop*, Thessaloniki, Greece, August 2007.
- [26] G. Chantas, N. Galatsanos, and A. Likas, "Bayesian restoration using a new non-stationary edge preserving image prior," *IEEE Transactions on Image Processing*, vol. 15, No. 10, pp: 2987-2997, October 2006.
- [27] G. Chantas, N. Galatsanos and A. Likas, "Maximum a posteriori image restoration based on a new directional continuous edge image Prior," in *Proc. IEEE ICIP 2005*, Genoa, Italy, vol. 1, pp. 941-944, September 11-14, 2005.
- [28] G. Chantas, N. Galatsanos, A. Likas, "Non Stationary Bayesian Image Restoration", *Proceedings of the IEEE International Conference on Pattern Recognition*, vol. 4, pp. 689-692, August 2004.
- [29] G. Chantas, N. P. Galatsanos, and N. Woods, A Super-Resolution Based on Fast Registration and Maximum a Posteriori Reconstruction, *IEEE Trans. on Image Processing*, vol. 16, no. 7, pp. 1821-1830, July 2007.
- [30] G. Chantas, N. P. Galatsanos, and N. Woods, "Maximum a Posteriori Super Resolution based on Simultaneous Non-Stationary Restoration, Interpolation and Fast Registration", *Proceedings of the European Signal Processing Conference, EUSIPCO'06*, Florence, September 2006.
- [31] P. Charbonnier, L. Blanc-Feraud, G. Aubert, and Michel Barlaud, "Deterministic Edge-Preserving Regularization in Computed Imaging", *IEEE Transactions on Image Processing*, vol. 6, pp. 298-311, February 1997.
- [32] S. Chaudhuri, Editor, *Super-Resolution Imaging*, Chapter 4, "Reconstruction of a high resolution image from multiple low resolution images", by B. Tom, N. Galatsanos and A. Katsaggelos, pp. 71-105, Kluwer, 2001.



- [33] C. Chaux, P. Combettes, J.-C. Pesquet, and V. Wajs. “Iterative image deconvolution using overcomplete representations,” *European Signal Processing Conference – EUSIPCO 2006*, Florence, Italy, 2006.
- [34] R. Chellapa, T. Simchony, Z. Lichtenstein, Image estimation using 2D noncausal Gauss-Markov random field models, in: Katsaggelos, A.K. (Ed.), *Digital Image Restoration*, Springer Series in Information Science, vol. 23, Springer, Berlin, 1991.
- [35] S. Chen, D. Donoho, and M. Saunders. “Atomic decomposition by basis pursuit,” *SIAM Journal of Scientific Computation*, vol. 20, pp. 33–61, 1998.
- [36] C. Constantinopoulos, *Statistical and Neural Methods for Machine Learning Problems*, PhD Thesis, University of Ioannina, Department of Computer Science, 2006.
- [37] A. L. Cunha, J. Zhou, and M. N. Do, “The nonsubsampling contourlet transform: Theory, design, and applications”, *IEEE Transactions on Image Processing*, vol. 15, no. 10, pp. 3089-3101, Oct. 2006.
- [38] I. Daubechies, M. Defriese, and C. De Mol, “An iterative thresholding algorithm for linear inverse problems with a sparsity constraint”, *Comm. Pure and Applied Math.*, vol. LVII, pp. 1413-1457, 2004.
- [39] A. H. Delaney and Y. Bresler, “Globally Convergent Edge-Preserving Regularized Reconstruction: An Application to Limited-Angle Tomography,” *IEEE Transactions On Image Processing*, Vol. 7, No. 2, February 1998.
- [40] G. Demoment, “Image restoration and reconstruction: overview of common estimation structures and problems,” *IEEE Trans. on Acoustics Speech and Signal Processing*, vol. 37, pp. 2024-2036, December 1989.
- [41] A. D. Dempster, N. M. Laird, and D. B. Rubin, “Maximum Likelihood from Incomplete Data via the EM algorithm”, *J. Roy. Stat. Soc.*, vol. B39, pp. 1-37, 1977.
- [42] M. N. Do and M. Vetterli, “Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance,” *IEEE Trans. on Image Processing*, vol. 11, no. 2, pp. 146–158, Feb 2002.
- [43] D. Donoho, “Nonlinear solution of linear inverse problems by wavelet-vaguelette decompositions,” *J. Appl. Comput. Harmon. Anal.*, vol. 1, pp. 100–115, 1995.

