

UNIVERSITY OF CALIFORNIA
Los Angeles

**Applications and Generalizations
of the Iterative Refinement Method**

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Mathematics

by

Lin He

2006

© Copyright by

Lin He

2006

The dissertation of Lin He is approved.

Luminita Vese

Stefano Soatto

Andrea Bertozzi

Stanley Osher, Committee Chair

University of California, Los Angeles

2006

*To my parents . . .
who—through the years—
have kept confidence in me
and have supported me in every way*

*To my fiancé . . .
whose company makes
this journey quite fascinating*

TABLE OF CONTENTS

1	Introduction	1
1.1	Some Total Variational Based Image Restoration Models	2
1.1.1	Rudin, Osher and Fatemi (ROF) Total Variation Minimization	2
1.1.2	Meyer's Analysis on the ROF model	4
1.1.3	Total Variational Models with Non-quadratic Fidelity	6
1.2	From TV Regularization to MR Image Reconstruction	7
1.3	The iterative Refinement Method	8
1.3.1	The Bregman Distance	8
1.3.2	The iterative Refinement Procedure	8
2	Iterative TV Regularization with Non-Quadratic Fidelity	11
2.1	Analysis of a Modified Iterative Regularization Procedure	12
2.1.1	Generalization of the Fidelity Term: A Modified Model	13
2.1.2	Well-Definedness of the New Iteration Model	15
2.1.3	Convergence Analysis	17
2.2	Numerical Results	22
2.2.1	L^2 Fidelity Model	23
2.2.2	L^1 Fidelity Model	24
3	Blind Deconvolution Using TV Regularization and Bregman Iteration	32

3.1	Blind Deconvolution Models	33
3.1.1	Existence of Minimizers	33
3.1.2	Analysis of the H^1 Regularization Blind Deconvolution Model	34
3.1.3	Analysis of TV Regularization Blind Deconvolution Model	36
3.2	New Model and Numerical Algorithm	37
3.2.1	New Model	37
3.2.2	Numerical Algorithm	38
3.2.3	Application of Bregman Distance	39
3.3	Explicit Numerical Scheme for the 2D Model	40
3.4	Numerical Results	43
4	MR Image Reconstruction from Sparse Radial Samples Using Bregman Iteration and Nonlinear Inverse Scale Space Methods	52
4.1	Prerequisite to Solve the Model (4.1)	54
4.1.1	The TV regularization	54
4.1.2	The Wavelet Transform	55
4.1.3	The Constraint Condition	57
4.2	The Sparse Representation for a Piecewise Smooth image	58
4.2.1	Wavelet + TV	58
4.2.2	Curvelet + TV	59
4.3	The Iterative Refinement Method and the Inverse Scale Space Method	60
4.3.1	The Bregman iteration	60
4.3.2	Nonlinear Inverse Scale Space Methods	61

4.4	Experimental Results	62
5	Conclusion and Future Work	70
	References	71

LIST OF FIGURES

2.1	Satellite reconstructions with the new iterative total variation method, Gaussian noise $\delta = 20$, $\lambda = 10$	26
2.2	Satellite reconstructions with the original iterative total variation method, Gaussian noise $\delta = 20$, $\lambda = 10$	27
2.3	Satellite reconstructions with the $BV + L^2$ T-N-V method, Gaus- sian noise $\delta = 20$, $\lambda = 10$	28
2.4	Restorations of a finger print obtained with the new iterative TV model, Gaussian noise, $\delta = 10$, $\lambda = 1$	29
2.5	Restorations of a finger print obtained with the original iterative TV model, Gaussian noise, $\delta = 10$, $\lambda = 1$	30
2.6	Restorations of a finger print obtained with the $BV + L^1$ T-N-V model, Gaussian noise, $\delta = 10$, $\lambda = 1$	31
3.1	Example of numerical algorithm procedure, $\text{SNR} = 1.47$. We ob- serve that the recovered kernel does improve after every round, and the Bregman iteration gives us more details of the image. However, the Bregman iteration brings back some noise in the first round.	45
3.2	Compare different α_1, α_2 . Notice that (d)~(f) with fewer details is smoother than (g)~(i).	46
3.3	Noise free case. from (b) to (e), we see that more details are restored.	48
3.4	Comparison 1: We show that a better initial guess for the kernel does not improve the result.	49

3.5	Comparison 2: We use the H^1 norm of the kernel. The result is comparable to that using the TV norm of kernel.	50
3.6	Comparison 3.	51
4.1	Conventional gridding algorithm.	63
4.2	A sequence images obtained from the iterative refinement method : top left, 1st Bregman iteration; top right, 2nd Bregman iteration; bottom left, 4th Bregman iteration; bottom right, 6th Bregman iteration.	64
4.3	A sequence images obtained from the relaxed inverse scale space flow : top left, iteration 100; top right, iteration 200; bottom left, iteration 400; bottom right, iteration 600.	66
4.4	Conventional gridding algorithm.	67
4.5	A sequence images obtained from the iterative refinement method with (4.9) where $\mu = 1.0$, $\nu = 0.1$, and $\lambda = 100$: top left, 1st Bregman iteration; top right, 2nd Bregman iteration; bottom left, 3rd Bregman iteration; bottom right, 4th Bregman iteration. . . .	68
4.6	A sequence images obtained from the relaxed inverse scale space flow with (4.9) where $\mu = 1.0$, $\nu = 0.1$, and $\lambda = 100$: top left, iteration 100; top right, iteration 200; bottom left, iteration 300; bottom right, iteration 400.	69

ACKNOWLEDGMENTS

I would first like to thank my adviser, Professor Stanley Osher, for his advice and guidance throughout my graduate studies. His encouragement, his enthusiasm and his patience sustains me to finish all my projects, motivates me to learn many different interesting topics and makes me feel grateful. I will never forget one day I came to his IPAM office during my first project, I was frustrated at my no-progress research work. But he said to me, “Don’t worry about it! It’s only research, it’s not life.”

Second I would like to express my gratitude to Professor Luminita Vese and Professor Andrea Bertozzi for their guidance and encouragement through these years and all of their insightful seminars I have attended.

I would like to extend my thanks to Professor Stefano Soatto for giving his time to be my committee member.

I would like to acknowledge all the co-authors I have worked with : Professor Antonio Marquina, Professor Martin Burger, Professor Chiu-yen Kao, Dr. Peter Speier and two colleagues Dr. Ti-Chiun Chang and Dr. Tong Fang from Siemens Corporate Research (SCR) in Princeton where I worked as an intern in 2005 Summer. I thank SCR for such a great opportunity which comprises an important part of my dissertation.

Now for my friends in IPAM 1129C : Jin-Jun Xu, Jason Chung, Tim Leung and Igor Yanovsky. I thank them for all the helpful and insightful discussions during lunch and tea time. That would lead my thanks to IPAM and all the staff working there for providing such a wonderful working environment to all the participants.

I am also grateful to all of the people who work in the mathematics depart-

ment here at UCLA for all of their help, and I would really like to extend my appreciation to Maggie Albert, Julie Honig and Babette Dalton.

This work was supported in part by the NIH through grants P20MH65166 and U54RR021813, and the NSF DMS Research through grants 0312222, ACI-0321917 and DMI-0327077. The financial support from UCLA is gratefully acknowledged.

Chapter 2 is a version of [HBO06]. Chapter 3 is a version of [HMO05].

VITA

- 1980 Born, Shaoyang, Hunan, China.
- 1997-2001 B.S. (Pure Mathematics)
 Peking University, China
- 2001-2002 Fellow
 Department of Mathematics, UCLA
- 2002 Summer Intern
 Advanced Computations Department, Stanford Linear Accelerator Center (SLAC), Stanford University
- 2002–2003 Teaching Assistant
 Department of Mathematics, UCLA.
- 2003 M.S. (Applied Mathematics)
 Department of Mathematics, UCLA.
- 2004–2005 Research Assistant
 Department of Mathematics, UCLA.
- 2005 Summer Intern
 Intelligent Vision and Reasoning Department, Siemens Cooperate Research (SCR), Princeton, New Jersey.
- 2005–2006 Research Assistant
 Department of Mathematics, UCLA.

PUBLICATIONS AND PRESENTATIONS

TI-CHIUN CHUNG, LIN HE, AND TONG FANG, *MR image reconstruction from sparse radial samples using Bregman iteration*, Oral presentation, International Society for Magnetic Resonance in Medicine (ISMRM) 14th Scientific Meeting and Exhibition (May 2006).

LIN HE, MARTIN BURGER, AND STANLEY OSHER, *Iterative total variation regularization with non-quadratic fidelity*, accepted by J. of Mathematical Imaging and Vision (2006).

LIN HE, ANTONIO MARQUINA, AND STANLEY OSHER, *Blind deconvolution using TV regularization and Bregman iteration*, International Journal of Imaging Systems and Technology, 15 (2005), pp. 74–83.

ABSTRACT OF THE DISSERTATION

Applications and Generalizations of the Iterative Refinement Method

by

Lin He

Doctor of Philosophy in Mathematics

University of California, Los Angeles, 2006

Professor Stanley Osher, Chair

An iterative refinement procedure has recently been introduced to the imaging denoising field. This procedure has significantly improved image quality. Here we first generalize this procedure to any variational model with non-quadratic convex fidelity terms. By using a suitable sequence of penalty parameters, we solve the issue of well-definedness of the minimization problems in each step of the iterative procedure. We also obtain rigorous convergence results for exact and noisy data. Secondly we apply the refinement method to solve blind deconvolution problems. We formulate a new time dependent variational model with additional constraints incorporated to preserve the mass and the nonnegativity of the signal and the kernel. We present an analytical study of the model by discussing the uniqueness of the solution, convergence to steady state and a priori parameter estimation. Finally we apply the iterative refinement method to MR image reconstruction from sparsely sampled data in the k -space (i.e. the Fourier transform domain). We formulate our cost functional by summing a fidelity term and an L^1 norm of a sparse representation of the reconstructed image. The fidelity term is used to satisfy the constraint condition of the undersampled data.

The sparse representation is realized by the total variation regularization and/or the wavelet transform. Furthermore, to speed up the iterative refinement method we use a nonlinear inverse scale space method. In addition to above, we present numerical experiments to show supporting evidence for all our work.

CHAPTER 1

Introduction

The field of image processing is full of many interesting tasks. Its applications are seen in many different fields such as astronomy, computer graphics, medical imaging and security identification. Among the tasks performed in image processing are image compression, image restoration, measurement extraction and etc. Our research has been focused on image extraction (i.e. segmentation), image restoration of noisy and/or blurry images and magnetic resonance (MR) image reconstruction from the k-space, i.e. the frequency space.

There are many different approaches to work on image processing. There are statistical methods for texture analysis. Also, there are Fourier and Wavelet (cf. [Mey93, Mal99]) transformations which have inspired the JPEG image representation. These methods have also led to new methods such as Ridgelets and Curvelets (cf. [CD02, CDD05, CG02]). Another approach to image processing is variational PDE based methods (cf. [ROF92, VO03, OSV03]). In our research we used the last two methods. First, by using Variational PDE methods it is proved that we can preserve edges well between homogeneous regions; Second, by using image representation, more precisely, by using image sparse representation it is possible for us to reconstruct MR image exactly from incomplete information about Fourier transform coefficients (cf. [CRT04, CR05, LLD05, CHF06]).

The organization of this dissertation is as follows. In this chapter we begin by reviewing some existing image restoration methods, discuss how total vari-

ational regularization is connected with MR image reconstruction and give a detailed derivation of the iterative refinement procedure. In Chapter 2, we focus on generalizing the iterative refinement procedure for any variational model with non-quadratic fidelity terms. In the following chapter we apply the iterative refinement method to solve the blind deconvolution problem. Next, we apply the iterative refinement method again to MR image reconstruction from sparsely sampled data in the Fourier transform domain. The numerical results are presented at the end of each chapter from Chapter 2 to Chapter 4. We conclude this dissertation by discussing future work in Chapter 5.

1.1 Some Total Variational Based Image Restoration Models

We consider a given observed image $f : \Omega \rightarrow \mathbb{R}$ as a function of $L^2(\Omega)$, where Ω is an open and bounded subset of \mathbb{R}^2 . For example, Ω could be the computer screen. The image f is comprised of a combination of homogeneous regions, contours and oscillatory patterns such as noise or texture. This scalar-valued or gray-scale image $f(x, y)$ at the pixel of (x, y) usually takes values between 0-black and 256-white.

1.1.1 Rudin, Osher and Fatemi (ROF) Total Variation Minimization

We assume the image f has the following form :

$$f = k * u + n,$$

where k is a convolution kernel with compact support (e.g., discrete Gaussian kernel), u is the original image and n is Gaussian white noise. Our task is to recover the unknown u and/or k from f .

Given knowledge of the kernel k , one of the most successful approaches to solve the above problem is the total variation (TV) regularization method proposed by Rudin, Osher and Fatemi (cf. [ROF92, RO94]). They solved the following constrained minimization problem:

$$u = \arg \min_{u \in BV(\Omega)} \left\{ \|u\|_{BV} + \frac{\lambda}{2} \|k * u - f\|_{L^2}^2 \right\}, \quad (1.1)$$

where λ is a positive parameter. The regularization functional is the BV -seminorm, defined as the following :

Definition 1 For any functional $u : \Omega \rightarrow \mathbb{R}$ as an element of $L^2(\Omega)$, the BV -seminorm of u

$$\|u\|_{BV} = \sup_{|g|_{\infty} \leq 1, g \in C_c^1(\Omega)^2} \int_{\Omega} u(\nabla \cdot g) dx,$$

where $|g| = \sqrt{(|g_1|^2 + |g_2|^2)}(x)$ and $C_c^1(\Omega)$ denotes the class of continuously differentiable functions of compact support in Ω .

The key feature of total variation regularization is the fact that it allows for (and even favors) discontinuous solutions, i.e., images with sharp edges. Nevertheless, this regularization suppresses oscillations and can still eliminate high-frequency noise.

For general convolution kernel k , the restoration of the image u in (1.1) is usually a numerically ill-posed problem. Vogel and Oman in [VO98] have demonstrated that the ill-conditioning in solving the Euler-Lagrange equation of (1.1) is due to the fact that the compact convolution operator, defined by k , has eigenvalues which cluster to zero.

Here we are interested in recovering the true image u and the kernel k without having any *a priori* knowledge of k and u . This so-called blind deconvolution problem (cf. [YK96, CW98, YK99, HMO05]) is studied in Chapter 3. We

not only propose a new time dependent model but also present an analytical study discussing the uniqueness of the solution and a priori parameter estimation. Furthermore, we apply the newly discovered idea of iterative regularization in [OBG05] to recover finer scales.

1.1.2 Meyer's Analysis on the ROF model

If k is the delta function, $\delta * u = u$, (1.1) becomes a pure denoising algorithm. It is usually referred as the ROF model:

$$u = \arg \min_{u \in BV(\Omega)} \left\{ \|u\|_{BV} + \frac{\lambda}{2} \|u - f\|_{L^2}^2 \right\}. \quad (1.2)$$

The ROF model is well known for its ability to remove noise while preserve sharp edges.

However, the ROF model (1.2) has certain limitations. The ideal result of minimizing ROF model (1.2) would be to decompose f into a true signal u and an additive noise $v := f - u$. In practice, however, we often find some signal in v . In [Mey01], Meyer has analyzed the ROF model by introducing the dual of the BV space to extract both the u component in BV and the v component as an oscillating function (texture or noise). For the topic of extracting texture from images using variational methods, we refer to [VO03, AC05].

The norm of the dual of the BV space is denoted as the $*$ norm:

Definition 2 *Let G denote the Banach space consisting of all generalized functions $v(x)$ which can be written as*

$$v(x) = \partial_1 g_1(x) + \partial_2 g_2(x) \quad (1.3)$$

where $g_1, g_2 \in L^\infty(\mathbb{R}^2)$.

The norm of v in G $\|v\|_*$ is defined as the lower bound of all L^∞ norms of the functions $|g|$ where $g = (g_1, g_2)$, $|g| = \sqrt{(|g_1|^2 + |g_2|^2)(x)}$ and where the infimum is computed over all decompositions (1.3) of v .

It is easy to derive the following lemma.

Lemma 1

$$\int uv dx \leq \|u\|_{BV} \|v\|_* \quad (1.4)$$

Based on (1.4), Meyer showed the following proposition.

Proposition 1 For the ROF denoising problem (1.2), if the norm of f in G does not exceed $\frac{1}{\lambda}$, then the optimal ROF decomposition of f is given by

$$u = 0, v = f. \quad (1.5)$$

Otherwise,

$$\|v\|_* = \frac{1}{\lambda}, \int u(x)v(x) dx = \frac{1}{\lambda} \|u\|_{BV}. \quad (1.6)$$

The Euler-Lagrange equation from the ROF minimization is :

$$-\nabla \cdot \frac{\nabla u}{|\nabla u|} = \lambda(f - u) = \lambda v. \quad (1.7)$$

The term v was previously thrown away even though it is an element in G with its $*$ norm = $\frac{1}{\lambda}$ when $\|f\|_* > \frac{1}{\lambda}$. In [OBG05], a correction step was taken which added v back to f and then $f + v$ was processed as the new noisy data by the ROF minimization procedure again. Their iterative regularization procedure is: decompose $f + v_{k-1} = u_k + v_k$ by the ROF model, where $k \geq 1$ and $v_0 = 0$, until the first k for which $\|f - u_k\| \leq \delta$ where δ is the L^2 norm of the noise. This iterative idea is equivalent to using the Bregman distance and Bregman iteration. This will be shown in Section 4.3.

1.1.3 Total Variational Models with Non-quadratic Fidelity

From above, we know that the term v that represents noise is usually ignored; however, it often contains textures. In order to preserve textures, Meyer suggested the modified variational problem

$$u = \arg \min_{u \in BV(\Omega)} \{ \|u\|_{BV} + \lambda \|f - u\|_* \}, \quad (1.8)$$

whose computational solution is a rather difficult task because of the nature of the norm $\|\cdot\|_*$ (cf. [AAB05, GY04]).

We also mention here another version of the total variation regularization model that has been studied recently by Chan and Esedoglu [CE05] (see also the references therein). They used the L^1 norm instead the square of L^2 norm as in ROF, as a measure of fidelity between the observed and denoised image. Given an observed image $f \in L^1(\Omega)$, this model is based on the following variational problem:

$$u = \arg \min_{u \in BV(\Omega)} \{ \|u\|_{BV} + \lambda \|u - f\|_{L^1} \}. \quad (1.9)$$

Even though this minimization problem may have a lack of uniqueness and the continuous dependence on data is not clear, it has many desirable and some unexpected consequences in applications such as reconstruction of binary images, multiscale image decomposition (cf. [YGO05, HBO06]), and data driven parameter selection.

Both generalizations of total variational models with respect to the fidelity functional can yield an improvement with respect to some aspects, but they share the systematic error yielding a decrease in the total variation and therefore a loss of information in either case. In order to overcome this issue, in Chapter 2 we generalize the iterative version of Tikhonov regularization introduced in [OBG05] that could be applied to any total variational model with non-quadratic fidelity

terms.

1.2 From TV Regularization to MR Image Reconstruction

Many applications in magnetic resonance imaging (MRI) require very short scan time while the image reconstruction can be performed off-line. To this end, during the scanning process it is necessary to sample the frequency plane (or k -space) very sparsely. This usually results in image artifacts and/or low signal to noise ratio (SNR).

As a result, in [CRT04, CR05] Candes et al. demonstrate the possibility to exactly recover signal from incomplete frequency information for numerical phantoms. Their method is based on minimizing a sparse representation of a target signal while enforcing the constraint so that the original sparse frequency samples are maintained. In their work, total variation regularization is used as one of the sparse transformations for piecewise constant functions. This regularization actually performs better than the wavelet transform (cf. [LLD05]) due to the capability of total variation to recover piecewise smooth functions without smoothing sharp discontinuities.

In Chapter 4, we propose a new formulation in terms of the sparse transformation while again apply the iterative refinement method to obtain finer scales for our reconstructed image. Furthermore, we also experiment a nonlinear inverse scale space method in addition to the iterative refinement procedure. Compared with the iterative refinement procedure, the nonlinear scale space method which provides similar results is shown to be a more straightforward and efficient algorithm.

1.3 The iterative Refinement Method

Consider a generalized minimization model:

$$u = \arg \min_u \{J(u) + \lambda H(u, f)\}, \quad (1.10)$$

where $J(u)$ is called the regularization term and $H(u, f)$ is called the fidelity term. Both functionals are required to be convex and $J(u)$ needs to be non-negative.

1.3.1 The Bregman Distance

Definition 3 *Given a differentiable function φ , the Bregman distance (cf. [Bre67]) is defined by*

$$D_\varphi(x, y) = \varphi(x) - \varphi(y) - \langle x - y, \partial\varphi(y) \rangle,$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product in \mathbb{R}^n and $\partial\varphi(y)$ is an element of the sub-gradient of φ at point y .

Since $J(u)$ is convex, $D_J(u, v)$ is also convex in u for each v . The quantity $D_J(u, v)$ is not a distance in the usual sense; e.g., in general, $D_J(u, v) \neq D_J(v, u)$ and also the triangle inequality does not hold. However, it is a measure of closeness in that $D_J(u, v) \geq 0$ and $D_J(u, v) = 0$ if $u = v$ (if and only if for strictly convex functionals).

1.3.2 The iterative Refinement Procedure

Denote $u_0 = 0$, an iterative procedure is given by the sequence of variational problems

$$u_m = \arg \min_{u \in BV(\Omega)} P(u_{m-1}, f, \lambda) =: \{D_J(u, u_{m-1}) + \lambda H(u, f)\} \quad (1.11)$$

We notice that the first iteration of the given procedure (1.11) is in fact the minimization of the generalized model (1.10). We present here a detailed analysis of the procedure by specifically considering the fidelity term $H(u, f) = \frac{1}{2} \|k * u - f\|_{L^2}^2$.

Take the sub-gradient of (1.11), we obtain the following for $m \geq 1$:

$$\lambda H_u(u_m, f) + \partial J(u_m) - \partial J(u_{m-1}) = 0. \quad (1.12)$$

So for $m = 1$,

$$\lambda \widehat{k} * (k * u_1 - f) + \partial J(u_1) = 0;$$

Write $f - k * u_1 = v_1$, i.e.

$$\partial J(u_1) = \lambda \widehat{k} * v_1.$$

Plug the above equation into (1.12) for $m = 2$, it becomes

$$\lambda \widehat{k} * (k * u_2 - f) + \partial J(u_2) - \lambda \widehat{k} * v_1 = 0.$$

It can be simplified as

$$\lambda \widehat{k} * (k * u_2 - f - v_1) + \partial J(u_2) = 0.$$

Write $f + v_1 - k * u_2 = v_2$, we then have

$$\partial J(u_2) = \lambda \widehat{k} * v_2.$$

From above we can see, if we denote $v_0 = 0$ and

$$v_m = f + v_{m-1} - k * u_m \quad \forall m \in \mathbb{N},$$

we will have

$$\partial J(u_m) = \lambda \widehat{k} * v_m.$$

In other words, to apply the Bregman iteration we only have to change the observed image from f to $f + v_{m-1}$ at the m -th iteration. Therefore, solving the

sequence of minimization problems of (1.11) is equivalent to solving a sequence of minimization problems

$$u_m = \arg \min_{u \in BV(\Omega)} Q(v_{m-1}, f, \lambda) =: \{J(u) + \lambda H(u, f + v_{m-1})\}. \quad (1.13)$$

Thus all the nice properties of the original minimization problem (1.10) are kept for (1.11), in particular this indicates the existence of the minimizer of (1.11). Furthermore, it has been shown that in [OBG05] for the pure denoising problem the sequence $\{u_k\}$ converges monotonically in L^2 to the noisy image f . More importantly, as k increases, for $k \leq \bar{k}$ and sufficiently small λ , the Bregman distance between u_k and the true noise free image \tilde{u} is decreasing. Here \bar{k} is defined as

$$\max\{k \in \mathbb{N} \mid H(u_k, f) > \tau H(\tilde{u}, f)\} \quad (1.14)$$

where $\tau > 1$. We refer to [OBG05] for more details of the analysis. Comparing with the denoised image from the ROF model (1.2), we see that this iterative algorithm improves the results significantly.

CHAPTER 2

Iterative TV Regularization with Non-Quadratic Fidelity

A key observation in the analysis of the iterative scheme is a rewritten version as the generalization of proximal point or Bregman iterations (cf. [Bre67, CT93]). These methods are usually used to solve problems of the form

$$\min\{g(u) : u \in \mathbb{R}^n\},$$

where $g : \mathbb{R}^n \mapsto (-\infty, \infty]$ is a proper, lower semi-continuous convex function. Proximal point methods generate a sequence $\{u_k\}$ from the minimization problem

$$u_k = \arg \min_{u \in \mathbb{R}^n} \{g(u) + \lambda_k^{-1} D_J(u, u_{k-1})\}. \quad (2.1)$$

In [CT93], the convergence analysis of the proximal point algorithms is based on two assumptions:

- $\liminf_{k \rightarrow \infty} \{\lambda_k : \geq 0\} > 0$;
- f is bounded below, and the iterative scheme generates a sequence $\{u_k\}$ such that $u_k \in \mathbb{R}^n$ for all k .

The second line assumes that a solution of the minimization problem (2.1) exists, which is however not straight-forward to show for many important cases. We shall introduce our a modified model with suitable decay of the penalty parameters

which allows us to prove the existence of the minimizers in Section 2.1.1 and 2.1.2. Furthermore, following the theoretical proof in [OBG05], monotonicity and convergence theorems are also obtained in Section 2.1.3. In Section 2.2, some numerical results are given using various fidelity functionals H in (1.10) while the regularization term J is kept as BV term. Comparisons with other methods of similar spirit are also presented.

2.1 Analysis of a Modified Iterative Regularization Procedure

In general it is quite easy to generalize the iterative regularization algorithm to variational models with a different regularization functional J , as long as this functional has suitable lower-semicontinuity and compactness properties in some topology (cf. [OBG05]). The analysis of the algorithm (with respect to well-definedness, convergence, regularization properties) can be transferred in a one-to-one way, only computational schemes have to be adapted to the specific form of the functional.

By far more challenging is the generalization of the procedure with respect to the fidelity term H . Of course, one can just write down (1.11) and try to solve the minimization problems in each step. However, the well-definedness of this procedure is not clear if H is not quadratic, since only in the quadratic each minimization problem in (1.11) can be rewritten in the same form as (??) and corresponding existence and uniqueness results for minimizers can be carried over.

2.1.1 Generalization of the Fidelity Term: A Modified Model

Inspired by Scherzer and Groetsch (cf.[SG01]) and Tadmor, Nezzar and Vese (cf.[TNV04]), who multiplied the parameter λ by two after each iteration step, we formulate a new iterative total variation regularization as

$$u_k = \arg \min_{u \in BV(\Omega)} E(u, u_{k-1}, f) := \left\{ H(u, f) + \frac{1}{2^{k-1}\lambda} D_J(u, u_{k-1}) \right\}, \quad (2.2)$$

where $J(u)$ satisfies the conditions assumed above, and $H(u, f) := h(u - f)$ with h being a nonnegative, convex, and positively homogeneous functional, which is continuous with respect to weak-* convergence in BV . Moreover, we assume that $h(c)$ does not vanish for constant functions $c \neq 0$, so that $h(u) + |u|_{BV}$ is indeed an equivalent norm on $BV(\Omega)$. For convenience we set $p = \partial J$, $u_{-1} = 0$ and $p_{-1} = 0$ so that u_0 is defined as the minimizer of $H(u, f) + 2\frac{J(u)}{\lambda}$. Using this new construction, we are able to overcome the problem of lower boundedness and thus obtain well-definedness in Section 2.1.2 and prove some monotonicity and convergence theorems in Section 2.1.3.

The models in [TNV04, SG01], which inspire our setting, did not use a proximal point iteration but instead used Tikhonov-Morozov iteration for which the following convergence analysis does not apply. The difference is that one replaces $D_J(u, u_{k-1})$ in (2.2) by $J(u - u_{k-1})$. We call this Tikhonov-Morozov iteration with the multiscale coefficient alone a T-N-V method even if $H(u, f)$ is not the L^2 square fidelity term as used in the original papers.

We mention that the multiscale decomposition of T-N-V concerned the function u directly, in [TNV04] the method was even set up as a sequence of minimization problems for the update $v_k = u_k - u_{k-1}$. For choosing λ constant, one would obtain $v_k \equiv 0$ for $k \geq 1$ and hence, λ has to be decreased in order to access small scale features that have been eliminated in the previous steps. Our model

(2.2) can rather be considered as a multiscale decomposition for the dual variable p_k . Note that the optimality condition for the variational problem in (2.2) reads

$$p_k - p_{k-1} \in -2^{k-1} \lambda \partial_u H(u_k, f),$$

and hence yields the update for k . In this case the dyadic choice is not as obvious. For an arbitrary variable choice λ_k we obtain (after summation of the optimality conditions and using the notation $q_j \in \partial_u H(u_j, f)$)

$$p_k = p_0 + \sum_{j=1}^{k-1} \lambda_j q_j.$$

In the quadratic case $H(u_j - f) = \|u_j - f\|^2$, the subgradient is single valued and given by $q_j = 2(u_j - f)$, and in particular it scales with the residual. Since the residual tends to zero, the sum $\sum_{j=1}^{k-1} \lambda_j q_j$ can also be bounded if λ is constant (and indeed is as can be deduced from the results in [OBG05]). In the case of positively one-homogeneous fitting functionals such the ones in (1.8) and (1.9), the norm of the subgradients of H will be of order one in the dual space to the fitting norm. E.g. for the L^1 -functional, we obtain that $\|q_j\|_\infty = 1$ for $q_j \in \partial_u H(u_j, f)$ (unless $u_j \equiv f$, one can think of q_j as the sign of $u_j - f$ almost everywhere). Thus, the update in the dual variable will actually increase during the iteration in these spaces (except $u_k \equiv f$, which leads to no subsequent changes in the iteration). In the worst case one has to expect that the norm of p_k is given by

$$\|p_k\|_\infty \sim \sum_{j=1}^k \lambda_j \|q_j\|_\infty = \sum_{j=1}^k \lambda_j.$$

In the iteration step from k to $k + 1$, the term $\frac{1}{\lambda_{k+1}} p_k$ appears in the Bregman distance and subsequently in the optimality condition. In order to obtain a reasonable scaling with the subgradient of H , also p_k should be of order one and this is not true for constant λ since from the above argument we would expect

that

$$\frac{1}{\lambda_{k+1}} \|p_k\|_\infty \sim k.$$

On the other hand, the scaling to order one can be achieved directly by the dyadic choice $\lambda_k = 2^{k-1}\lambda$, for which we can expect

$$\frac{1}{\lambda_{k+1}} \|p_k\|_\infty \sim \frac{1}{2^k \lambda} \sum_{j=1}^k 2^{k-1} \lambda = 1.$$

We mention that the same type of scaling argument is possible for other positively one-homogeneous functionals. As we shall prove below, the dyadic choice leads to a well-defined model for rather general fidelity functionals, but from the scaling one might argue that it is not the optimal one if H is not positively one-homogeneous (and one could find the optimal one from the scaling of subgradients). However, positively one-homogeneous fidelity terms are by far the most interesting examples in total variation based image restoration (except the previously analyzed quadratic ones), so that we focus on the dyadic choice here. Possible extensions for other functionals can then be carried out along the lines of our analysis.

2.1.2 Well-Definedness of the New Iteration Model

We now derive the existence of minimizers for the modified iterative method (2.2):

Theorem 1 *Under the above conditions, the iteration scheme (2.2) yields a well-defined sequence $u_k \in BV(\Omega)$.*

Proof: With u_k defined via (2.2) we obtain the Euler-Lagrange equations

$$q_k + \frac{1}{2^{k-1}\lambda}(p_k - p_{k-1}) = 0, \quad q_k \in \partial_u H(u_k, f), \quad p_j \in \partial J(u_j), \quad (2.3)$$

Using (2.3), first we express $p_k \in \partial J(u_k)$ in terms of the subgradients of the functionals $H(u_j, f)$, $0 \leq j < k$, which are denoted by $q_j \in \partial_u H(u_j, f)$:

$$\begin{aligned}
\frac{1}{2^{k-1}\lambda} p_k &= \frac{1}{2^{k-1}\lambda} p_{k-1} - q_k \\
&= \frac{1}{2^{k-1}\lambda} p_{k-2} - \frac{1}{2} q_{k-1} - q_k \\
&= \dots \\
&= \frac{1}{2^{k-1}\lambda} p_0 - \sum_{j=1}^k \frac{1}{2^{k-j}} q_j = - \sum_{j=0}^k \frac{1}{2^{k-j}} q_j. \tag{2.4}
\end{aligned}$$

Next we define $v := f - u$, thus $H(u, f) = h(v)$. Due to the convexity, nonnegativity, and homogeneity of the functional h , we have for any $w \in BV$, $q \in \partial h(w)$, and $t > 0$,

$$\begin{aligned}
t\langle v, q \rangle &= \langle (tv + w) - w, q \rangle \leq h(tv + w) - h(w) \\
&\leq h(tv) = th(v).
\end{aligned}$$

Since $t > 0$, we obtain the inequality

$$\langle v, q \rangle \leq h(v), \quad \forall v, w, q \in \partial h(w). \tag{2.5}$$

In particular, if we choose $w = f - u_j$, then (2.5) becomes

$$\langle u - f, q_j \rangle \geq -H(u, f). \tag{2.6}$$

Using (2.4) and (2.6) and defining $\{w_n\}$ as a minimizing sequence to u_k , we know there exist $M > 0$ such that

$$\begin{aligned}
M &\geq H(w_n, f) + \frac{1}{2^{k-1}\lambda}J(w_n) - \frac{1}{2^{k-1}\lambda}J(u_{k-1}) - \langle w_n - u_{k-1}, \frac{1}{2^{k-1}\lambda}p_{k-1} \rangle \\
&= H(w_n, f) + \frac{1}{2^{k-1}\lambda}J(w_n) - \frac{1}{2^{k-1}\lambda}J(u_{k-1}) + \langle u_{k-1} - f, \frac{1}{2^{k-1}\lambda}p_{k-1} \rangle \\
&\quad + \frac{1}{2}\langle w_n - f, \sum_{j=1}^{k-1} \frac{1}{2^{k-1-j}}q_0 \rangle \\
&\geq \frac{1}{2^{k-1}\lambda}J(w_n) - \frac{1}{2^{k-1}\lambda}J(u_{k-1}) + \langle u_{k-1} - f, \frac{1}{2^{k-1}\lambda}p_{k-1} \rangle \\
&\quad + H(w_n, f) - H(w_n, f) \sum_{j=0}^{k-1} \frac{1}{2^{k-j}} \\
&\geq \frac{1}{2^{k-1}\lambda}J(w_n) - \frac{1}{2^{k-1}\lambda}J(u_{k-1}) + \langle u_{k-1} - f, \frac{1}{2^{k-1}\lambda}p_{k-1} \rangle.
\end{aligned}$$

Since both u_{k-1} and f are fixed and bounded independent of n , the sequence $\{J(w_n)\}$ is bounded. From the compactness and lower semi-continuous properties of the functional J , we can conclude that the iteration model (2.2) is well-defined. \square

From the proof one observes that the choice 2^{k-1} is crucial but not the only choice for the penalization parameters, but it suffices to choose a sequence λ_k such that $\sum_{j=1}^{\infty} \frac{1}{\lambda_j} \leq 1$.

2.1.3 Convergence Analysis

We now study some convergence properties of the new iterative regularization process. Our analysis below basically follows the lines in [OBG05]. In particular we shall see that all desirable properties like monotonicity of residual and convergence for the exact image and for the noisy image still hold for our new model.

Proposition 2 (Monotonicity) *Let $u_k \in BV(\Omega)$ be a sequence defined by*

(2.2). Then the sequence $H(u_k, f)$ is monotonically non-increasing and satisfies:

$$H(u_k, f) \leq H(u_k, f) + \frac{1}{2^{k-1}\lambda} D(u_k, u_{k-1}) \leq H(u_{k-1}, f). \quad (2.7)$$

Moreover, if $J(u) < \infty$, then we have:

$$D(u, u_k) + D(u_k, u_{k-1}) + 2^{k-1}\lambda H(u_k, f) \leq 2^{k-1}\lambda H(u, f) + D(u, u_{k-1}). \quad (2.8)$$

Proof: Using the definition of subgradient and the fact that u_k minimizes $E(u, u_{k-1}, f)$, we have

$$\begin{aligned} H(u_k, f) &\leq H(u_k, f) + \frac{1}{2^{k-1}\lambda} D(u_k, u_{k-1}) = E(u_k, u_{k-1}, f) \\ &\leq E(u_{k-1}, u_{k-1}, f) = H(u_{k-1}, f). \end{aligned}$$

To prove (2.8), we use the convexity of the function $H(u, f)$ and a standard decomposition of the Bregman distance (with the same notation for subgradients as above).

$$\begin{aligned} D(u, u_k) - D(u, u_{k-1}) + D(u_k, u_{k-1}) &= J(u) - J(u_k) - \langle u - u_k, p_k \rangle \\ &\quad - J(u) + J(u_{k-1}) + \langle u - u_{k-1}, p_{k-1} \rangle \\ &\quad + J(u_k) - J(u_{k-1}) - \langle u_k - u_{k-1}, p_{k-1} \rangle \\ &= \langle u - u_k, p_{k-1} - p_k \rangle = \langle u - u_k, 2^{k-1}\lambda q_k \rangle \\ &\leq 2^{k-1}\lambda [H(u, f) - H(u_k, f)]. \square \end{aligned}$$

The above proposition states that the fidelity term $H(u_k, f)$ is decreasing. Furthermore, if we choose $u = f$ in (2.8) and use the fact that $H(f, f) = 0$, then we obtain for "exact data" satisfying $J(f) < \infty$ that

$$\begin{aligned} D(f, u_k) &\leq D(f, u_k) + D(u_k, u_{k-1}) \\ &\leq D(f, u_{k-1}) + 2^{k-1}\lambda (H(f, f) - H(u_k, f)) \\ &\leq D(f, u_{k-1}). \end{aligned} \quad (2.9)$$

Hence, the Bregman distance between the solution u_k of (2.2) at the k -th iteration step and the image f is decreasing, too. In fact, as illustrated by computations in Section 2.2, both the Bregman distance $D(u_k, f)$ and the fidelity term $H(u_k, f)$ decrease as k increases.

Theorem 2 (Exact Data) *Let f satisfy $J(f) < \infty$ and let u_k be a sequence generated by (2.2) with data f . Then*

$$H(u_k, f) \leq \frac{J(f)}{2^k \lambda} \quad (2.10)$$

and in particular $\{u_k\}$ is a minimizing sequence of $H(\cdot, f)$.