- [44] M. Elad, "Why simple shrinkage is still relevant for redundant representations?" *IEEE Transactions on Information Theory*, vol. 52, pp. 5559-5569, 2006.
- [45] M. Elad, B. Matalon, and M. Zibulevsky, "Image denoising with shrinkage and redundant representations", *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition – CVPR'2006*, New York, 2006.
- [46] L. Evans and R. Gariepy, *Measure Theory and Fine Properties of Functions*, CRC Press, 1992.
- [47] S. Farsiu, M. Elad, P. Milanfar, "Video-to-Video Dynamic Superresolution for Grayscale and Color Sequences", *EURASIP Journal of Applied Signal Processing, Special Issue on Super-resolution Imaging*, Volume 2006, Article ID 61859, January 2006.
- [48] S. Farsiu, M. Elad, P. Milanfar, "Multiframe demosaicing and super-resolution of color images", *IEEE Transactions on Image Processing*, Vol.15, Issue 1, pp. 141-159, January 2006.
- [49] S. Farsiu, M. D. Robinson, M. Elad, P. Milanfar, "Fast and robust multi-frame super resolution," *IEEE Trans. on Image Processing*, Vol. 13, No. 10, pp. 1327 – 1344, October 2004.
- [50] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, "Advances and challenges in super-resolution," *International Journal of Imaging Systems and Technology*, vol. 14, no. 2, pp. 47-57, August 2004.
- [51] S. Farsiu, D. Robinson, M. Elad, P. Milanfar, "Robust Shift and Add Approach to Super-Resolution", *International Symposium on Optical Science and Technology, SPIE's 48th Annual Meeting, August 2003*, San Diego, CA, Volume 5203, Pages: 121-130.
- [52] M. Figueiredo, "Adaptive sparseness for supervised learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence - PAMI*, vol. 25, no. 9 pp.1150-1159, September 2003.
- [53] M. Figueiredo, J. Bioucas-Dias, R. Nowak, "Majorization-minimization algorithms for wavelet-based image restoration", *IEEE Transactions on Image Processing*, vol. 16, no. 12, pp. 2980-2991, 2007.

- [54] M. Figueiredo and R. Nowak, “Wavelet-based adaptive image deconvolution,” *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing – ICASSP’2002*, Orlando, FL, 2002.
- [55] M. Figueiredo and R. Nowak. “An EM algorithm for wavelet-based image restoration.” *IEEE Transactions on Image Processing*, vol. 12, pp. 906–916, 2003.
- [56] W. T. Freeman, T. R. Jones, and E. C. Pasztor, “Example-based super-resolution”, *IEEE Computer Graphics and Applications*, , Vol. 22, No. 2, pp. 56 – 65, March-April 2002.
- [57] H. Fu, M. Ng, M. Nikolova, and J. Barlow, “Efficient minimization methods of mixed  $l^1$  -  $l^1$  and  $l^1$  -  $l^2$  norms for image restoration,” *SIAM Journal of Scientific Computing*, Volume 27, Issue 6, pp. 1881-1902, 2006.
- [58] N. P. Galatsanos, V. N. Mesarovic, R. M. Molina and A. K. Katsaggelos, “Hierarchical Bayesian Image Restoration from Partially-Known Blurs,” *IEEE Trans. on Image Processing*, Vol. 9, No. 10, pp. 1784-1797, October 2000.
- [59] N. P. Galatsanos and A. K. Katsaggelos, “Methods for Choosing the Regularization Parameter and Estimating the Noise Variance in Image Restoration”, *IEEE Trans. Image Proc.*, vol 1., pp. 322-336, July 1992.
- [60] N. Galatsanos, V. N. Mesarovic, R. M. Molina, J. Mateos, and A. K. Katsaggelos, “Hyper-parameter Estimation Using Gamma Hyper-priors in Image Restoration from Partially-Known Blurs,” *Optical Engineering*, 41(8), pp. 1845-1854, August 2002.
- [61] S. Geman and D. Geman, “Stochastic Relaxation, Gibbs distribution, and the Bayesian Restoration of Images”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, pp. 228-238, November 1984.
- [62] D. Geman and C. Yang, “Non-linear Image Recovery with Half-Quadratic Regularization and FFT’s”, *IEEE Transactions on Image Processing*, vol.4, pp. 932-946, July 1995.
- [63] G. H. Golub and C. F. Van Loan. *Matrix Computations*. The Johns Hopkins University Press, Baltimore, third edition, 1996.

- [64] J. Guerrero-Colon, J. Portilla, “Deblurring-by-denoising using spatially adaptive Gaussian scale mixtures in overcomplete pyramids.” Proceedings of the *IEEE International Conference on Image Processing – ICIP’2006*, Atlanta, GA, USA, 2006.
- [65] H. He, L. P. Kondi, “An image super-resolution algorithm for different error levels per frame”, *IEEE Transactions on Image Processing*, Vol. 15, Issue 3, pp. 592-603, March 2006.
- [66] K. Hirakawa and X.-L. Meng, “An Empirical Bayes EM-Wavelet Unification For Simultaneous Denoising, Interpolation, And/Or Demosaicing”, in *Proceedings IEEE International Conference on Image Processing*, ICIP 2006, Atlanta Georgia, September 2006.
- [67] T.S. Huang, R.Y. Tsai, “Multi-frame image restoration and registration,” *Adv Comput Vision Image Process*, 1: 317–339, 1984.
- [68] B. R. Hunt, “The Application of Constrained Least Squares to Image Restoration by Digital Computer”, *IEEE Trans. Computers*, vol. 22, pp. 805-812, September 1973.
- [69] J. Idier, “Convex Half-Quadratic Criteria and Interacting Auxiliary Variables for Image Restoration”, *IEEE Transactions on Image Processing*, vol. 10, pp. 1001-1009, no. 7, July 2001.
- [70] A. K. Jain, *Fundamentals of Digital Image Processing*, Prentice Hall, 1988.
- [71] A. Jalobeanu, N. Kingsbury, and J. Zerubia, “Image deconvolution using hidden Markov tree modeling of complex wavelet packets,” in *Proc. IEEE Int. Conf. Image Proc.—ICIP’2001*, Thessaloniki, Greece, 2001.
- [72] F. C. Jeng, Compound Gauss-Markov random fields for image estimation and restoration, Ph.D. Thesis, Rensselaer, Polytechnic Institute, 1988.
- [73] F. C. Jeng, J. W. Woods, Compound Gauss-Markov models for image processing, in: Katsaggelos, A.K. (Ed.), *Digital Image Restoration*, Springer Series in Information Science, vol. 23, Springer, Berlin, 1991.
- [74] F. C. Jeng and J. W. Woods, Simulated annealing in compound Gaussian random fields, *IEEE Trans. Inform. Theory* 36, (1988), 94-107.
- [75] M. V. Joshi, S. Chaudhuri and R. Panuganti, “A learning-based method for image super-resolution from zoomed observations,” *IEEE Trans. Systems, Man and Cybernetics*, Part B, Vol. 35, No. 3, pp. 527 – 537, June 2005.

- [76] J. Kalifa, S. Mallat, and B. Rougé, “Deconvolution by thresholding in mirror wavelet bases,” *IEEE Transactions on Image Processing*, to be published.
- [77] A. Kanemura, S.-i. Maeda and S. Ishii, “Hyperparameter Estimation in Bayesian Image Superresolution with a Compound Markov Random Field Prior”, in *Proceedings of IEEE Int. Workshop on Machine Learning for Signal Processing*, pp. 181-186, Thessaloniki, Greece, August 2007.
- [78] A. K. Katsaggelos “Iterative Image Restoration”, in *Handbook on Image and Video Processing*, Editor A.I. Bovik, pp. 191-206, Academic Press 2000.
- [79] A. K. Katsaggelos, *Constrained Iterative Image Restoration Algorithms*, Ph.D. Thesis, Georgia Institute of Technology, August 1985.
- [80] A. Katsaggelos, J. Biemond, R. M. Mersereau, and R. W. Schafer, “Non Stationary Iterative Image Restoration”, *IEEE Proceedings of International Conference on Acoustics Speech and Signal Processing (ICASSP)*, Tampa FL, pp. 696-699, 1985.
- [81] A. K. Katsaggelos, J. Biemond, R. W. Schafer, and A. R. Mersereau, “A regularized Iterative Image Restoration Algorithm”, *IEEE Trans. Acoust., Speech, Signal Proc.*, vol. 39, pp. 914-929, April 1991.
- [82] A. K. Katsaggelos and M. G. Kang, “Spatially adaptive iterative algorithm for the restoration of astronomical images,” *International Journal of Imaging Systems and Technology*, vol. 6, no. 4, pp. 305-313, 1995.
- [83] J. C. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright, “Convergence properties of the Nelder-Mead simplex method in low dimensions,” *SIAM Journal of Optimization*, Vol. 9, No. 1, pp. 112-147, 1998.
- [84] R.L. Lagendijk, J. Biemond and D.E. Boeke “Regularized Iterative Image Restoration with Ringing Reduction”, *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 36, pp. 1874-1887, December 1988.
- [85] K. Lange, *Optimization*, Springer Texts in Statistics, Springer-Verlag, 2004.
- [86] S. Z. Li, *MRF modelling in Computer Vision*, Springer-Verlag 2005.
- [87] Likas and N. Galatsanos, “A Variational Approach For Bayesian Blind Image Deconvolution”, *IEEE Transactions on Signal Processing*, Volume: 52, Issue: 8, pp: 2222 – 2233, August 2004.