Moreover, $u_k \rightarrow f$ in the weak- $$ topology of $BV(\Omega)$.*

Proof: Summing all inequalities (2.8) from 1 to k , we have:

$$\sum_{j=1}^k [D(u_j, u_{j-1}) + 2^{j-1} \lambda H(u_j, f)] \leq D(f, u_0) - D(f, u_k). \quad (2.11)$$

From the fact that $H(u, f)$ is a convex function, we also have the following inequality:

$$J(u_0) + \lambda H(u_0, f) \leq J(f) - D(f, u_0). \quad (2.12)$$

Now adding (2.12) to (2.11) and using both $D(u_j, u_{j-1}) \geq 0$ and the monotonicity property of $H(u_j, f)$, we conclude

$$2^k \lambda H(u_k, f) \leq J(f).$$

Based on (2.11) and (2.12), we also obtain

$$\begin{aligned}
J(f) &\geq \sum_{j=1}^k D(u_j, u_{j-1}) + J(u_0) \\
&= J(u_k) - \sum_{j=1}^k \langle p_{j-1}, u_j - u_{j-1} \rangle \\
&= J(u_k) - \langle p_{k-1}, u_k - f \rangle + \sum_{j=1}^{k-1} \langle p_j - p_{j-1}, u_j - f \rangle \\
&= J(u_k) + \lambda \langle H_u(u_0, f) + \sum_{j=1}^{k-1} 2^{j-1} H_u(u_j, f), u_k - f \rangle \\
&\quad - \sum_{j=1}^{k-1} \langle 2^{j-1} \lambda H_u(u_j, f), u_j - f \rangle \\
&\geq J(u_k) - \lambda 2^{k-1} H(u_k, f) - \sum_{j=1}^{k-1} 2^{j-1} \lambda H(u_j, f) \\
&\geq J(u_k) - \frac{3}{2} J(f).
\end{aligned}$$

Therefore $J(u_k) \leq \frac{5}{2} J(f)$. Because the level sets of $\{u \in \mathcal{U} \mid J(u) \leq M\}$ are compact, the further assertions then follow by standard weak-* convergence techniques analogous to the arguments in [OBG05]. \square

The above result is important from a theoretical point of view since it verifies convergence of the method. In practice however, the given data do not represent the exact but rather a noisy version of the image, since otherwise one would not need to denoise it. Therefore we consider the case where f contains noise in the following. It is well-known for iterative methods that a regularizing effect is obtained only via appropriate stopping in dependence of the noise level, which is given for the fidelity functional H as

$$H(u_*, f) \leq \delta. \tag{2.13}$$

We again inspect the decrease of the distance between the iterations and the noisy image, which is now guaranteed only until the residual becomes to small:

Proposition 3 *Let u_* be the true noise free image and let f be a given noisy version satisfying (2.13). Then, as long as $H(u_k, f) > \delta$, the Bregman distance between u_k and the true solution u_* decreases, i.e.,*

$$D(u_*, u_k) \leq D(u_*, u_k) + D(u_k, u_{k-1}) < D(u_*, u_{k-1}). \quad (2.14)$$

Proof: Plugging the noise free image u_* into (2.8), we obtain

$$D(u_*, u_k) + D(u_k, u_{k-1}) + 2^{k-1}\lambda H(u_k, f) \leq D(u_*, u_{k-1}) + 2^{k-1}\lambda H(u_*, f)$$

With the assumption $H(u_k, f) > \delta$ and (2.13), we can conclude (2.14). \square

From Proposition 3 we can deduce that the *generalized discrepancy principle* (cf. [Pla96]) is a good candidate as a stopping rule for (2.2), i.e., the iteration is stopped at the index

$$k_* = \max\{k \in N | H(u_k, f) \geq \tau\delta\} \quad (2.15)$$

where $\tau > 1$. Important features of k_* are studied. For given $\delta > 0$, we shall consider u_{k_*} as the regularized solution, i.e., the result of our iterative scheme. With such a stopping criterion, we can obtain a so-called *semi-convergence property*, i.e., the regularized solutions converge to u_* as $\delta \rightarrow 0$, more precisely:

Theorem 3 (Noisy Data) *The stopping index k_* is well-defined by (2.15) for any $\delta > 0$, and $k_*(\delta) = \mathcal{O}(\log \delta)$.*

Moreover, let f_m denotes a sequence of noisy data satisfying (2.13) with noise level $\delta_m \rightarrow 0$. If we denote by $u^m = u_{k_(\delta_m)}$ the regularized solutions obtained for data f_m , then there exists a subsequence u^{m_ℓ} that converges in the weak-* topology of BV , and the limit of each convergent subsequence is a minimizer of $H(\cdot, u_*)$. Furthermore, if u_* is the unique minimizer of $H(\cdot, u_*)$ then the whole sequence u^m converges to u_* in the weak-* topology.*

Proof: It is easy to see that k_* is well-defined because $H(u_k, f)$ is monotonically decreasing and $H(u_k, f) \rightarrow 0$ as k increases. Furthermore, if we sum the inequalities (2.8) from 1 to k , and then add the following inequality derived from the convexity of the functional $H(u, f)$

$$J(u_0) + D(u_*, u_0) + \lambda H(u_0, f) \leq J(u_*) + \lambda H(u_*, f), \quad (2.16)$$

we obtain:

$$2^{k_*} \lambda H(u_{k_*}, f) \leq J(u_*) + \sum_{j=0}^{j=k_*} 2^{j-1} \lambda H(u_j, f) \leq 2^{k_*} \lambda \delta + J(u_*).$$

This means

$$\tau \delta \leq H(u_{k_*}, f) \leq \delta + \frac{J(u_*)}{2^{k_*} \lambda},$$

i.e. $2^{k_*} \leq \frac{J(u_*)}{\lambda(\tau-1)\delta}$.

In order to prove the weak-* convergence of subsequences, it suffices to prove that the BV-norm of u^m is uniformly bounded with respect to m , which can be obtained by analogous arguments to the proofs of Theorem 2 and the semiconvergence result in [OBG05]. \square

2.2 Numerical Results

In this section, we will present some numerical results obtained from two different fidelity term models by using our iterative regularization procedure. We shall also compare them with the results from the original iterative regularization procedure (cf. [OBG05]) and the hierarchical decomposition algorithm by Tadmor-Nezzar-Vese(T-N-V)(cf. [TNV04, SG01]). For the sake of appropriate comparison, we change the coefficient $\frac{1}{2^{k-1}\lambda}$ to $\frac{1}{2^k\lambda}$ in the model (2.2), which is equivalent to doubling λ , and finally allows to choose the same parameter λ in all three methods.

2.2.1 L^2 Fidelity Model

We start with the iterative total variation regularization using an L^2 fidelity term (but not the square of the L^2 norm as in [OBG05], i.e.,

$$u_k \in \arg \min_{u \in BV(\Omega)} \left\{ \|u - f\|_{L^2} + \frac{1}{2^k \lambda} D(u, u_{k-1}) \right\}. \quad (2.17)$$

We start with a noisy satellite image with Gaussian white noise with $\delta = 20.0$ and $SNR = 11.5$. In Figures 2.1, 2.2 and 2.3 provide a comparison of the three methods mentioned above with $\lambda = 10.0$.

In Figure 2.1, (d)-(f) and (j)-(l) display the recovered image u_k resulted from our new Bregman iteration, and (g)-(i) and (m)-(o) display the corresponding residual image $u_k - f$. The choice of a small value of λ yields an oversmoothed iteration u_0 , and u_k gradually improves as k increases until the stopping criterion becomes valued at $k_* = 3$. The plot of the residual in (p) confirms the monotonicity theorem, $H(u_k, f)$ simply decreases for all k . Moreover, at the k th step, $H(u_k, f)$ is bounded above by $\frac{J(f)}{2^k \lambda}$. The residual $H(u_k, f) = \|f - u_k\|$. The plots in (q) and (r) showing these error $\|u_* - u_k\|$ and $D(u_*, u_k)$ as a function of the iteration index are in good agreement with the noisy data theorem, which states that both the L^2 distance and the Bregman distance between u_k and the true solution u_* decrease as long as $H(u_k, f) > \delta$.

Figure 2.2 is obtained from the iterative total variation regularization model proposed in [OBG05], i.e, we do not multiply 2 to λ after each iteration. We observe similar but slightly worse results as in Figure 2.1. From (m) one observes that the residual $H(u_k, f)$ is decreasing slower than in the modified version, which is not too surprising since the residual is multiplied by increasing parameters in the modified minimization problems.

Figure 2.3 displays the results obtained from the T-N-V algorithm (i.e., Tikhonov-

Morozov) by using the $BV + L^2$ model. The recovered u_3 in (g) seems to be the visually best one. From the rescaled version of $u_3 - f$ in (j) one observes that almost all the important features of the satellite including the antenna are removed from the residual. The plot of the L^2 -error in (n) shows a very similar behaviour to the iterative total variation regularization; the closest distance between u_k and the exact image u_* , $\|u_3 - u_*\|_{L^2} = 12.3$, is even better than $\|u_{k_*} - u_*\|_{L^2} = 14.9$ in Figure 2.1. However, at the moment there are no rigorous convergence and monotonicity results for the T-N-V algorithm using the $BV + L^2$ norm. Moreover, it is not clear how the T-N-V algorithm behaves with respect to Bregman distances, which might provide more information about the convergence speed of certain features such as edges (cf. [BO04]). Unfortunately one cannot even compute suitable Bregman distances during the T-N-V algorithm since it does not provide any subgradients.

2.2.2 L^1 Fidelity Model

In this section we apply the new iterative regularization procedure using the L^1 fidelity term model

$$u_k \in \arg \min_{u \in BV(\Omega)} \left\{ \|u - f\|_{L^1} + \frac{1}{2^k \lambda} D(u, u_{k-1}) \right\}. \quad (2.18)$$

Corresponding to the motivation of the L^1 -model as a suitable fidelity for binary images, we consider the denoising of a black-and white finger print image as a test case

Figures 2.4, 2.5 and 2.6 show restored images with the three different methods (same order as before), for Gaussian white noise $\delta = 10.0$ and $SNR = 14.8$. The results show similar properties as noticed in the L^2 -case in the previous subsection. At the earlier stage, u_k is over-smoothed. As k increases, u_k is getting closer to the exact image u_* in terms of L^1 distance, until k_* is reached,

and subsequently noise is added to u_k . Comparing the value of $\|u_{k_*} - u_*\|_{L^1}$ among the three models we see that the differences are very small this time, less than 0.1. However, visually the result of the T-N-V model seems more noisy than the ones obtained from the iterative total variation method.

A difference from the L^2 -case is the behaviour of the original iterative total variation regularization. In the case of the L^2 fidelity term (see (m) in Figure 2.2), $H(u_k, f)$ was decreasing rather slowly, which is not the case for the L^1 fidelity term model (see (m) in Figure 2.5). In fact, the fidelity term $\|u_k - f\|_{L^1}$ seems to converge to zero even though we cannot prove that the original iteration model based on the L^1 fidelity term is well defined. A detailed analysis of this effect might be an interesting problem for future study.

Finally we draw attention to possible cartoon-texture decompositions (cf. [CE05, YGO05]) for the $BV + L^1$ model, which corresponds to (d), (g) and (e), (h) of Figure 2.4 for the first and second iteration. One observes that the main cartoon is already incorporated in the image after the first step, while the texture remains in the residual $v = f - u$. In the later steps the texture is gradually incorporated into the image. Hence, as a by-product of our algorithm we obtain cartoon-texture decompositions at different scales.

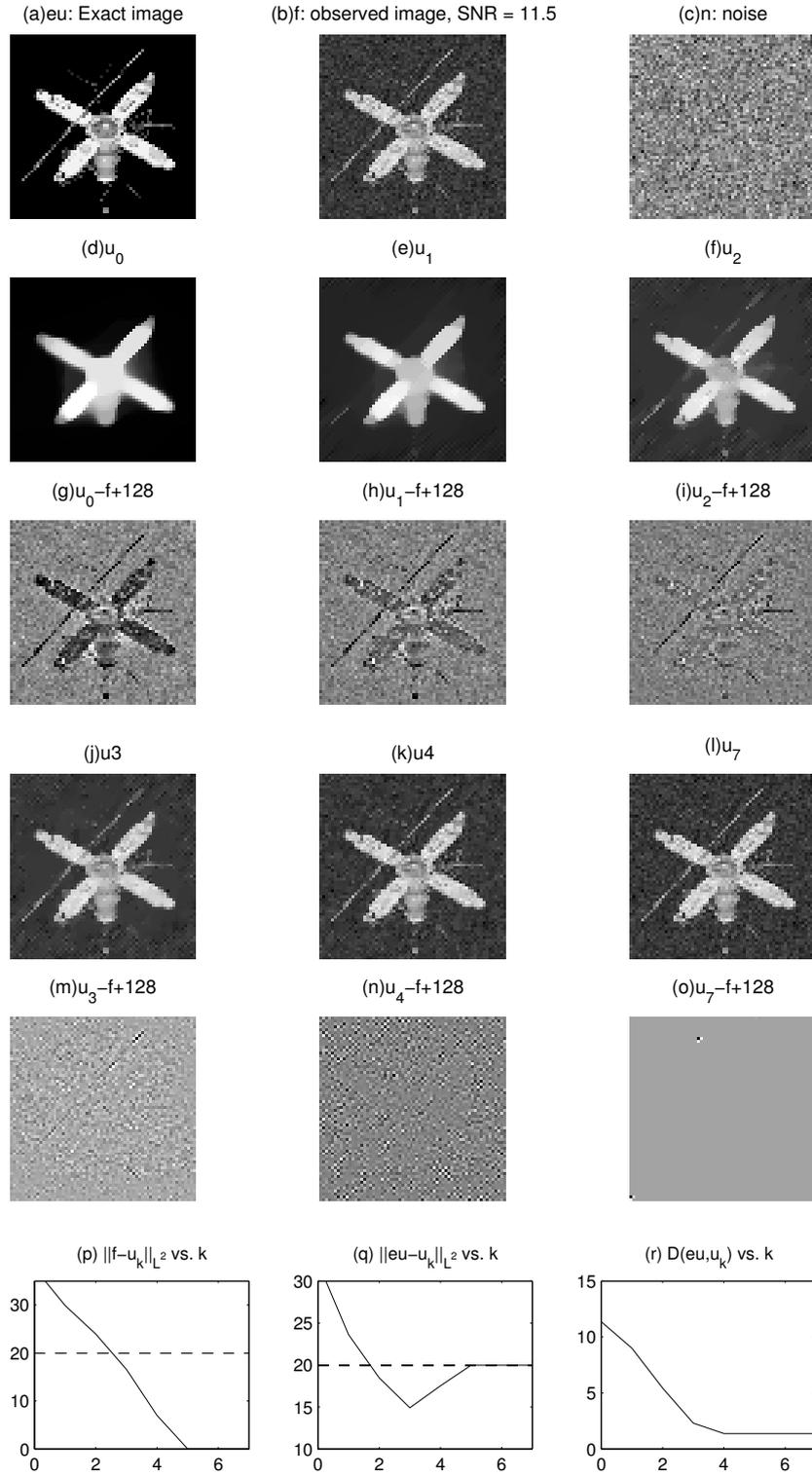


Figure 2.1: Satellite reconstructions with the new iterative total variation method, Gaussian noise $\delta = 20$, $\lambda = 10$.

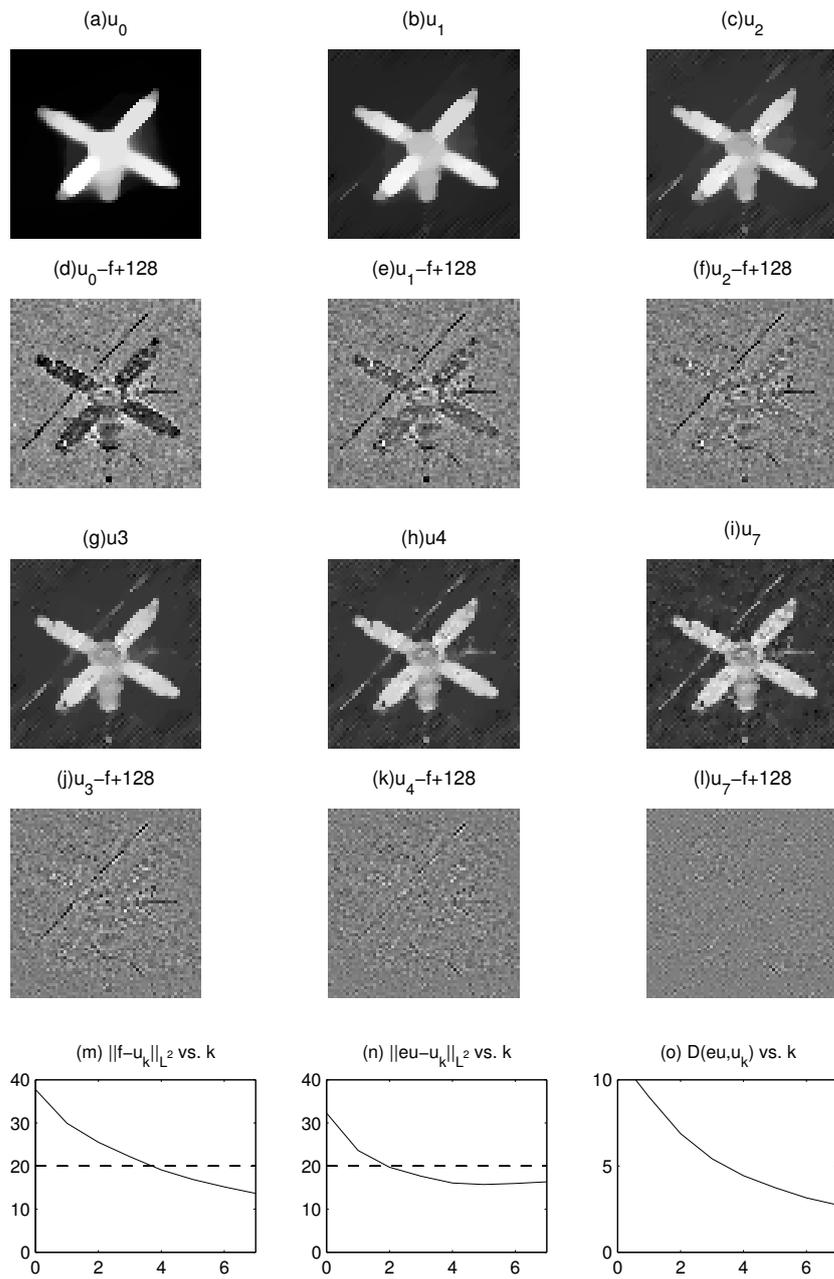


Figure 2.2: Satellite reconstructions with the original iterative total variation method, Gaussian noise $\delta = 20$, $\lambda = 10$.