- [88] J. Liu and P. Moulin. "Complexity-Regularized Image Restoration," *Proceedings of the IEEE International Conference on Image Processing – ICIP'1998*, vol. 1, pp. 555–559, Chicago, IL, USA, 1998.
- [89] C. Liu and D. B. Rubin, "ML estimation of the t distribution using EM and its extensions", *ECM and ECME, Statistica Sinica*, 5, 19-39, 1995.
- [90] D. J. MacKay, *Information Theory, Inference and Learning Algorithms*, Cambridge University Press, 2003.
- [91] A. Mairgiotis, N. Galatsanos and Y. Yang, "New Detectors for Watermarks with Unknown Power Based on Student-t Image Priors", *Proc. IEEE International Conference on Multimedia Signal Processing, MMSP 2007*, Chania, Crete 2007.
- [92] R. Molina, "On the Hierarchical Bayesian Approach to Image Restoration. Applications to Astronomical Images", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 11, 1122-1128, November 1994.
- [93] R. Molina, Aggelos K. Katsaggelos, J. Mateos, A. Hermoso, and C. A. Segall, "Restoration of severely blurred high range images using stochastic and deterministic relaxation algorithms in compound Gauss-Markov random fields," *Pattern Recognition*, vol. 33, no. 4, pp. 555-571, 2000.
- [94] R. Molina, A. K. Katsaggelos, J. Mateos, "Bayesian and regularization methods for hyper-parameter estimation in image restoration," *IEEE Trans. on Image Processing*, Vol: 8 No: 2, pp. 231 -246, Feb. 1999.
- [95] R. Molina, and B. D. Ripley, "Using spatial models as priors in astronomical images analysis", *J. Appl. Stat.*, vol.16, pp.193-206, 1989.
- [96] P. Moulin and J. Liu. "Analysis of multiresolution image denoising schemes using generalized-Gaussian and complexity priors," *IEEE Transactions on Information Theory*, vol. 45, pp. 909–919, 1999.
- [97] S. G. Nash and A. Sofer, *Linear and Nonlinear Programming*, Mc Graw Hill, 1996.
- [98] M. Z. Nashed, "Operator Theoretic and Computational Approaches to Ill-Posed Problems with Applications to Antenna Theory," *IEEE Trans. Antennas Prop.*, vol. 29, pp. 220-231, March 1981.

- [99] R. M. Neal, and G. E. Hinton, "A View of the E-M Algorithm that Justifies Incremental, Sparse and Other Variants", In M. I. Jordan (ed.) *Learning in Graphical Models*, pp. 355-368. Cambridge, MA: MIT Press 1998.
- [100] R. Neelamani, H. Choi, and R. Baraniuk, "ForWaRD: Fourier-wavelet regularized deconvolution for ill-Conditioned systems," *IEEE Transactions on Signal Processing*, Vol. 52, No. 2, pp. 418-433, February 2004.
- [101] C. Nikias and M. Shao, *Signal Processing with Alpha-Stable Distributions and Applications*. Wiley, 1995.
- [102] S. Osher, L. Rudin and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D.*, vol. 60, pp. 259-268, 1992.
- [103] S. Osher, A. Sole, and L. Vese, "Image decomposition and restoration using total variation minimization and the  $h^1$  norm," *SIAM Multiscale Modeling and Simulation*, vol. 1, no. 2, pp. 349-370, 2003.
- [104] R. Pan, and S. Reeves, "Efficient Huber-Markov Edge Preserving Image Restoration", *IEEE Transactions on Image Processing*, Vol. 15, No. 12, pp. 3728-3735, Dec. 2006.
- [105] S. Park, M. Park, and M. Kang, "Super-resolution image reconstruction: a technical overview," *IEEE Signal Processing Magazine*, vol. 20, no. 3, pp 21-36, May 2003.
- [106] C. C. Paige and M. A. Saunders, "LSQR: An algorithm for sparse linear equations and sparse least squares", *ACM Trans. Math. Software*, Vol. 8, no. 1, pp. 43-71, 1982.
- [107] C. C. Paige and M. A. Saunders, "Solution of sparse indefinite systems of linear equations", *SIAM Journal on Numerical Analysis*, Vol. 12, pp. 617-629, 1975.
- [108] L. C. Pickup, D. P. Capel, S. J. Roberts, and A. Zisserman, "Bayesian image super-resolution, continued," in *Advances in Neural Information Processing Systems 19*, pp. 1089-1096, Cambridge, Mass, USA, December 2006.
- [109] I. Pollak, A. Willsky, and Y. Huang, "Nonlinear evolution equations as fast and exact solvers of estimation problems," *IEEE Transactions on Signal Processing*, vol. 53, no. 2, pp. 484-498, 2005.
- [110] J. Portilla, V. Strela, M. Wainwright, E. Simoncelli. "Image denoising using scale mixtures of Gaussians in the wavelet domain." *IEEE Transactions on Image Processing*, vol. 12, pp. 1338-1351, 2003.

- [111]A. Rangarajan and R. Chellappa, *Markov random field models in image processing*, Arbib M. (Editor), in *The Handbook of Brain Theory and Neural Networks*, MIT Press, 1995.
- [112]P. de Rivaz and N. Kingsbury, "Bayesian image deconvolution and denoising using complex wavelets," in *Proc. IEEE Int. Conf. Image Proc, ICIP'2001*, Thessaloniki, Greece, 2001.
- [113]D. Robinson, P. Milanfar, "Statistical performance analysis of super-resolution", *IEEE Trans. on Image Processing*, Vol. 15, Issue 6, pp. 1413-1428, June 2006.
- [114]S. Roth, M. J. Black, "Fields of Experts: A Framework for Learning Image Priors", *IEEE Conf. on Computer Vision and Pattern Recognition*, vol. II, pp. 860-867, June 2005.
- [115]J. Ruanaidh, and W. Fitzgerald, *Numerical Bayesian Methods Applied to Signal Processing*, Springer Verlag, 1996.
- [116]Y. Saad, *Iterative Methods for Sparse Linear Systems*, Second Edition, Society for Industrial and Applied Mathematics, 2000.
- [117]R. R. Schultz and R. L. Stevenson, "A Bayesian Approach to Image Expansion with Improved Resolution", *IEEE Trans. on Image Processing*, Vol. 3, No. 5, pp. 233-242, May 1994.
- [118]C. A. Segall, A. K. Katsaggelos, R. Molina, J. Mateos, "Bayesian resolution enhancement of compressed video", *IEEE Trans. on Image Processing*, Vol. 13, Issue 7, pp. 898-911, July 2004.
- [119]L. Sendur and I. Selesnick. "Bivariate shrinkage with local variance estimation," *IEEE Signal Processing Letters*, vol. 9, pp. 438-441, 2002.
- [120]H. R. Sheikh and A. C. Bovik, "Image information and visual quality", *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 430-444, February 2006.
- [121]E. Simoncelli, E. Adelson, "Noise removal via Bayesian wavelet coring", *Proceedings of the IEEE International Conference on Image Processing – ICIP'96*, vol. I, pp. 379-382, Lausanne, Switzerland, 1996.
- [122]D. Sun, and W.-K. Cham, "Postprocessing of Low Bit-Rate Block DCT Coded Images Based on a Fields of Experts Prior", *IEEE Trans. on Image Processing*, Vol. 16, No. 11, pp. 2743-2751, November 2007.