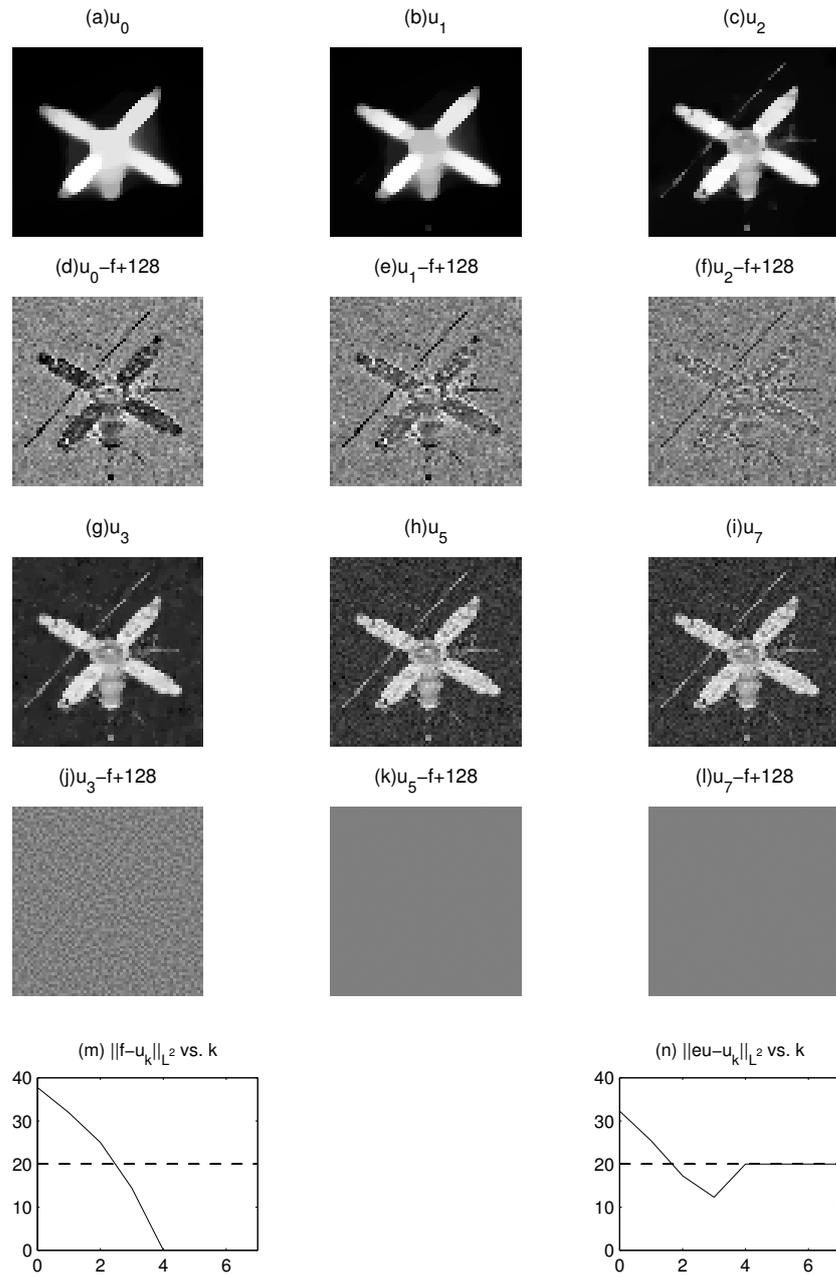


Figure 2.3: Satellite reconstructions with the $BV + L^2$ T-N-V method, Gaussian noise $\delta = 20$, $\lambda = 10$.

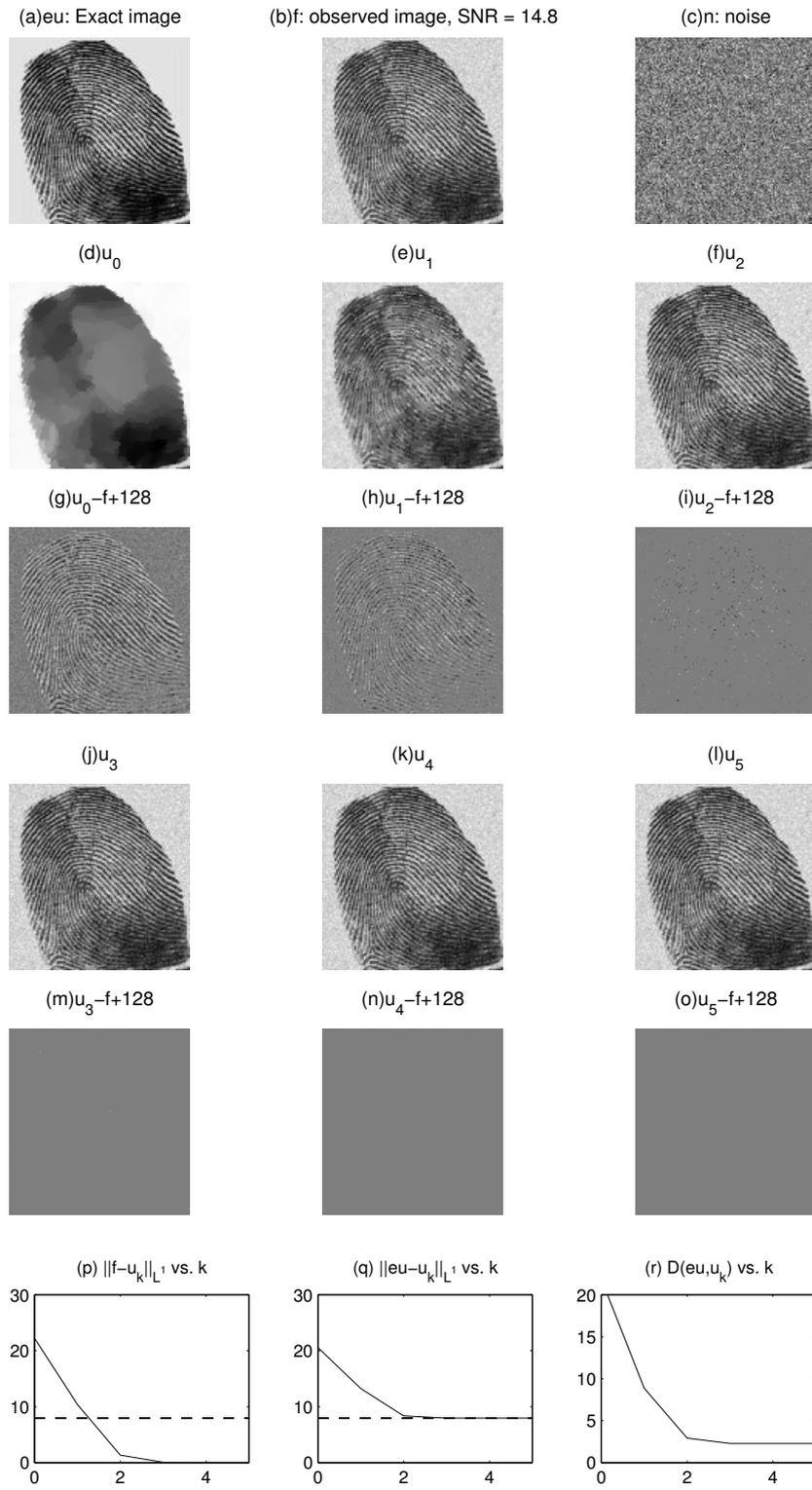


Figure 2.4: Restorations of a finger print obtained with the new iterative TV model, Gaussian noise, $\delta = 10$, $\lambda = 1$.

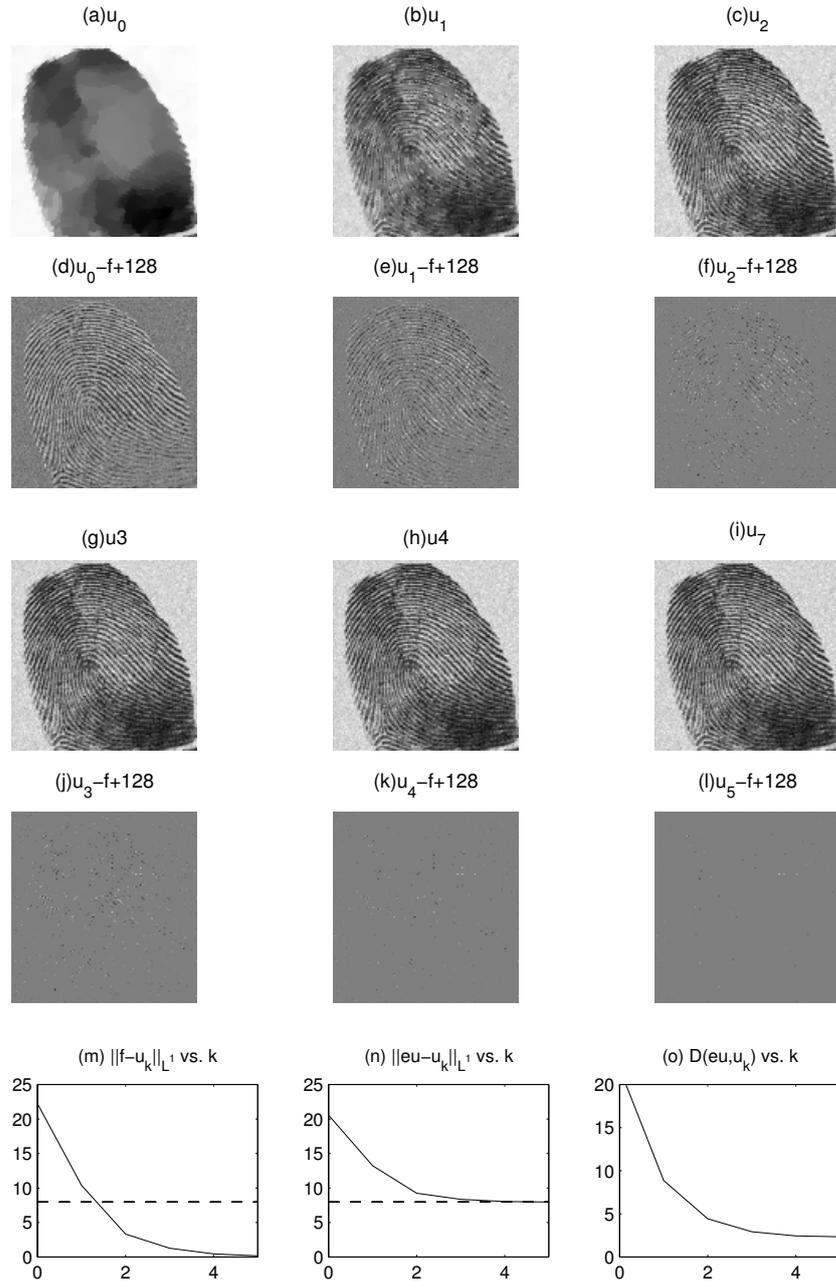


Figure 2.5: Restorations of a finger print obtained with the original iterative TV model, Gaussian noise, $\delta = 10$, $\lambda = 1$.

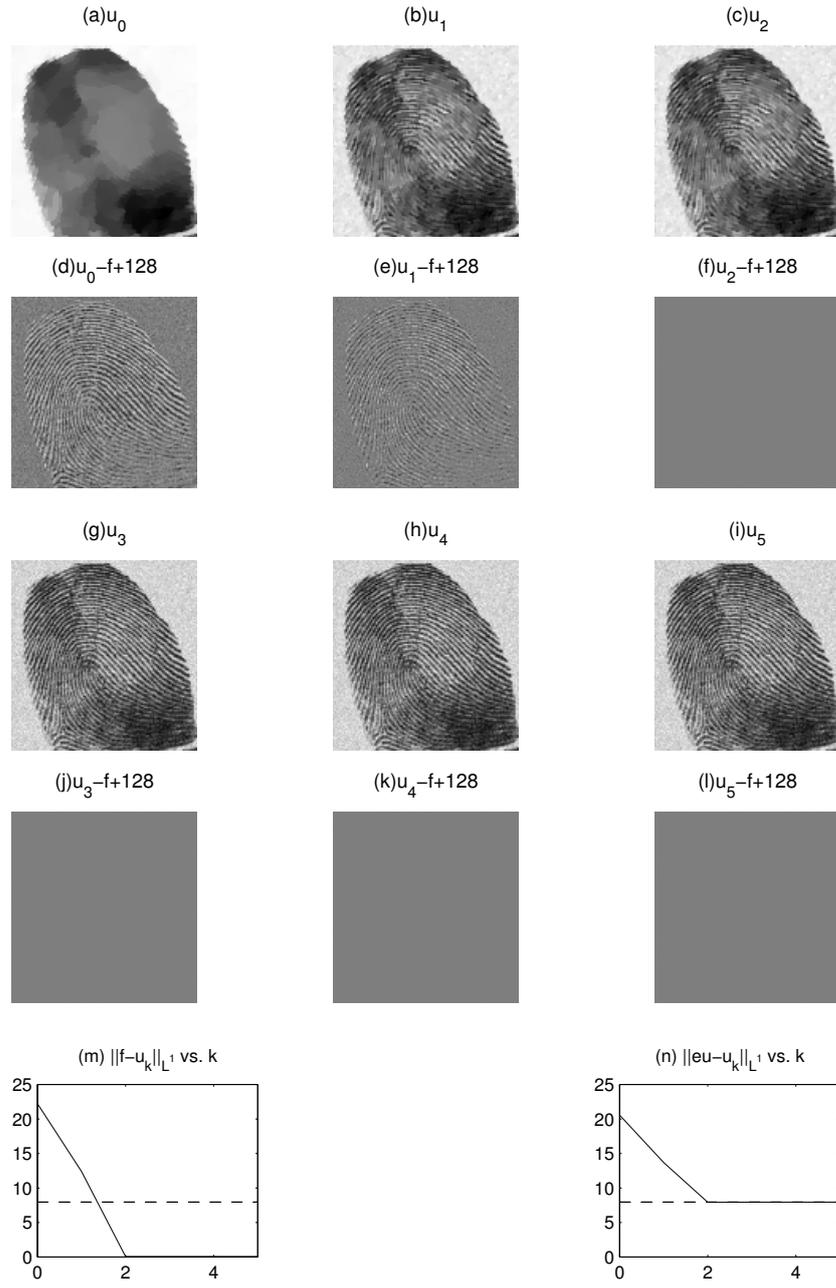


Figure 2.6: Restorations of a finger print obtained with the $BV + L^1$ T-N-V model, Gaussian noise, $\delta = 10$, $\lambda = 1$.

CHAPTER 3

Blind Deconvolution Using TV Regularization and Bregman Iteration

Instead of only minimizing u in (1.1), the blind deconvolution problem is viewed as a joint minimization problem with respect to both u and k . For example, You and Kaveh [YK96] proposed the following model

$$(u, k) = \arg \min_{u, k} \left\{ G(u, k) := \frac{1}{2} \|k * u - f\|_{L^2}^2 + \alpha_1 \|u\|_{H^1}^2 + \alpha_2 \|k\|_{H^1}^2 \right\}. \quad (3.1)$$

However, the model (3.1) cannot preserve sharp edges in the minimizer u . Moreover, as we will prove in Section 3.1.2, it also does not yield a local minimizer. In a later work [YK99] they, and independently Chan and Wong in [CW98], proposed a similar blind deconvolution method using *TV* regularization. The minimization problem was formulated as

$$(u, k) = \arg \min_{u, k} \left\{ F(u, k) := \frac{1}{2} \|k * u - f\|_{L^2}^2 + \alpha_1 \|u\|_{BV} + \alpha_2 \|k\|_{BV} \right\}. \quad (3.2)$$

The function $F(u, k)$ depends on both u and k . If we fix u or k and view F as a one-variable function depending only on k or u , respectively, then F is convex. But F is not jointly convex if it is viewed as a two-variable function. In [CW98], an alternating minimization (AM) algorithm was devised to solve the two Euler-Lagrange equations of u and k alternatively. The idea is to fix k first, solve u

from $F_u(u, k) = 0$, and then fix u , solve k from $F_k(u, k) = 0$, repeatedly:

$$\begin{cases} F_u(u, k) &= \hat{k} * (k * u - f) - \alpha_1 \nabla \cdot \frac{\nabla u}{|\nabla u|} = 0, \\ F_k(u, k) &= \hat{u} * (u * k - f) - \alpha_2 \nabla \cdot \frac{\nabla k}{|\nabla k|} = 0, \end{cases} \quad (3.3)$$

where \hat{k}, \hat{u} are the L^2 -conjugate of k and u respectively. The cosine preconditioned conjugate gradient method and the fixed-point method (cf. [CCW99]) were applied to find the solution of each Euler-Lagrange equation. Although it has been numerically shown that the AM algorithm converges quickly and recovers good image features after only a few AM iterations, time delay effects are noticed in the convergent process since u and k are not updated at the same time. Here we deal with (3.3) as a coupled system and evolve it by a time marching method, which is based on the gradient descent method. Moreover the results in [CW98, YK99, CW00] did not include the difficult case of Gaussian blur where we will handle below.

Also, as noticed in [CW98], if (u, k) is a solution, so are $(-u, -k)$, $(\frac{\alpha_2}{\alpha_1}k, \frac{\alpha_1}{\alpha_2}u)$ and $(u(x \pm c), k(x \mp c))$ as well, for any constant vector $c \in \Omega$. We will modify the energy function (3.2) by adding additional constraints for the kernel and the image, removing this source of nonuniqueness.

3.1 Blind Deconvolution Models

3.1.1 Existence of Minimizers

First, we will prove the existence of minimizers of $F(u, k)$ in (3.2). We will only present the sketch of the proof derived from the Kondrachov compactness theorem and lower semicontinuous property of the function $F(u, k)$.

Take a minimizing sequence function $\{u_j, k_j\}$ of (3.2), where $F(u_j, k_j)$ de-

creases to $\min_{u,k} F(u, k)$ which is denoted as M . Since $\{u_j\}$ has bounded BV norm and L^1 norm, by the Kondrachov compactness theorem, this sequence is precompact in $L^1(\Omega)$, which means that up to an extraction, there exists a function u that $u_j \rightarrow u$ in $L^1(\Omega)$. Then up to an extraction again, we can actually find $u_j \rightarrow u$ *a.e.* Also from lower semicontinuity of BV norm, we have $\liminf_{j \rightarrow \infty} |u_j|_{BV} \geq |u|_{BV}$.

Now among this subsequence of $\{u_j\}$, correspondingly consider for $\{k_j\}$, the same conclusion holds that up to an extraction $k_j \rightarrow k$ *a.e.* and $\liminf_{j \rightarrow \infty} |k_j|_{BV} \geq |k|_{BV}$.

Since we have $u_j \rightarrow u$ *a.e.* and $k_j \rightarrow k$ *a.e.*, applying the Fatou Lemma, we obtain

$$\begin{aligned} \liminf_{j \rightarrow \infty} \int (k_j * u_j - f)^2 dx &\geq \int \liminf_{j \rightarrow \infty} (k_j * u_j - f)^2 dx \\ &= \int (k * u - f)^2 dx, \end{aligned}$$

which means that (u, k) is one of the minimizers of (3.2).

Remark *If we replace the TV regularization term of u and k by $\int |\nabla u|^2 dx$ and $\int |\nabla k|^2 dx$ respectively, i.e., H^1 norm, then we still have the existence of minimizers from the Rellich compactness theorem and the fact that $L^2(\Omega) \subset L^1(\Omega)$ for bounded Ω .*

3.1.2 Analysis of the H^1 Regularization Blind Deconvolution Model

As we have mentioned, even though $F(u, f)$ is convex with respect to u and k respectively (if we fix one of them and consider F as a one-variable function), it is not generally convex for (u, k) . To study the behavior of (3.1) we will use Fourier Transform to solve the Euler-Lagrange equations. We will conclude from simple analysis that there are no local minimizers for the kernel k and the image

u , if $k(x) = k(-x)$.

We assume that there is a local minimizer (u, k) from (3.1), then (u, k) should be the solution of the Euler-Lagrange equations:

$$\begin{cases} G_u(u, k) := \widehat{k} * (k * u - f) - \alpha_1 \Delta u = 0 \\ G_k(u, k) := \widehat{u} * (u * k - f) - \alpha_2 \Delta k = 0. \end{cases}$$

These two equations can be solved by using Fourier Transform. Solve for $G_u(u, k)$ at ξ point, we have

$$-\alpha_1 \xi^2 F(u) = |F(k)|^2 F(u) - F(\widehat{k}) F(f),$$

i.e.

$$F(u) = \frac{F(\widehat{k}) F(f)}{|F(k)|^2 + \alpha_1 \xi^2}.$$

Similarly, we have for k

$$F(k) = \frac{F(\widehat{u}) F(f)}{|F(u)|^2 + \alpha_2 \xi^2}.$$

Plugging $F(u)$ into $F(k)$, using

$$F(\widehat{u}) = \widehat{F(u)} = \frac{F(k) F(\widehat{f})}{|F(k)|^2 + \alpha_1 \xi^2},$$

and denoting

$$X = |F(k)|^2 + \alpha_1 \xi^2,$$

then we obtain

$$F(k) \xi^2 (\alpha_2 X^2) = F(k) \xi^2 (\alpha_1 |F(f)|^2). \quad (3.4)$$

(3.4) means that, when $F(k)$ and ξ are not zero, $X^2 = \frac{\alpha_1}{\alpha_2} |F(f)|^2$.

Since f is the observed image, α_1 and α_2 are constants, then

$$|F(k)|^2 = -\alpha_1 \xi^2 + \sqrt{\frac{\alpha_1}{\alpha_2} |F(f)|^2},$$

or

$$|F(k)|^2 = -\alpha_1 \xi^2 - \sqrt{\frac{\alpha_1}{\alpha_2}} |F(f)|^2.$$

To guarantee $|F(k)|^2 \geq 0$, we can only choose the first expression when

$$\sqrt{\frac{\alpha_1}{\alpha_2}} |F(f)|^2 \geq \alpha_1 \xi^2,$$

which means ξ is bounded by above. However, $|\xi|$ could go to ∞ , which contradicts the assumption for $f \in L^2$. Hence, we proved that if there are minimizers then they are not local minimizers.

Remark *The above conclusion does not mean there are no local minimizers for (3.2).*

3.1.3 Analysis of TV Regularization Blind Deconvolution Model

Following (1.5) and (1.6) from Meyer's book (cf. [Mey01]), we obtain a similar result for the deblurring ROF model:

$$u = \arg \min_u \left\{ F(u) := \frac{1}{2} \|k * u - f\|_2^2 + \alpha_1 \|u\|_{BV} \right\}.$$

Lemma 2 *If the norm of $\widehat{k} * f$ in G does not exceed α_1 , then the above energy model gives a minimizer u such that*

$$u = 0, v = f.$$

Otherwise,

$$\|\widehat{k} * v\|_* = \alpha_1, \int (k * u)v = \alpha_1 \|u\|_{BV}.$$

Applying this lemma to our blind deconvolution TV regularization formula (3.2), we will also have the same conclusion for k .

Corollary 1 *If the norm of $\widehat{u} * f$ in G does not exceed α_2 , then (3.2) gives the minimization on k that*

$$k = 0, v = f.$$

Otherwise,

$$\|\widehat{u} * v\|_* = \alpha_2, \int (u * k)v = \alpha_2 \|k\|_{BV}.$$

Combining these two interesting facts, if (3.2) has a minimizer u and k and for sufficiently small α_1 and α_2 , we will get

$$\begin{cases} \int (k * u)v = \alpha_1 \|u\|_{BV}, \\ \int (u * k)v = \alpha_2 \|k\|_{BV}. \end{cases}$$

Using $\int (k * u)v = \int (u * k)v$, we have

$$\alpha_1 \|u\|_{BV} = \alpha_2 \|k\|_{BV}. \quad (3.5)$$

(3.5) will be a criteria which we use to choose α_1 and α_2 .

3.2 New Model and Numerical Algorithm

3.2.1 New Model

As we have noted above, the model (3.2) may not yield reasonable solutions due to the obvious fact that it does not have a unique solution. In order to obtain a unique reasonable solution, we include the following constraints on u and k :

$$\int k \, dx = 1, \quad \int u \, dx = w = \int f \, dx, \quad u(x), k(x) \geq 0 \quad (3.6)$$

This means that $(-u, -k)$ and $(\frac{\alpha_2}{\alpha_1}k, \frac{\alpha_1}{\alpha_2}u)$ do not satisfy as a solution. It is natural to impose these constraints to preserve mass conservation and nonnegativity for

the kernel and the signal. Thus our model is modified as the following:

$$\begin{aligned}
(u, k) = \arg \min_{u, k} F(u, k) := & \frac{1}{2} \|k * u - f\|_2^2 + \alpha_1 \|u\|_{TV} + \alpha_2 \|k\|_{TV} \\
& + \frac{\lambda_1}{2} (\int u \, dx - w)^2 + \frac{\lambda_2}{2} (\int |u| \, dx - w)^2 \\
& + \frac{\lambda_3}{2} (\int k \, dx - 1)^2 + \frac{\lambda_4}{2} (\int |k| \, dx - 1)^2.
\end{aligned} \tag{3.7}$$

For $\lambda_1, \lambda_2, \lambda_3$ and λ_4 large, the constraints of (3.6) are approximately satisfied. We still need to prevent $(u(x \pm c), k(x \mp c))$ from being a solution. Thus we continue to assume the symmetry of the kernel k . Due to the fact that $\widehat{u} * (u * k)$ is symmetric if k is symmetric, we state that if we solve the Euler-Lagrange equations (3.3) with a symmetric kernel as the initial guess, the kernel won't deviate too much when the noise n is not too large relative to the noisy and blurry data f .

We have no proof yet that (3.7) has a unique solution when k is symmetric, but in practice (3.7) yields a great improvement in quality of image and stability and convergence of the algorithm. The idea is just to preserve the mass and the positivity of the kernel and the image. Also numerically this model is very easy to implement since the variational derivatives of these four additional terms are constants.

In general, (3.7) yields very accurate solutions. In [CW98], these constraints were imposed on (u, k) by modifying the results after every AM iteration from solving (3.3). Thus after every AM iteration the accuracy was degraded and some information was lost in the procedure.

3.2.2 Numerical Algorithm

It was proved that for the H^1 blind deconvolution model (3.1) the AM procedure converges globally (cf. [CW98]), but the solution depends on the initial guess(cf.

[CW00]). Also as we have discussed before, because we are not updating u and k simultaneously we observe time delay effects in the convergent process. Furthermore, the AM algorithm fails to improve pure Gaussian blurred image efficiently.

Therefore we propose a time marching method to update u and k after every small time step, based on a straightforward gradient descent idea. Suppose we have k^n and u^n , plugging them into (3.8), gives us k^{n+1} ; plugging them into (3.9), gives us u^{n+1} . Thus we obtain u^{n+1} and k^{n+1} .

- Solve for k^{n+1}

$$\begin{aligned} \frac{k^{n+1}-k^n}{dt} = & -\widehat{u}^n * (u^n * k^n - f) + \alpha_2 \nabla \cdot \frac{\nabla k^n}{|\nabla k^n|} \\ & -\lambda_3 (\int |k^n| dx - 1) - \lambda_4 (\int |k^n| dx - 1) \operatorname{sgn}(k^n) \end{aligned} \quad (3.8)$$

- Solve for u^{n+1}

$$\begin{aligned} \frac{u^{n+1}-u^n}{dt} = & -\widehat{k}^n * (k^n * u^n - f) + \alpha_1 \nabla \cdot \frac{\nabla u^n}{|\nabla u^n|} \\ & -\lambda_1 (\int |u^n| dx - w) - \lambda_2 (\int |u^n| dx - w) \operatorname{sgn}(u^n) \end{aligned} \quad (3.9)$$

3.2.3 Application of Bregman Distance

Following Section 1.3.2, we will only apply the Bregman iteration on the image u . The general procedure of our numerical algorithm using the Bregman iteration is this: starting from a guess for the image u and the kernel k which are the observed image f and δ function correspondingly, we solve the Euler-Lagrange equation (3.8) and (3.9), and denote the solution as k_1 and u_{10} . Fixing the kernel k_1 , we apply the Bregman iteration on u m times, and write the result as u_{1m} .

This is called a round. We expect this to give a better image than u_{10} because $\|u_{1m} - f\|_2^2$ will always decrease with m (cf. [OBG05]). Furthermore, $\|u_{1m} - f\|_2^2$ converges to zero for a pure deblurring problem, and for noisy data, the Bregman distance between u_{1m} and the true solution \tilde{u} is decreasing as long as the residual $\|u_{1m} - f\|_2^2$ lies above the noise level. These results were obtained in [OBG05].

We next go back to the Euler-Lagrange equation (3.8) and (3.9) with what we expect to be a better guess k_1 and u_{1m} , and repeat the iteration for another round. A better kernel than k_1 is expected. We repeat these rounds and stop as soon as the recovered image becomes noisy.

3.3 Explicit Numerical Scheme for the 2D Model

We can write our 2D model as:

$$\left\{ \begin{array}{l} u_t = -\widehat{k} * (k * u - f) + \alpha_1 \frac{u_{xx}(u_y^2 + \beta) - 2u_{xy}u_xu_y + u_{yy}(u_x^2 + \beta)}{(u_x^2 + u_y^2 + \beta)^{\frac{3}{2}}} \\ \quad + \lambda_1 (\int |u| dx - w) + \lambda_2 \operatorname{sgn}(u) (\int |u| dx - w) \\ k_t = -\widehat{u} * (u * k - f) + \alpha_2 \frac{k_{xx}(k_y^2 + \beta) - 2k_{xy}k_xk_y + k_{yy}(k_x^2 + \beta)}{(k_x^2 + k_y^2 + \beta)^{\frac{3}{2}}} \\ \quad + \lambda_3 (\int |k| dx - 1) + \lambda_4 \operatorname{sgn}(k) (\int |k| dx - 1). \end{array} \right.$$

and use f and the delta function as initial guesses for u and k respectively. Homogeneous Neumann boundary condition is used for f . The parameter $\beta > 0$ is the regularization parameter for the sgn function. It is chosen to be very small for a pure deblurring problem. We usually pick it as 0.01, and change it to $1.0e-6$ after the recovered image and kernel are improved. This usually happens after a couple rounds of the procedure.

We consider u_{ij}^n and k_{ij}^n the approximations to $u(x_i, y_j, t_n)$ and $k(x_i, y_j, t_n)$,

respectively where $x_i = idx$, $y_j = jdy$, $t_n = ndt$.

We denote by

$$v_{ij}^n = [\widehat{k}^n * (k^n * u^n - f)]_{ij} + \lambda_1 \left[\int |u^n| dx - w \right] + \lambda_2 \operatorname{sgn}(u_{ij}^n) \left[\int |u^n| dx - w \right]$$

and

$$w_{ij}^n = [\widehat{u}^n * (u^n * k^n - f)]_{ij} + \lambda_3 \left[\int |k^n| dx - 1 \right] + \lambda_4 \operatorname{sgn}(k_{ij}^n) \left[\int |k| dx - 1 \right],$$

where the convolutions are computed using the discrete cosine transform (DCT) to enforce the homogeneous Neumann boundary condition. The explicit scheme for the simultaneous evolution of the signal and the kernel reads as follows:

$$\frac{u_{ij}^{n+1} - u_{ij}^n}{dt} = -v_{ij}^n + \alpha_1 s_{ij}^n(u) \quad (3.10)$$

$$\frac{k_{ij}^{n+1} - k_{ij}^n}{dt} = -w_{ij}^n + \alpha_2 s_{ij}^n(k). \quad (3.11)$$

And the second order term $s_{ij}^n(u)$, (resp. $s_{ij}^n(k)$) is computed by means of

$$s_{ij}^n(u) := \frac{g_{ij}^{xx}(u)(g_{ij}^y(u)^2 + \beta) - 2g_{ij}^{xy}(u)g_{ij}^x(u)g_{ij}^y(u) + g_{ij}^{yy}(u)(g_{ij}^x(u)^2 + \beta)}{(g_{ij}^x(u)^2 + g_{ij}^y(u)^2 + \beta)^{\frac{3}{2}}},$$

where

$$\begin{aligned} g_{ij}^x(u) &= \frac{u_{i+1,j}^n - u_{i-1,j}^n}{2dx}, \\ g_{ij}^y(u) &= \frac{u_{i,j+1}^n - u_{i,j-1}^n}{2dy}, \\ g_{ij}^{xx}(u) &= \frac{u_{i+1,j}^n - 2u_{ij}^n + u_{i-1,j}^n}{dx^2}, \\ g_{ij}^{yy}(u) &= \frac{u_{i,j+1}^n - 2u_{ij}^n + u_{i,j-1}^n}{dy^2}, \end{aligned}$$

$$g_{ij}^{xy}(u) = \frac{u_{i+1,j+1}^n - u_{i-1,j+1}^n - u_{i+1,j-1}^n + u_{i-1,j-1}^n}{4dxdy}.$$

The expressions for $s_{ij}^n(k)$ are the same using k instead of u .

To speed up the convergence of k to the true kernel solution (which is a Gaussian kernel in our examples), we will compute the following equation in place of (3.10). This model is based on level set motion and was formulated by Marquina and Osher in [MO01]. It is very efficient and fast. Our model for k is expressed in terms of explicit partial derivatives as

$$k_t = -\sqrt{k_x^2 + k_y^2} \{ \widehat{u}^n * (u^n * k^n - f) + \alpha_2 \frac{k_{xx}(k_y^2 + \beta) - 2k_{xy}k_xk_y + k_{yy}(k_x^2 + \beta)}{(k_x^2 + k_y^2 + \beta)^{3/2}} \\ + \lambda_3 (\int |k| dx - 1) + \lambda_4 \operatorname{sgn}(k) (\int |k| dx - 1) \}.$$

The first order scheme reads as follows:

$$\frac{k_{ij}^{n+1} - k_{ij}^n}{dt} = -\sqrt{kg_{ij}^x{}^2 + kg_{ij}^y{}^2} (w_{ij}^n + \alpha_2 s_{ij}^n(k)),$$

where kg_{ij}^x and kg_{ij}^y are defined as follows from Rouy and Tourin (cf. [RT92]),

$$kg_{ij}^x{}^2 = \max(\max(\frac{k_{ij}^n - k_{i-1,j}^n}{dx}, 0)^2, \min(\frac{k_{i+1,j}^n - k_{i,j}^n}{dx}, 0)^2) \\ kg_{ij}^y{}^2 = \max(\max(\frac{k_{ij}^n - k_{i,j-1}^n}{dy}, 0)^2, \min(\frac{k_{i,j+1}^n - k_{i,j}^n}{dy}, 0)^2)$$

if $w_{ij}^n > 0$, and

$$kg_{ij}^x{}^2 = \max(\min(\frac{k_{ij}^n - k_{i-1,j}^n}{dx}, 0)^2, \max(\frac{k_{i+1,j}^n - k_{i,j}^n}{dx}, 0)^2) \\ kg_{ij}^y{}^2 = \max(\min(\frac{k_{ij}^n - k_{i,j-1}^n}{dy}, 0)^2, \max(\frac{k_{i,j+1}^n - k_{i,j}^n}{dy}, 0)^2)$$

if $w_{ij}^n < 0$. For more details, please see [MO01].

The speed up procedure is only applied to the kernel evolution equation. The signal evolution equation (3.11) remains unchanged. Since we start with the blurred and noisy signal f , we need to increase the total variation of the signal to restore u . If we use the speed up procedure for the signal it is impossible to recover local extrema because of the maximum principle.

3.4 Numerical Results

In this section we present some of the results we have obtained using our blind deconvolution TV regularization model with the iterated regularization procedure. We also compare this with the results obtained using the H^1 norm of k instead of the BV norm of k .

We are given a scaled noisy and blurred gray image f that takes value from $[0,1]$. We take the image f as the initial guess for u . We take a δ function as the initial guess for the kernel k . Inside one round described above, we run 1000 iterations to solve the Euler-Lagrange equations for k and u , and then another 1000 iterations to update u by the Bregman iteration. For a 64 by 64 matrix image, this takes 90 seconds for a computer with a Mobile Pentium 4 and 1.8Ghz processor. Generally the full algorithm converges after 3 rounds, that means all the work takes less than five minutes.

We usually pick $\lambda_1 = \lambda_2 = 0.001$ and $\lambda_3 = \lambda_4 = 10$ to enforce the positivity of the kernel and the image and preserve their mass. Also, from (3.5), we know $\alpha_1 \|u\|_{BV} = \alpha_2 \|k\|_{BV}$, i.e., the choice of α_1 and α_2 are dependent on the proportion of the TV norm of the image and the kernel. Of course, we do not know the value of both norms until we try some preliminary experiments.

We use white Gaussian noise $\eta \sim \mathcal{N}(0, \sigma^2)$, $\|\eta\|_{L^2} \approx \sigma$.

$$SNR := 20 * \log_{10} \left(\frac{\|f - \bar{f}\|_{L^2}}{\|\eta - \bar{\eta}\|_{L^2}} \right)$$

is the signal-to-noise ratio, measured in decibels(dB). \bar{f} and $\bar{\eta}$ are the means of f and η over Ω .

Figure 3.1 with $SNR = 1.47$ from (a) to (i) displays the results of our numerical algorithm including the Bregman iteration. We have run this experiment for three rounds. In every round we only apply the Bregman iteration to u once,

thus, including the kernel, we obtain nine pictures. We observe that the iterated refinement procedure gives us more details of the image. It is sometimes the case that in the earlier rounds it also brings back some noise. However, as we do more rounds the recovered kernel gets better.

Figure 3.2 shows us how to choose parameters α_1, α_2 . The smaller the parameter, the more details we have of the picture in the very early rounds. This sometimes unfortunately includes noise, see (g)~(i). The bigger the parameter, the more we smooth the image u and obtain less detail, see (d) ~ (f). Since the main purpose of applying the Bregman iteration is to bring back details without noise, the bigger parameter is preferred; however, one must remember that the noise will eventually come back in the later rounds, for proof and examples see [OBG05].

Figure 3.3 is the noise free case. From [OBG05] we know that the Bregman distance $D(u_m, \tilde{u})$ converges to zero, and we do see this. We have recovered more and more details as we repeatedly apply the iterated refined procedure. Comparing (e), (f) and (g), we see that (g) from deconvolution with the Bregman iteration is the best, but (e) from blind deconvolution with the Bregman iteration is as good as (f) from deconvolution without the Bregman iteration.

Figure 3.4 are all obtained from using blind deconvolution with the iterated refinement method. However, (c) and (d) are obtained from using the exact kernel as the initial kernel. Comparing (c) and (d) with (a) and (b) where the δ function is used as the initial kernel, we do not observe very remarkable differences. It is not surprising to see that the algorithm converges faster for a better initial kernel k . It is possible that this model, with the symmetry assumption, converges globally, and the result does not depend on the initial guess.

In Figure 3.5, we show the results obtained by regularizing the square of the

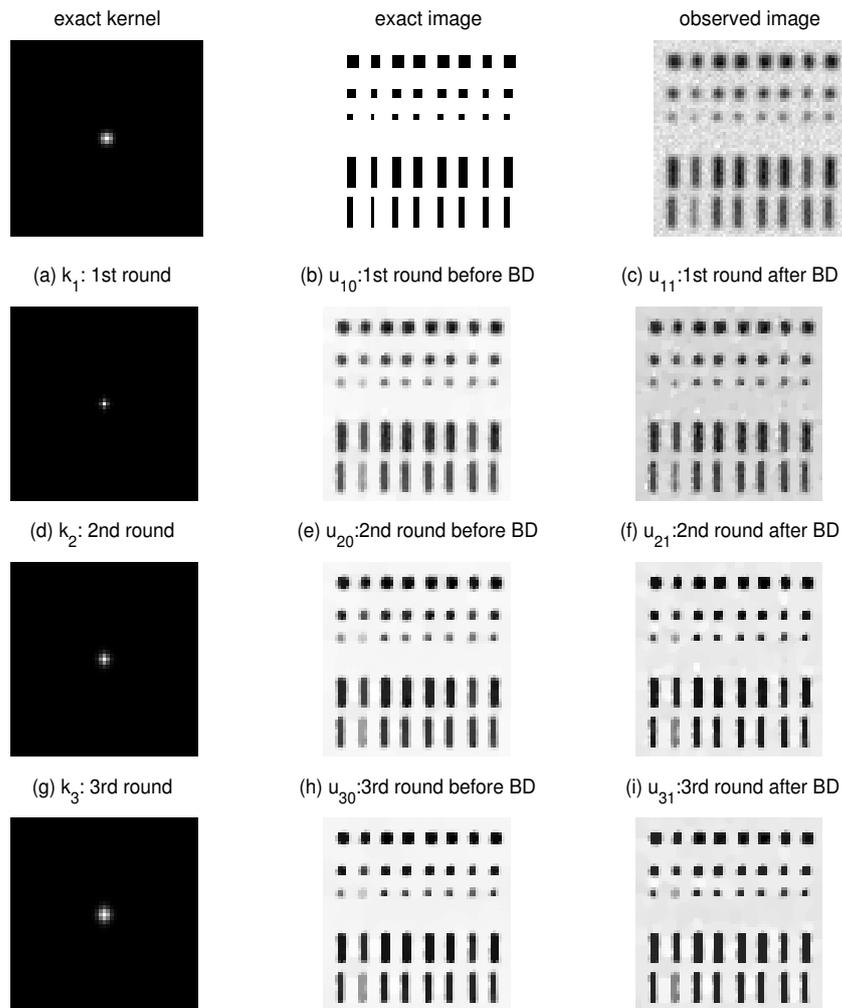


Figure 3.1: Example of numerical algorithm procedure, $\text{SNR} = 1.47$. We observe that the recovered kernel does improve after every round, and the Bregman iteration gives us more details of the image. However, the Bregman iteration brings back some noise in the first round.