- [123]J. L. Starck, E. Candes, and D. Donoho. “Astronomical image representation by the curvelet transform”, *Astronomy and Astrophysics*, vol. 398, pp. 785–800, 2003.
- [124]H. Stark and Y. Yang, *Vector Space Projections Methods a Numerical Approach to Signal and Image Processing Neural Networks and Optics*, John Wiley 1998.
- [125]A. M. Tekalp, H. Kaufman, and J.W. Woods, “Edge Adaptive Image Filtering for Image Restoration with Ringing Suppression”, *IEEE Trans. Acoust., Speech, and Signal Proc.*, vol. 37, no. 6, pp. 892-899, June 1989.
- [126]R. Tibshirani “Regression shrinkage and selection via the LASSO”, *Journal of the Royal Statistical Society (B)*, vol. 58, no. 1, pp. 267–288, 1996.
- [127]A. N. Tikhonov and V.Y Arsenin, *Solutions of Ill-Posed Problems*, New York: Wiley, 1977.
- [128]M. E. Tipping “Sparse Bayesian Learning and the Relevance Vector Machine” *Journal of Machine Learning Research* 1, 211-244, 2001.
- [129]D. Tzikas, A. Likas, and N. Galatsanos, “Life After the EM Algorithm: The Variational Approximation for Bayesian Inference”, *IEEE Signal Processing Magazine*, to appear.
- [130]D. Tzikas, A. Likas and N. Galatsanos, “Variational Bayesian Blind Image Deconvolution with Student-t Priors”, in *Proceedings of IEEE Int. Conference on Image Processing*, S. Antonio, TX, September 2007.
- [131]A. Wilsky, “Multiresolution Markov Models for Signal and Image Processing”, *Proceedings of the IEEE*, Vol. 90, No. 8, August 2002, pp. 1397-1458.
- [132]D. Wipf and B. Rao, “Sparse Bayesian Learning for Basis Selection”, *IEEE Trans. on Signal Processing*, Vol. 52, No. 8, pp. 2153-2164, August 2004.
- [133]N. A. Woods, N.P. Galatsanos and A.K. Katsaggeloss, “Stochastic methods for joint restoration, interpolation and registration of multiple under sampled images,” *IEEE Transactions on Image Processing*, vol. 15, no. 1, pp. 201-213, January 2006.
- [134]C. R. Vogel and M. E. Oman, “Fast, robust total variation-based reconstruction of noisy, blurred images,” *IEEE Transactions on Image Processing*, vol. 7, pp. 813–824, June 1998.

[135]L. Yaroslavsky, "Boundary effect free and adaptive discrete signal sinc-interpolation algorithms for signal and image resampling ", *Applied Optics - IP*, Vol. 42, Issue 20, pp. 4166-4175, July 2003.

## PUBLISHED WORK

---

### Journal publications:

- G. Chantas, N. P. Galatsanos, A. Likas, and M. Saunders, "Variational Bayesian Image Restoration Based on a Product of  $t$ -Distributions Image Prior", *IEEE Transactions on Image Processing*, vol. 17, no. 10, pp. 1795-1805, October 2008.
- G. Chantas, N. P. Galatsanos, and N. Woods, "Super-Resolution Based on Fast Registration and Maximum a Posteriori Reconstruction," *IEEE Trans. on Image Processing*, vol. 16, no. 7, pp. 1821-1830, July 2007.
- G. Chantas, N. Galatsanos and A. Likas, "Maximum a posteriori image restoration based on a new directional continuous edge image Prior," in *Proc. IEEE ICIP 2005*, Genoa, Italy, vol. 1, pp. 941-944, September 11-14, 2005.

### Conference publications:

- G. Chantas, N. P. Galatsanos, and A. Likas, "Bayesian Image Restoration Based on Variational Inference and a Product of Student-t Priors", *Proceedings of the IEEE International MLSP Workshop*, Thessaloniki, Greece, August 2007.
- G. Chantas, N. P. Galatsanos, and N. Woods, "Maximum a Posteriori Super Resolution based on Simultaneous Non-Stationary Restoration, Interpolation and Fast Registration", *Proceedings of the European Signal Processing Conference, EUSIPCO'06*, Florence, September 2006.
- G. Chantas, N. Galatsanos and A. Likas, "Maximum a posteriori image restoration based on a new directional continuous edge image Prior," in *Proc. IEEE ICIP 2005*, Genoa, Italy, vol. 1, pp. 941-944, September 11-14, 2005.
- Chantas, G., Galatsanos, N., Likas, A., "Non Stationary Bayesian Image Restoration", *Proceedings of the IEEE International Conference on Pattern Recognition*, vol. 4, pp. 689-692, August 2004.

- A. Mairgiotis, G. Chantas, N. Galatsanos, K. Blekas, and Yongyi Yang, “New Detectors for Watermarks with Unknown Power Based on Student-t Image Priors,” in *Proc. IEEE 9th Workshop on Multimedia Signal Processing, 2007, MMSP 2007*, pp.353-356, October 2007.

## **SHORT CURRICULUM VITAE**

---

Ioannis Chantas received the bachelor degree and the MSc. Degree from the Department of Computer Science of University of Ioannina in 2002 and 2004 respectively. Since 2004 he has been a PhD student in the same department.

His research interests are in the area of statistical image processing, concentrating on statistical image modeling for solution of inverse problems, such as image restoration and super-resolution problems. Also, his research interests include machine-learning, neural networks and computer-vision.