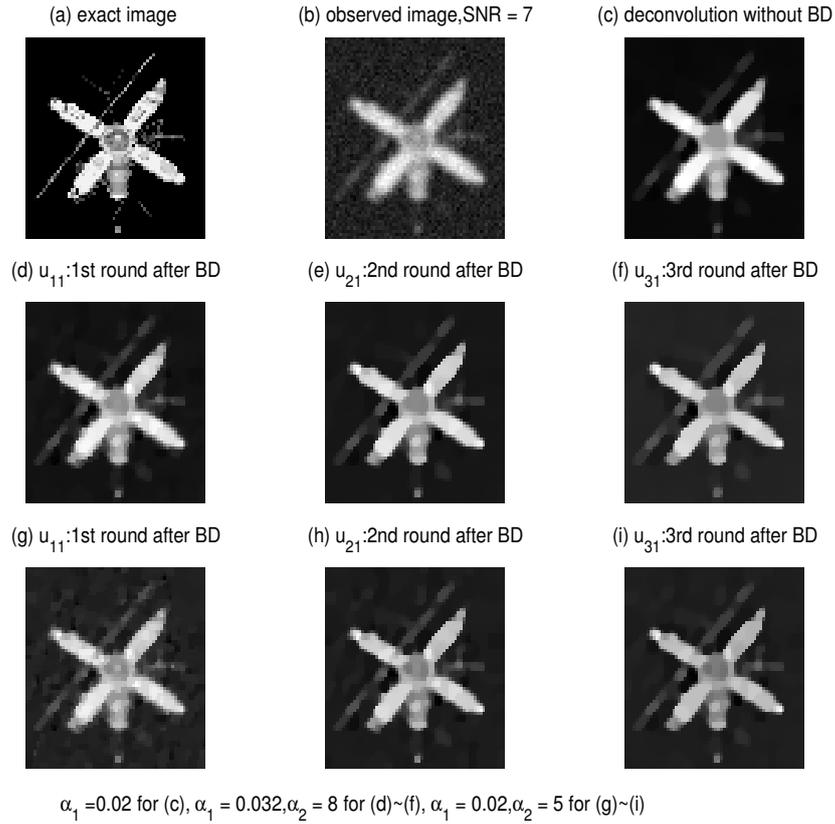


Figure 3.2: Compare different α_1, α_2 . Notice that (d)~(f) with fewer details is smoother than (g)~(i).

H^1 norm of the kernel instead of using the TV norm of kernel. The observed image is taken from Figure 3.4 with $SNR = 2.6$. Everything else is kept the same, including the constraint terms and our algorithm. Figures (a) and (b) are from blind deconvolution without the Bregman iteration. Figure (c) and (d) are from blind deconvolution with the Bregman iteration. Even though we can not tell the differences between (b) and (d) by eye, the L^2 norm of the distance from the exact image u_{exact} shows us the effectiveness of the Bregman iteration. Comparing the results with Figure 3.4, we see that using the H^1 norm of the kernel is almost as good as using the TV norm of the kernel. However there is a

small disadvantage in using the H^1 norm of the kernel, which is that we do not know how to choose good parameters for α_1 and α_2 .

Figure 3.6 displays the results of applying our new model with the Bregman iteration to a blurred and noisy galaxy picture. The picture itself is complicated, because the bright stars could be treated as noise. The result from using blind deconvolution with the iterative regularization method (b), shows that we still recover a good image as in (a), where we know the exact kernel. (c) and (d) are presented to show again, that our model with the Bregman iteration does not depend on the initial guess, and using the H^1 norm of the kernel k can recover a good picture as well if we know how to choose the value of the parameters α_1 and α_2 .

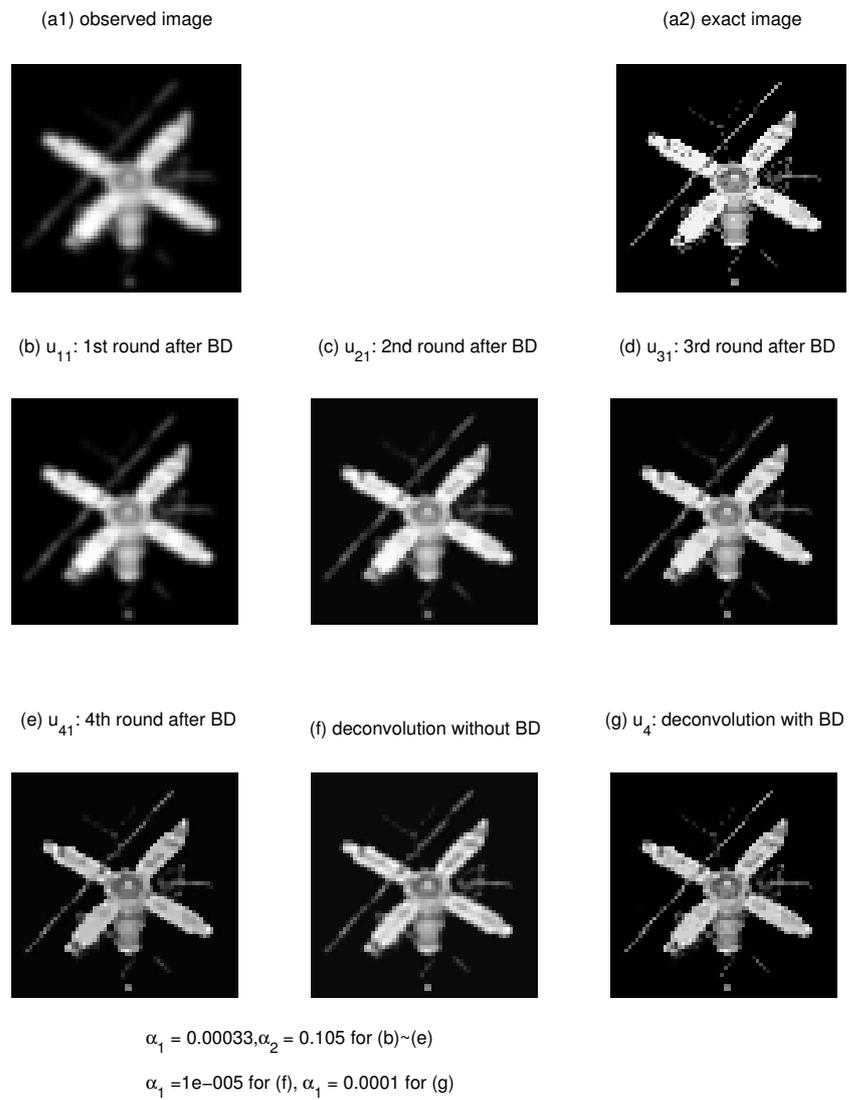


Figure 3.3: Noise free case. from (b) to (e), we see that more details are restored.

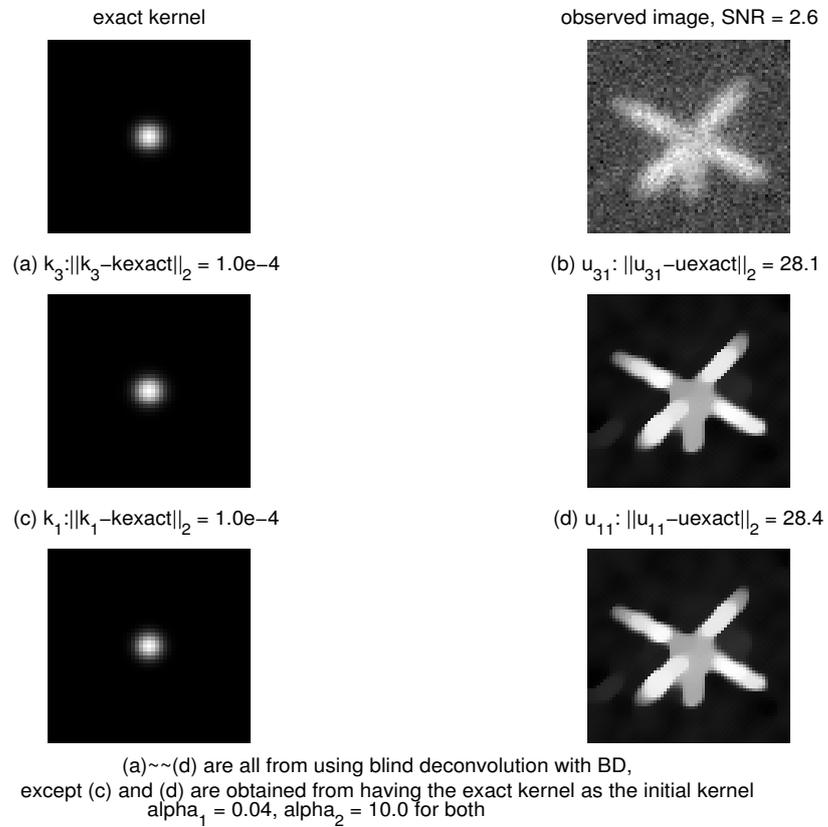


Figure 3.4: Comparison 1: We show that a better initial guess for the kernel does not improve the result.

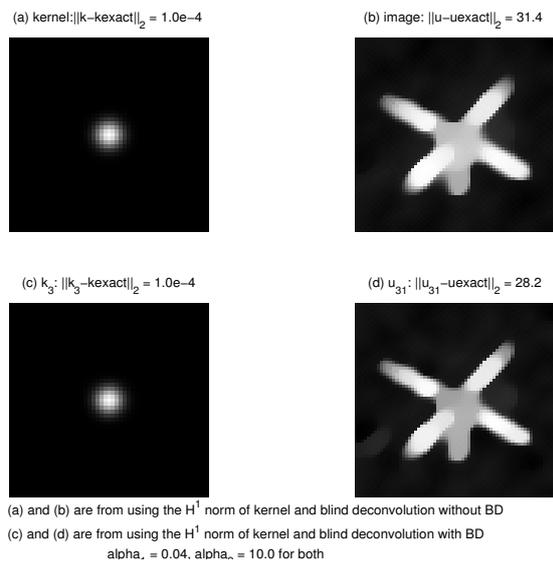


Figure 3.5: Comparison 2: We use the H^1 norm of the kernel. The result is comparable to that using the TV norm of kernel.

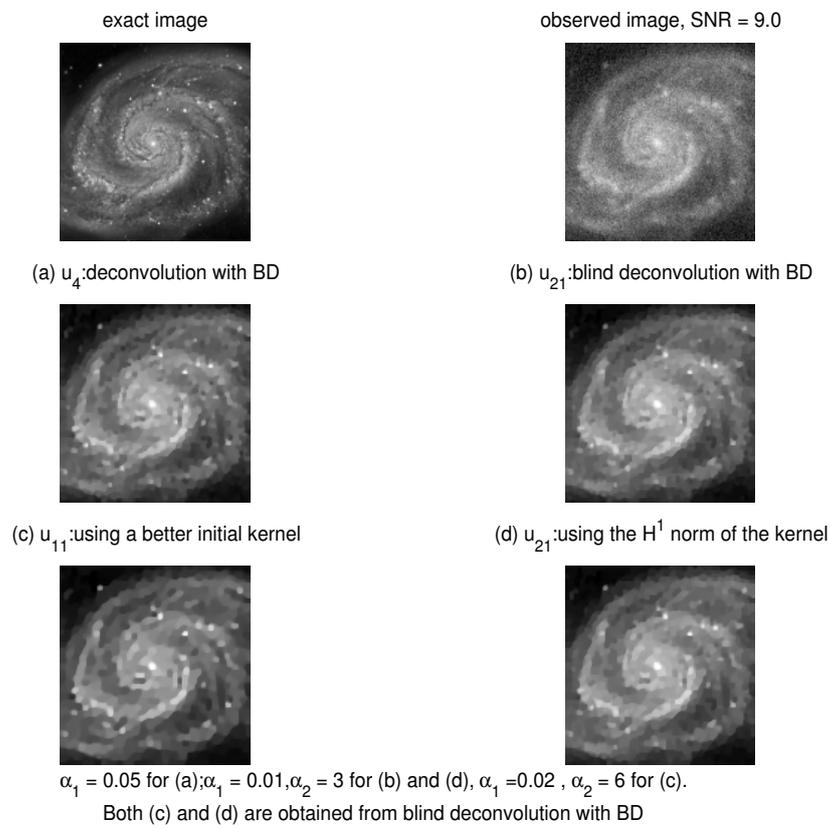


Figure 3.6: Comparison 3.

CHAPTER 4

MR Image Reconstruction from Sparse Radial Samples Using Bregman Iteration and Nonlinear Inverse Scale Space Methods

In a recent paper [CRT04] by Candes et al., a general problem was proposed to address the problem of image artifacts and/or low SNR resulted from MR image reconstruction from sparsely sampled k -space data. They consider for a discrete complex signal n of length N and a randomly chosen set of frequencies Ω of mean size τN with $\tau < 1$, whether it is possible to reconstruct n from the partial knowledge of its Fourier coefficients on the set Ω . The answer is yes,

Proposition 4 *if the signal n obeys*

$$\#\{t, n(t) \neq 0\} \leq \alpha(M) \cdot (\log N)^{-1} \cdot \#\Omega, \quad \forall M \in \mathbb{N}$$

then with the probability at least $1 - O(N^{-M})$, n can be reconstructed exactly as the solution to the L^1 minimization problem

$$\min_m \sum_{t=0}^{N-1} |m(t)|, \quad \text{s.t. } \hat{m}(\omega) = \hat{n}(\omega) \text{ for all } \omega \in \Omega,$$

where m is the reconstructed signal.

The assumption of the signal n being sparse is the key point in the proof of the above proposition (cf. [CRT04]). However, in practice many applications

deal with more complicated signals rather than sparse ones. Therefore, Candes et al. in [CR05] choose to solve an extended L^1 minimization problem by assuming that the signal n has a sparse representation. The L^1 minimization problem is reformulated as

$$\min_m \|\psi(m)\|_1 \text{ s.t. } Fm = y, \quad (4.1)$$

where m is the reconstructed signal/image, ψ transforms the image m into a sparse representation, F is an undersampled Fourier matrix which satisfies $Fm = \hat{m}|_\Omega$, and $y = \hat{n}|_\Omega$ is the undersampled measured k-space data. There is no theoretical proof on the extended L^1 minimization problem yet. But the numerical solutions of (4.1) solved from numerical phantoms (cf. [CRT04]) or from practical signals (cf. [CR05, LLD05]) have shown that the reconstructed MR image quality surpasses the conventional techniques.

Based on their work, we propose here to solve (4.1) with an iterative refinement method (cf. [OBG05, HMO05, HBO06, XO06, CHF06]). Furthermore, a time-continuous inverse scale-space formulation (cf. [BGO06]) is applied to solve (4.1). In the context of processing noisy images, scale space methods which start with the noisy image can separate the noise or so called small scales from large scales by smoothing small scale features faster than large ones. However, inverse scale space methods start with the image $u_0 = 0$ and approach the noisy image f as time increases, with large scales converge faster than small ones. Thus, if the method is stopped at a suitable time, large scale features may already be incorporated into the reconstruction, while small scale features (including the noise) are still missing. Therefore, the inverse scale space method may be treated as the time continuous version of the iterative refinement method.

4.1 Prerequisite to Solve the Model (4.1)

In this section, we describe related tools in more details to solve the given problem (4.1). This includes the use of sparse transformations and the implementation of maintaining the constraint condition. Following the work in [CR05, LLD05], the total variation regularization or the wavelet transform are used for the MR image reconstruction from raw k -space data of a numerical phantom.

4.1.1 The TV regularization

The total variation regularization or the so-called bounded variation (BV) norm was first introduced in [ROF92] in the context of image denoising. Afterwards this approach has been generalized to other image processing problems, such as deconvolution (cf. [CW98]), decomposition (cf. [CE05, HBO06, YGO05]), computer tomography (cf. [ZF05]), etc. It is proved to be particularly relevant in recovering piecewise smooth functions without smoothing sharp discontinuities while also being a sparse transformation for piecewise constant functions.

Normally, the bounded variation norm of an two dimensional real image m is defined as

$$\|\psi(m)\|_1 = \|m\|_{BV} := \sum_{i,j} \sqrt{|\nabla_x m(i,j)|^2 + |\nabla_y m(i,j)|^2}. \quad (4.2)$$

However, our reconstructed image m from (4.1) equals the magnitude of a complex image (m_r, m_i) from the inverse Fourier transform of the undersampled data y , i.e., $m = \sqrt{m_r^2 + m_i^2}$. We can not just simply define the BV norm of image m as (4.2) due to the non-convexity of this formulation. Instead, we define the BV

norm of m as

$$\|m\|_{BV} = \sum_{i,j} \sqrt{|\nabla_x m_r(i,j)|^2 + |\nabla_x m_i(i,j)|^2 + |\nabla_y m_r(i,j)|^2 + |\nabla_y m_i(i,j)|^2}. \quad (4.3)$$

Another alternatives is also used. The basic idea is similar to (4.3), that is, to separate the real and imaginary part,

$$\|m\|_{BV} = \sum_{i,j} \sqrt{|\nabla_x m_r(i,j)|^2 + |\nabla_y m_r(i,j)|^2} + \sqrt{|\nabla_x m_i(i,j)|^2 + |\nabla_y m_i(i,j)|^2}. \quad (4.4)$$

The minimization of (4.3) and (4.4) subject to the constraint condition $Fm = y$ is guaranteed to converge to a point according to [Bre65]. Indeed, our experiments show that they both produce similar and reasonable results.

4.1.2 The Wavelet Transform

The wavelet transform is tremendously popular in the signal and image processing communities, due in large part to its ability to provide parsimonious representations for signals that are smooth away from isolated discontinuities. The wavelet transform is widely used in image processing, and it generally represents a sparse representation of the signal/image g . The result from the wavelet transform is compared with the result from the total variation regularization subject to the constraint condition in [LLD05]. The wavelet transform does a fairly good job although not as good as the BV norm.

In the following we will first define the L^1 /wavelet norm $\|\psi(g)\|_1$ and then derive the sub-gradient of the L^1 /wavelet norm. For the reason of simplicity, we consider a 1D signal g of length $N = 2^J$. At the level L of the wavelet transform,

the signal g is decomposed as

$$g = \sum_{p=0}^{2^L-1} \alpha_{L,p} \varphi_{L,p} + \sum_{q=L+1}^J \sum_{p=0}^{2^{q-1}-1} \alpha_{q,p} \phi_{q,p}, \quad (4.5)$$

where the $\varphi_{L,p}$ are scaling functions, the $\phi_{q,p}$ are oscillatory wavelets, see [Mey93, BGG98, Mal99]. (Note, when $L = J$, for Haar wavelets, $\alpha_{L,p}$ is just the pixel value.) Then the L^1 /wavelet norm $\|\psi(g)\|_1$ is defined as

$$\|\psi(g)\|_1 = \|g\|_W := \sum_{p=0}^{2^L-1} |\alpha_{L,p}| + \sum_{q=L+1}^J \sum_{p=0}^{2^{q-1}-1} |\alpha_{q,p}|. \quad (4.6)$$

From (4.5), for the signal g and a variable signal h , we assume that

$$\begin{cases} g = \sum_{p=0}^{2^L-1} \alpha_{L,p} \varphi_{L,p} + \sum_{q=L+1}^J \sum_{p=0}^{2^{q-1}-1} \alpha_{q,p} \phi_{q,p}, \\ h = \sum_{p=0}^{2^L-1} \beta_{L,p} \varphi_{L,p} + \sum_{q=L+1}^J \sum_{p=0}^{2^{q-1}-1} \beta_{q,p} \phi_{q,p}. \end{cases}$$

The sub-gradient of $\|g\|_W$ is defined as:

$$\begin{aligned} & \lim_{s \rightarrow 0} \frac{1}{s} (\|g + sh\|_W - \|g\|_W) \\ &= \lim_{s \rightarrow 0} \frac{1}{s} \left(\sum_{p,q} |\alpha_{p,q} + s\beta_{p,q}| - \sum_{p,q} |\alpha_{p,q}| \right) \\ &= \sum_{p,q} \operatorname{sgn}(\alpha_{p,q}) \beta_{p,q} \\ &= \left(\sum_{p=0}^{2^L-1} \operatorname{sgn}(\alpha_{L,p}) \varphi_{L,p}, \sum_{p=0}^{2^L-1} \beta_{L,p} \varphi_{L,p} \right) \\ & \quad + \left(\sum_{q=L+1}^J \sum_{p=0}^{2^{q-1}-1} \operatorname{sgn}(\alpha_{q,p}) \phi_{q,p}, \sum_{q=L+1}^J \sum_{p=0}^{2^{q-1}-1} \beta_{q,p} \phi_{q,p} \right) \\ &= \left(\sum_{p=0}^{2^L-1} \operatorname{sgn}(\alpha_{L,p}) \varphi_{L,p} + \sum_{q=L+1}^J \sum_{p=0}^{2^{q-1}-1} \operatorname{sgn}(\alpha_{q,p}) \phi_{q,p}, h \right) \end{aligned} \quad (4.7)$$

The last two steps are valid because the scaling functions φ and the wavelet functions ϕ are an orthonormal basis. Thus, the sub-gradient of $\|g\|_W$ is

$$\sum_{p=0}^{2^L-1} \operatorname{sgn}(\alpha_{L,p}) \varphi_{L,p} + \sum_{q=L+1}^J \sum_{p=0}^{2^{q-1}-1} \operatorname{sgn}(\alpha_{q,p}) \phi_{q,p},$$

which is implemented by the inverse wavelet transformation.

4.1.3 The Constraint Condition

Depending on the accuracy for which the locations of the original k -space samples are maintained, there are two ways to deal with the constraint condition.

After we have the measurement data from the MR scanner which for example consists of 63 radial lines with 512 samples each, we can round off the sample locations to the integer grids. Then the nonzero integer grids become the constraint set Ω . This corresponds to 0 order interpolation which is not accurate. But we have the advantage of imposing the condition $Fm = y$ directly by an inverse fast Fourier transform (FFT). Combining with descending at the gradient of $\|\psi(m)\|_{L^1}$, m converges to the solution of (4.1) according to the well-known algorithm of projection onto convex set (POCS) (cf. [Bre65]).

A better interpolation is to retain the locations of the sampled data and use non-uniform FFT operator to approach y . We achieve this by minimizing $\|Am - y\|_2^2$, where A denotes the non-uniform FFT operator, which we use NFFT from [KP04] or NUFFT from [FS03]. We have found great convenience in modifying the NFFT package that is written in C into C++. However, it is faster to operate the interpolation by NUFFT package written in matlab because the NUFFT is designed under the optimal min-max distance. To clarify it, the interpolation coefficients of NUFFT depend on the location of the data in the spatial domain only while the interpolation coefficients of NFFT also depend on the value of the data in the spatial domain. Thus for NUFFT only one time pre-computation of the interpolation coefficients is needed, while for NFFT this computation needs to be repeated because the value of the data Am is changing.

By adding a parameter λ to control the L^2 distance between Am and y , our

problem is reformulated as

$$\min_m \left\{ \|\psi(m)\|_1 + \frac{\lambda}{2} \|Am - y\|_2^2 \right\}. \quad (4.8)$$

The sub-gradient of the second term in (4.8) is $\lambda\hat{A}(Am - y)$, where \hat{A} is the adjoint of A .

4.2 The Sparse Representation for a Piecewise Smooth image

As it is proven in [CRT04], the reason that we can reconstruct the exact signal from incomplete frequency information is based on the assumption that the image has a sparse representation ψ . For a piecewise constant image, both the BV norm and the wavelet transform are good sparse representations. However when we are given raw measurement k -space data scanned from a real object, which is often piecewise smooth, the TV regularization or the L^1 /wavelet alone will not be a good sparse transformation. We have to look for other means.

4.2.1 Wavelet + TV

In [CDS98], Chen et al. proposed a dictionary merger *jump+wavelet*. As they claim, the *jump+wavelet* dictionary based on a merger of wavelets with tapered Heavisides (The Heaviside is equivalent to total variation norm in one dimension) will lead to a sparse representation for a piecewise smooth image. Based on such an over-complete dictionary we will decompose the image m into two components where h is composed of a Heaviside orthonormal basis and w is composed of a wavelet orthonormal basis. To be more general, when $m = \mu h + \nu w$, the following

problem will be considered:

$$\min_{h,w,m=\mu h+\nu w} \left\{ \|m - \nu w\|_{BV} + \|\nu w\|_W + \frac{\lambda}{2} \|Am - y\|_2^2 \right\}, \quad (4.9)$$

where μ and ν are nonnegative parameters. (4.9) is still a convex formulation due to the linear relationship between m , h and w . Furthermore, when $\mu = 0$ or $\nu = 0$, (4.9) is simplified to the case using wavelet transformation or total variation regularization alone.

We also want to draw a connection to the work [CR05], where $\|\psi(m)\|_1 = \|m\|_{BV} + \nu \|m\|_W$ is considered to solve the minimization problem (4.8), where the parameter ν prescribes the importance of the solution having small L^1 norm in the wavelet domain versus having small BV norm in the spatial domain. However, our formulation (4.9) is more faithful to finding a sparse representation over an over-complete dictionary *jump+wavelet* for the piecewise smooth functions.

4.2.2 Curvelet + TV

The curvelet transformation (cf. [CD02, CDD05, CG02]) was developed by Candes et al. in the last few years in an attempt to overcome the inherent limitations of traditional multiscale representations such as wavelets. Conceptually, the curvelet transform is a multiscale pyramid with many directions and positions at each length scale, and needle-shaped elements at fine scales. Thus those curvelets including new tight frames of curvelets (cf. [CD02]) are able to address the problem of finding optimally sparse representations of objects with discontinuities along C^2 edges, for which wavelets are far from ideal. Based on the good behavior of curvelet+TV in denoising images in [CG02] we believe that it is worth implementing curvelet+TV as the sparse representation.

4.3 The Iterative Refinement Method and the Inverse Scale Space Method

4.3.1 The Bregman iteration

By substituting the nonuniform FFT operator A into the convolution kernel k in Section 1.3.2 and following the calculus there, our Bregman iteration is in the same manner. That is, at the k th iteration, we add the residual v_{k-1} back to y .

There is an intuitive perspective to explain the reason why we apply the Bregman iteration. We consider a much simpler case where we full sample the Fourier coefficients of a disk on the integer grids. That means we have the complete information of $y = F(\alpha\chi_{(x_0,y_0)}^R)$, where F denotes the uniform Fourier transform, α is the grey value of the image and

$$\chi_{x_0,y_0}^R(x,y) = \begin{cases} 1, & \text{if } (x-x_0)^2 + (y-y_0)^2 \leq R^2; \\ 0, & \text{otherwise.} \end{cases} \quad (4.10)$$

To study the link between the Bregman iteration and our model (4.8) we consider the following specific minimization problem:

$$\min_{m \in BV} \|m\|_{BV} + \frac{\lambda}{2} \|Fm - y\|_2^2. \quad (4.11)$$

i.e. $\|\psi(m)\|_1 = \|m\|_{BV}$ and non-uniform fast Fourier transform A becomes uniform Fourier transform F . Based on the fact $\hat{F}Fm = m$ and Lemma 3.1.3 we obtain the solution of (4.11) $m_1^\lambda = (\alpha - \frac{2}{\lambda R})\chi_{(x_0,y_0)}^R$ when $\|\hat{F}y\|_* = \frac{\alpha R}{2} > \frac{1}{\lambda}$. Thus the residual $v_1^\lambda = y - Fm_1^\lambda = F\frac{2}{\lambda R}\chi_{(x_0,y_0)}^R$. Therefore, at the second Bregman iteration, we end up minimizing (4.11) with $y = F(\alpha + \frac{2}{\lambda R})\chi_{(x_0,y_0)}^R$, see Equation (1.13). This will reconstruct the image m exactly, i.e. $m_2^\lambda = \alpha\chi_{(x_0,y_0)}^R$.

4.3.2 Nonlinear Inverse Scale Space Methods

The nonlinear inverse scale space method described in [BGO06] is derived as a limit of the iterative refinement procedure for $\lambda \rightarrow 0$.

The key point is to reinterpret the fidelity parameter $\lambda = \Delta t$ as a time step and divide by Δt on both sides of the Euler-Lagrange equation (1.12), then

$$\frac{\partial J(m_k^{\Delta t}) - \partial J(m_{k-1}^{\Delta t})}{\Delta t} = -\widehat{A}(Am_k^{\Delta t} - y).$$

Setting $t_k = k\Delta t$, $p^{\Delta t}(t_k) = \partial J(m_k^{\Delta t})$ and $m^{\Delta t}(t_k) = m_k^{\Delta t}$, then we have

$$\frac{p^{\Delta t}(t_k) - p^{\Delta t}(t_k - \Delta t)}{\Delta t} = -\widehat{A}(Am^{\Delta t}(t_k) - y).$$

Letting $\Delta t \rightarrow 0$ and dropping the subindex k we arrive at the differential equation

$$\partial_t p(t) = -\widehat{A}(Am(t) - y), \quad p(t) \in \partial J(m(t)), \quad (4.12)$$

with initial values given by $m(0) = p(0) = 0$.

The concise formulation of (4.12) is not straightforward to compute, as the relations between p and m are quite complicated in nonlinear cases. Here we resort to a relaxed version (cf. [BGO06]) which aims at having a flow with qualitatively similar properties to that of (4.12) by using standard variational formulations, which are simple to compute. We compute the following coupled equations :

$$\begin{cases} \partial_t m(t) &= -p(t) - \lambda(\widehat{A}Am(t) - \widehat{A}y - v(t)) \\ \partial_t v(t) &= -\alpha(\widehat{A}Am(t) - \widehat{A}y), \end{cases} \quad (4.13)$$

where $m(0) = v(0) = 0$ and $\alpha > 0$ is a constant. It is easy to see that the steady state of these equations is $u = f$, $v = p/\lambda$. In order for the solution of (4.13) to converge to the steady state, $\alpha \leq \lambda/4$ is required. Please see [BGO06] for detailed analysis.

To apply the Bregman iteration, the minimization problem has to be solved several times before a stopping criterion is satisfied. This could be very time-consuming. The inverse scale-space method is shown to be more straightforward and more efficient than Bregman iteration. Our numerical experiments indicate that the nonlinear inverse scale space method works much faster than the iterative refinement method with similar results.

4.4 Experimental Results

The raw MR measurement data were obtained from a Siemens Magnetom Avanto 1.5T scanner. These data samples have a radial trajectory in k -space. Figure 4.1, 4.2 and 4.3 are reconstructed from the raw measurement k -space data scanned from a numeric phantom. The k -space data is composed of a total of 63 radial lines with 512 samples each. During the MR scanning, three coils/channels were used, the pulse sequence is trueFISP, the scanning parameters are TR=4.8ms, TE=2.4ms, flip angle $\alpha=60^\circ$, FOV=206mm with a resolution of 256 pixels; and the final image is obtained by taking the square root of the sum of each channel, which separately went through the proposed iterative procedure or the non-linear inverse scale space method.

We solved the minimization problem (4.8) by conjugate gradient descent method (cf. [She94]) and back-tracking line search (cf. [BV04]). We chose the sparse representation and the parameter λ to reach the best visual effects.

Figure 4.1 is obtained from the conventional gridding algorithm. We see a lot noise and image artifacts because the data is sampled sparsely.

Compared with Figure 4.1, the top left image in Figure 4.2 obtained from solving (4.8) with $\|\phi(m)\|_1 = \|m\|_{BV}$ and $\lambda = 100.0$ demonstrates superior noise

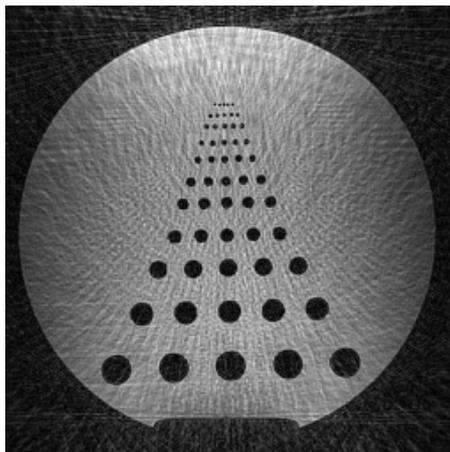


Figure 4.1: Conventional gridding algorithm.

deduction. In order to enforce the constrained condition $Am = y$ we require a small value of the fidelity term $\|Am - y\|_2^2$. We could use a large λ , however this would result in a noisy image with a large bounded variation norm. Therefore, we choose to apply the Bregman iteration in solving the minimization problem (4.8). The sequence images of the 1st, 2nd, 4th and 6th Bregman iteration are shown in figure 4.2. Each Bregman iteration takes 200-500 steps to converge. As the iteration k increases, we observe a recovery of fine details, particularly including the black dots in the top row.

In Figure 4.3, we present another four images corresponding to the 100th, 200th, 400th and 600th iteration results obtained from the time continuous flow of the relaxed inverse scale space method with $\|\phi(m)\|_1 = \|m\|_{BV}$ and $\lambda = 100.0$. As mentioned in (4.13), it starts from $m(0) = 0$ and $v(0) = 0$. Thus we notice image artifacts at the early step, see the top two images of Figure 4.3. The artifacts disappear at the later step and the image of the 600th steps from the non-linear inverse scale space flow (see the bottom right of Figure 4.3) is comparable with the result of the 6th Bregman iteration (see the bottom right of Figure

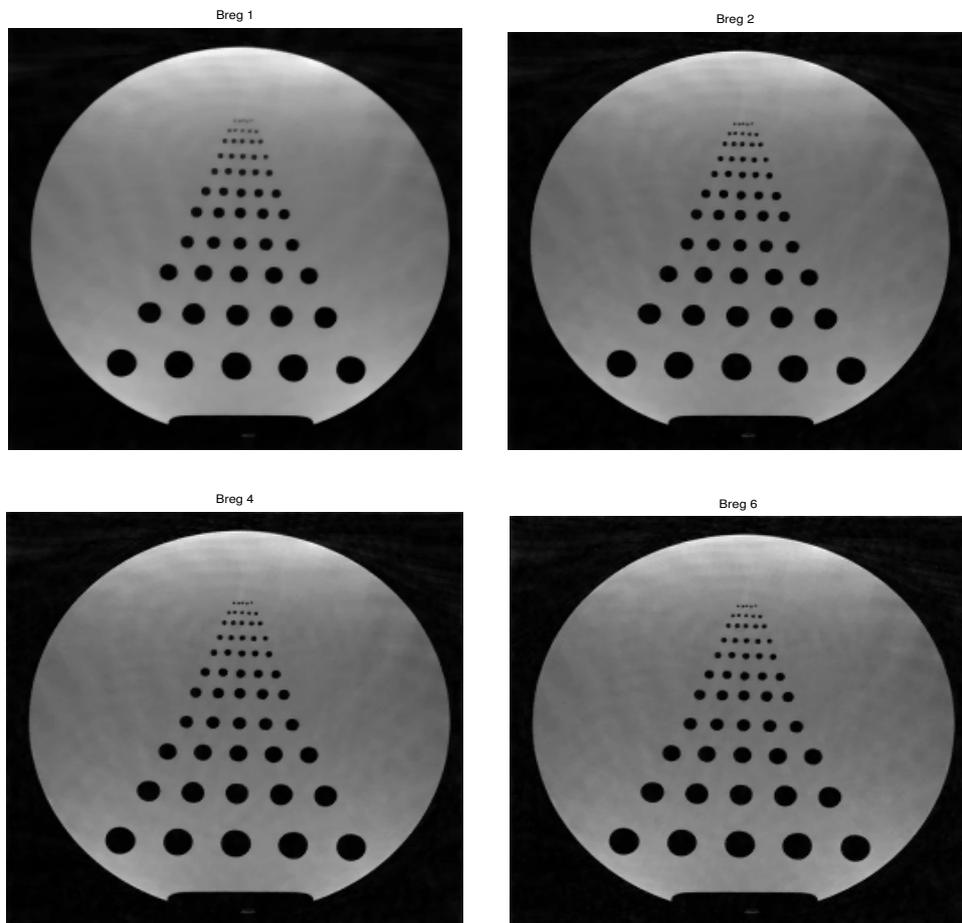


Figure 4.2: A sequence images obtained from the iterative refinement method : top left, 1st Bregman iteration; top right, 2nd Bregman iteration; bottom left, 4th Bregman iteration; bottom right, 6th Bregman iteration.

4.2). However, the iterative refinement procedure is around 3-4 times more time-consuming.

Figure 4.4, 4.5 and 4.6 are reconstructed from 2 out of 8 interleaves of the k -space data scanned from a head. Each interleave has 31 radial lines with 512 samples each and 4 channels are used during the scan. Scanning parameters are TR=4.46ms, TE=2.23ms, flip angle $\alpha=50^\circ$, FOV=250mm with a resolution of 256 pixels.

Figure 4.4 is obtained from conventional gridding algorithm. Again we see a lot of noise and image artifacts.

The minimization problem (4.9) is solved for Figure 4.5 and 4.6 with $\mu = 1.0$, $\nu = 0.1$ and $\lambda = 100.0$. This is because, as we mentioned above, with the wavelet and heaviside merged over-complete dictionary we will have a better sparse representation. The Bregman iteration (see Figure 4.5) and the nonlinear inverse scale space method (see Figure 4.6) are carried out as in the above section. We observe that Figure 4.5 and 4.6 have less noise and are smoother compared with Figure 4.4. However, in terms of sharpness, Figure 4.4 is the best. This might be resulted from the ability of total variation to suppress oscillations. The nonlinear inverse scale space method is two times faster than the iterative refinement method.

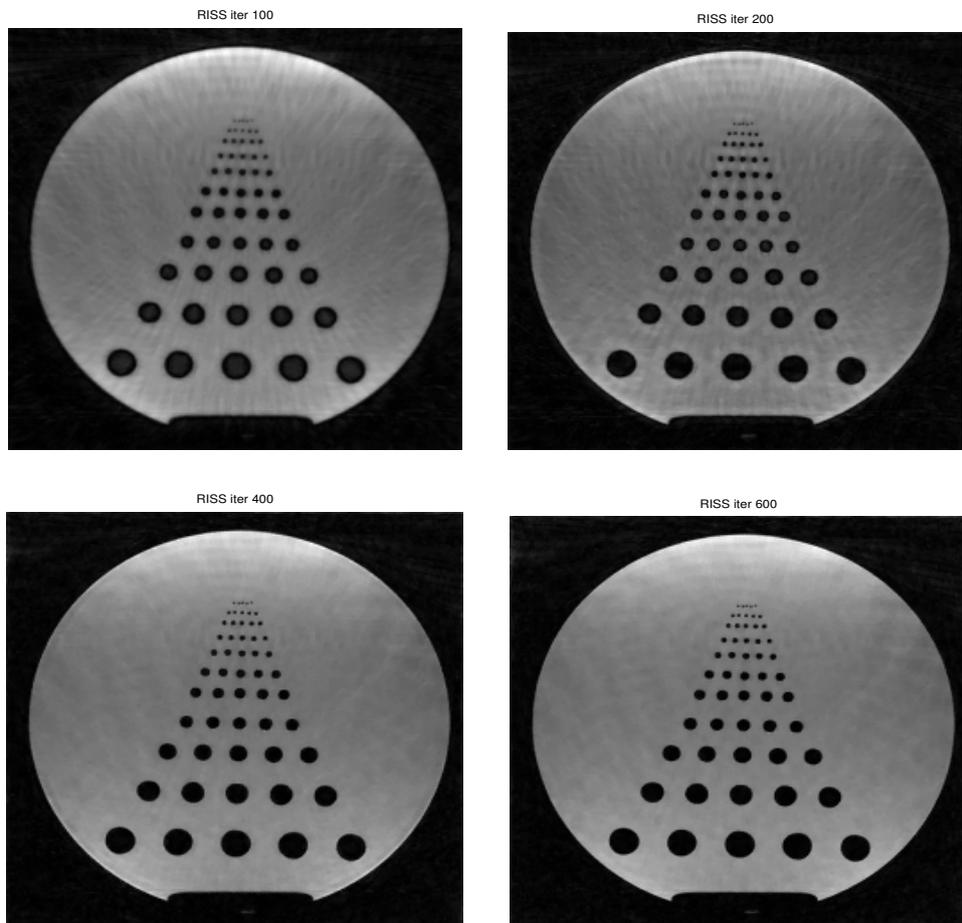


Figure 4.3: A sequence images obtained from the relaxed inverse scale space flow : top left, iteration 100; top right, iteration 200; bottom left, iteration 400; bottom right, iteration 600.

Gridding method

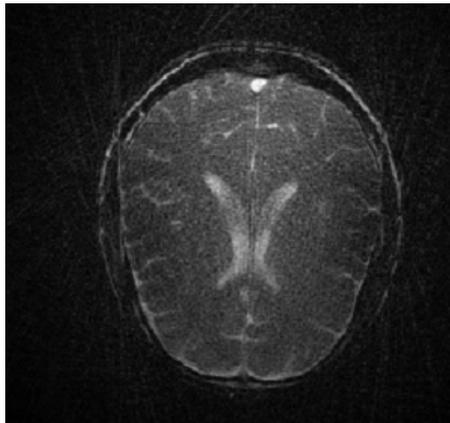


Figure 4.4: Conventional gridding algorithm.

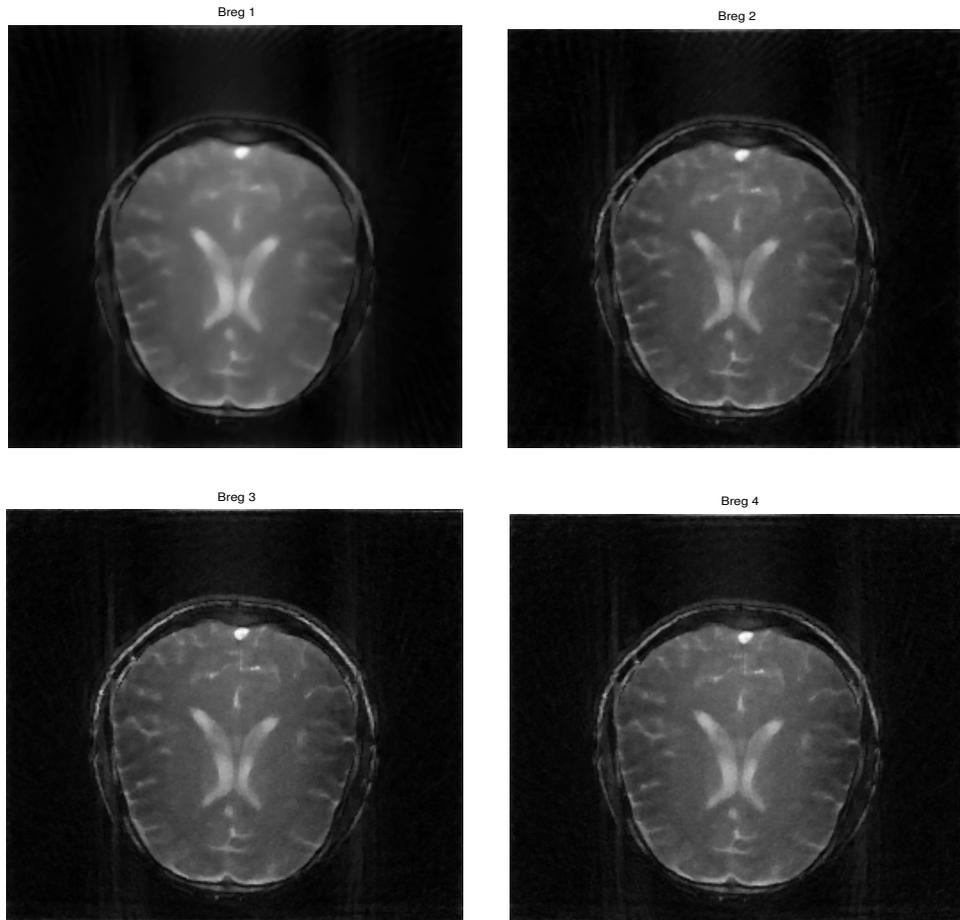


Figure 4.5: A sequence images obtained from the iterative refinement method with (4.9) where $\mu = 1.0$, $\nu = 0.1$, and $\lambda = 100$: top left, 1st Bregman iteration; top right, 2nd Bregman iteration; bottom left, 3rd Bregman iteration; bottom right, 4th Bregman iteration.

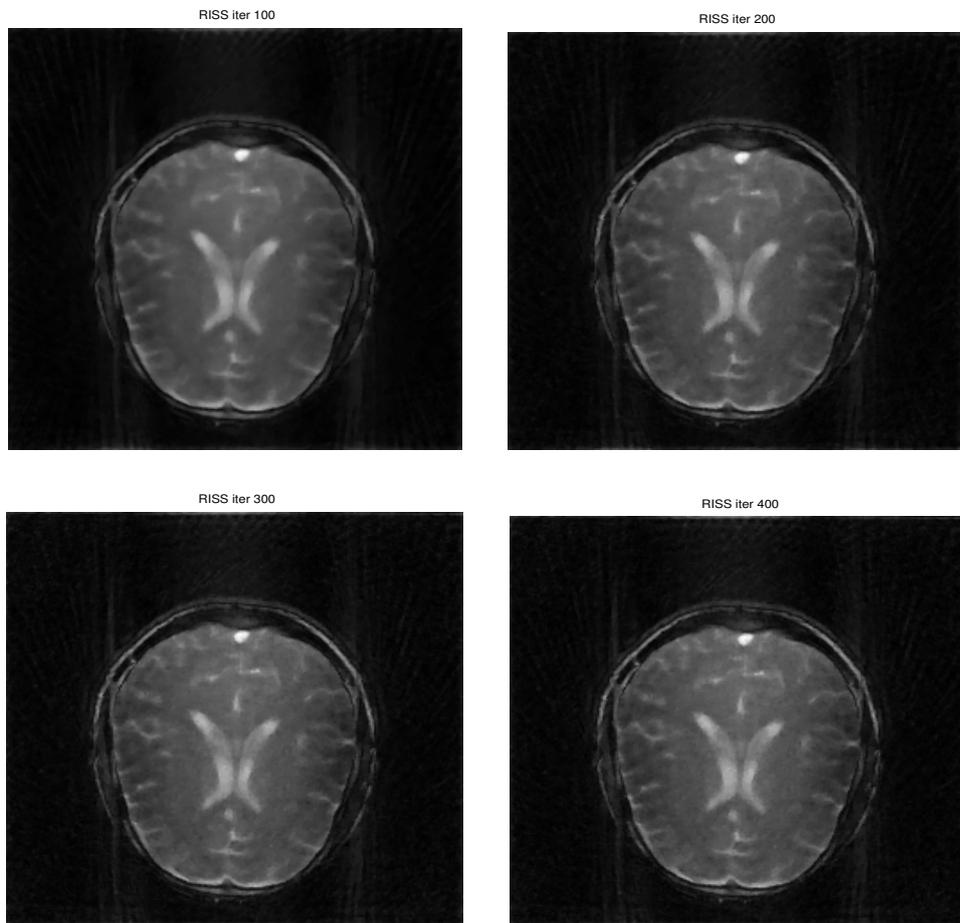


Figure 4.6: A sequence images obtained from the relaxed inverse scale space flow with (4.9) where $\mu = 1.0$, $\nu = 0.1$, and $\lambda = 100$: top left, iteration 100; top right, iteration 200; bottom left, iteration 300; bottom right, iteration 400.

CHAPTER 5

Conclusion and Future Work

In conclusion, we have proposed a modified iterative total variation regularization procedure (cf. [HBO06]) that can be applied to any variational models with non-quadratic convex fidelity terms and we have applied the original iterative regularization procedure (cf. [OBG05]) into the blind deconvolution problem (cf. [HMO05]) and MR image reconstruction from undersampled k-space data (cf. [CHF06]). Experimental results of good quality have illustrated that finer scales are recovered along the Bregman iteration.

The ongoing work for MR image reconstruction is to find a better sparse representation for piecewise smooth functions. One possible option we are investigating right now is curvelet + TV.

Another possible future study is to apply the modified iterative regularization procedure to the BV + G model (1.8) suggested by Meyer. Since the G norm is a Banach norm satisfying the assumptions nonnegativity, convexity and positive homogeneity, it seems possible that our approach could improve the denoising performance of the model.

REFERENCES

- [AAB05] Jean-Francois Aujol, Gilles Aubert, Laure Blanc-Feraud, and Antonin Chambolle. “Image Decomposition Into a Bounded Variation Component and an Oscillating Component.” *Journal of Mathematical Imaging and Vision*, **22**(1):71–88, 2005.
- [AC05] Jean-Francois Aujol and Antonin Chambolle. “Dual Norms and Image Decomposition Models.” *Internat. J. Comput. Vision*, **63**:85–104, 2005.
- [BGG98] C.Sidney Burrus, R.A. Gopinath, and Haitao Guo. “Introduction to Wavelets and Wavelet Transformations.” 1998. A Primer, Prentice-Hall.
- [BGO06] Martin Burger, Guy Gilboa, Stanley Osher, and Jin-Jun Xu. “Nonlinear Inverse Scale Space Methods.” *Comm. Math. Sci.*, **4**(1):175–208, 2006.
- [BO04] Martin Burger and Stanley Osher. “Convergence Rates of Convex Variational Regularization.” *Inverse Problems*, **20**:1411–1422, 2004.
- [Bre65] L. M. Bregman. “Finding the Common Point of Convex Sets by the Method of Successive Projection.” *Dokl. Akad. Nauk. USSR*, **162**:487–490, 1965.
- [Bre67] L. M. Bregman. “The Relaxation Method for Finding the Common Point of Convex Sets and its Application to the Solution of Problems in Convex Programming.” *USSR Comp. Math. and Math. Phys.*, **7**:200–217, 1967.
- [BV04] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.
- [CCW99] Raymond H. Chan, Tony F. Chan, and Chiu-Kwong Wong. “Cosine Transform Based Preconditioners for Total Variation Deblurring.” *IEEE Transactions on Image Processing*, **8**:10:1472–1478, 1999.
- [CD02] Emmanuel J. Candes and David L. Donoho. “New Tight Frames of Curvelets and Optimal Representations of Objects with Smooth Singularities.” Technical Report, Stanford University, Submitted, 2002.
- [CDD05] Emmanuel J. Candes, Laurent Demanet, David L. Donoho, and Lexing Ying. “Fast Discrete Curvelet Transforms.” 2005.

- [CDS98] Scott Shaobing Chen, David L. Donoho, and Michael A. Saunders. “Atomic Decomposition By Basis Pursuit.” *SIAM J. Scientific Computing*, **20**:33–61, 1998.
- [CE05] Tony F. Chan and Selim Esedoglu. “Aspects of Total Variation Regularized L^1 Function Approximation.” *SIAM J. Appl. Math.*, **65**(5):1817–1837, 2005.
- [CG02] Emmanuel J. Candes and Frank Guo. “New Multiscale Transforms, Minimum Total Variation Synthesis: Applications to Edge-Preserving Image Reconstruction.” *Signal Processing*, **82**:1519–1543, 2002.
- [CHF06] Ti-Chiun Chang, Lin He, and Tong Fang. “MR Image Reconstruction from Sparse Radial Samples Using Bregman Iteration.” ISMRM, 2006.
- [CR05] Emmanuel J. Candes and Justin Romberg. “Practical Signal Recovery from Random Projections.” 2005.
- [CRT04] Emmanuel J. Candes, Justin Romberg, and Terence Tao. “Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information.” *submitted to IEEE Trans. Inform. Theory*, June 2004.
- [CT93] Gong Chen and Marc Teboulle. “Convergence Analysis of a Proximal-Like Minimization Algorithm Using Bregman Functions.” *SIAM J. Optim.*, **3**:538–543, 1993.
- [CW98] Tony F. Chan and Chiu-Kwong Wong. “Total Variation Blind Deconvolution.” *IEEE Trans. Image Process.*, **7**:370–375, 1998.
- [CW00] Tony F. Chan and Chiu-Kwong Wong. “Convergence of the Alternating Minimization Algorithm for Blind Deconvolution.” *Linear Algebra and its Applications*, **316**(1–3):259–285, 2000.
- [FS03] Jeffrey A. Fessler and Bradley P. Sutton. “Nonuniform Fast Fourier Transforms Using Min-Max Interpolation.” *IEEE Trans. Signal Process.*, **51**(2), February 2003.
- [GY04] Donald Goldfarb and Wotao Yin. “Second Order Cone Programming Methods for Total Variation-Based Image Restoration.” *2005 Society for Industrial and Applied Mathematics*, **27**(2):622–645, 2004.
- [HBO06] Lin He, Martin Burger, and Stanley Osher. “Iterative Total Variation Regularization with Non-Quadratic Fidelity.” *J. of Mathematical Imaging and Vision*, 2006.

- [HMO05] Lin He, Antonio Marquina, and Stanley Osher. “Blind Deconvolution Using TV Regularization and Bregman Iteration.” *International Journal of Imaging Systems and Technology*, **15**:74–83, 2005.
- [KP04] Stefan Kunis and Daniel Potts. “NFFT2.0.Beta.” technical report, University of Lubeck, Germany, 2004.
- [LLD05] Michael Lustig, Jin Hyung Lee, David L. Donoho, and John M. Pauly. “Faster Imaging with Randomly Perturbed Undersampled Spirals and L1 Reconstruction.” In *Proc. Of the ISMRM '05*, 2005.
- [Mal99] Stephane Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, 1999.
- [Mey93] Yves Meyer. *Wavelets Algorithms and Applications*. SIAM, Philadelphia, 1993.
- [Mey01] Yves Meyer. *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations*. AMS, Providence, RI, 2001.
- [MO01] Antonio Marquina and Stanley Osher. “Explicit Algorithms for a New Time Dependent Model Based on Level Set Motion for Nonlinear Deblurring and Noise Removal.” *IAM Journal on Scientific Computing*, **22**:387–405, 2001.
- [OBG05] Stanley Osher, Martin Burger, Donald Goldfarb, Jin-Jun Xu, and Wotao Yin. “An Iterative Regularization Method for Total Variation Based Image Restoration.” *Multiscale Modeling and Simulation*, **4**(2):460–489, 2005.
- [OSV03] Stanley Osher, Andres Sole, and Luminita Vese. “Image Decomposition and Restoration Using Total Variation Minimization and the H^{-1} Norm.” *Multiscale Model. Simul.*, **1**:349–370, 2003.
- [Pla96] Robert Plato. “On The Discrepancy Principle for Iterative and Parametric Methods to Solve Linear Ill-Posed Equations.” *Numer. Math.*, **75**:99–120, 1996.
- [RO94] Leonid I. Rudin and Stanley Osher. “Total Variation Based Image Restoration with Free Local Constraints.” In *Proceedings of the IEEE International Conference on Image Processing*, pp. 31–35, Austin, TX, 1994.

- [ROF92] Leonid I. Rudin, Stanley J. Osher, and Emad Fatemi. “Nonlinear Total Variation Based Noise Removal Algorithms.” *Phys. D*, **60**:259–268, 1992.
- [RT92] Elisabeth Rouy and Agnes Tourin. “A Viscosity Solutions Approach to Shape-from-Shading.” *SIAM. J. Numer. Anal.*, **29**(3):867–884, 1992.
- [SG01] Otmar Scherzer and Chuck Groetsch. “Inverse Scale Space Theory for Inverse Problems.” In M.Kerckhove, editor, *Scale-Space and Morphology in Computer Vision, Lecture Notes in Comput. Sci. 2106*, pp. 317–325. Springer, New York, 2001.
- [She94] Jonathan Richard Shewchuk. “An Introduction to the Conjugate Gradient Method Without the Agonizing Pain.” August 1994.
- [TNV04] Eitan Tadmor, Suzanne Nezzar, and Luminita Vese. “A Multiscale Image Representation Using Hierarchical (BV, L^2) Decompositions.” *Multiscale Model. Simul.*, **2**:554–579, 2004.
- [VO98] Curtis R. Vogel and Mary E. Oman. “Fast, Robust Total Variation-Based Reconstruction of Noisy, Blurred Images.” *IEEE Transaction on Image Processing*, **7**(6), June 1998.
- [VO03] Luminita Vese and Stanley Osher. “Modeling Textures with Total Variation Minimization and Oscillatory Patterns in Image Processing.” *J. Sci. Comput.*, **19**:553–572, 2003.
- [XO06] Jin-Jun Xu and Stanley Osher. “Iterative Regularization and Nonlinear Inverse Scale Space Applied to Wavelet Based Denoising.” 2006. submitted to IEEE Trans. Image Proc.
- [YGO05] Wotao Yin, Donald Goldfarb, and Stanley Osher. “Image Cartoon-Texture Decomposition and Feature Selection Using The Total Variation Regularized L1 Functional.” In *Lecture Notes in Computer Science 3752*, pp. 73–84. Variational, Geometric, and Level Set Methods in Computer Vision, Springer, 2005.
- [YK96] Yu-Li You and M. Kaveh. “A Regularization Approach to Joint Blur Identification and Image Restoration.” *IEEE Transaction on Image Processing*, **5**:416–428, 1996.
- [YK99] Yu-Li You and M. Kaveh. “Blind Image Restoration by Anisotropic Regularization.” *IEEE Trans. Image Processing*, **8**:396–407, 1999.

- [ZF05] Xiao-Qun Zhang and Jacques Froment. “Total Variation Based Fourier Reconstruction and Regularization for Computerized Tomography.” Nuclear Science Symposium and Medical Imaging Conference, 2005.