University of California
Los Angeles

Iterative Regularization and Nonlinear Inverse Scale Space Methods in Image Restoration

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

by

Jinjun Xu

2006

The dissertation of Jinjun Xu is approved.

Luminita A. Vese

Stefano Soatto

Andrea L. Bertozzi

Stanley J. Osher, Committee Chair

University of California, Los Angeles

2006

To my parents

Table of Contents

List of Figures

Acknowledgments

I wish to express my sincere and deepest gratitude to my advisor, Professor Stanley Osher, for his dedicated guidance and generous support throughout my studies at UCLA. Had I not met him in Beijing five years ago, I might have not been able to present this dissertation now. His enthusiasm, intelligence and patience have helped me mature in my studies and research. I greatly appreciate his care and help in my life and career development. I would also like to thank my other committee members: Professor Andrea Bertozzi, Professor Stefano Soatto and Professor Luminita Vese for their great help, encouragement and valuable suggestions on my studies and research. I thank my other collaborators: Professor Martin Burger, Dr. Guy Gilboa, Professor Donald Goldfarb, Dr. Stefan Kindermann and Wotao Yin. The work of this dissertation owes to our stimulating discussions and effective collaborations. I also thank Professor Tony Chan and Professor Antonio Marquina for their help in my studies and career development.

I thank my parents and my brother and sister for their continuous and constant support through the years. 亲爱的爸爸妈妈，谢谢您们的教导和支持，祝福您们幸福安康。

I thank many people for their help on my studies as well as my life at UCLA: Lin He, Shing-Yu Leung, Jason Chung, Triet Le, Lok-Ming Liu, Frederick Park, Igor Yanovsky, Youri Bae, Eric Chung, David Shao, Chiu-Yen Kao, Edward Fan, Xiaosheng Li, Wei Zhu, Ming-Ham Yip, Maggie Albert, Rocie Carrillo, Babette Dalton and many more. I also thank IPAM and its staff for providing me a wonderful working environment.

This work was supported in part by the NIH through grant U54RR021813, and the NSF through grants DMS-0312222, ACI-0321917 and DMI-0327077. Plus the financial support from UCLA and the Department of Mathematics is also gratefully acknowledged.

Collaborations

All the chapters are co-authorized with Professor Stanley Osher. Chapter 2 is also a joint work with Professor Martin Burger, Professor Donald Goldfarb and Wotao Yin and is a version of [OBG05]. Chapter 3 is also a joint work with Professor Martin Burger and Dr. Guy Gilboa and is a version of [BOX05, BGO06]. Chapter 3 is a version of [XO06].

Vita

1978	Born, Fujian, P.R. China.
1998	B.S., Applied Mathematics Beijing University, P.R. China
2001	M.S., Applied Mathematics Beijing University, P.R. China
2002	M.A., Mathematics University of California, Los Angeles
2001-2006	Teaching Assistant and Research Assistant Department of Mathematics University of California, Los Angeles

Publications

MARTIN BURGER, GUY GILBOA, STANLEY OSHER, AND JINJUN XU. "Nonlinear Inverse Scale Space Methods." *Comm. Math. Sci.*, Volume 4(1), Pages 175–208, 2006.

MARTIN BURGER, STANLEY OSHER, JINJUN XU, AND GUY GILBOA. "Nonlinear Inverse Scale Space Methods for Image Restoration." *Lecture Notes in Computer Science*, Volume 3752, Pages 25–36, 2005.

Stefan Kindermann, Stanley Osher, and Jinjun Xu. "Denoising by BV-Duality." *UCLA-CAM-Report 05-21, Los Angeles, CA, USA*, 2005. to appear in J. of Sci. Comput., 2006.

Stanley Osher, Martin Burger, Donald Goldfarb, Jinjun Xu, and Wotao Yin. "An Iterative Regularization Method for Total Variation Based Image Restoration." *Multiscale Model. and Simul.*, Volume 4, Pages 460–489, 2005.

Abstract of the Dissertation

Iterative Regularization and Nonlinear Inverse Scale Space Methods in Image Restoration

by

Jinjun Xu

Doctor of Philosophy in Mathematics

University of California, Los Angeles, 2006

Professor Stanley J. Osher, Chair

Total variation regularization is a popular and important technique for image restoration. In this dissertation, first we develop a new iterative regularization procedure for inverse problems based on the use of generalized Bregman distances, with particular focus on the problems arising in total variation based image restoration. We obtain rigorous convergence results and effective stopping criterion for the general procedure. Specifically, when a discrepancy principle is used as the stopping criterion, the error measured by the Bregman distance between the reconstruction and the noise-free image decreases until termination, The numerical results for denoising appear to give significant improvement over standard models.

Then we generalize this discrete iterative regularization procedure to a time-continuous inverse scale space formulation, which arises as reinterpreting the scale parameter in the iterative regularization procedure as a time-step and letting it go to zero. Similar properties as for iterative regularization procedure hold for the inverse flow. For one-dimensional signal processing, the inverse flow can be computed directly through an integro-differential equation, yielding high qual-

ity restoration. In higher spatial dimensions, a relaxation technique has been introduced to solve the inverse flow.

Finally, the iterative regularization and inverse scale space techniques are successfully generalized from total variation based image restoration to wavelet based image restoration.

CHAPTER 1

Introduction

Due to its increasing applications, digital image processing has become a more and more important and active research field in the past decade. For a recent review of image denoising algorithms please refer to [CS05, BCM05, TO05, CEP05c].

1.1 Images

Mathematically an image can be represented as a function $f : \Omega \to \mathbb{R}^N$, where Ω is a bounded subset of $\mathbb{R}^d$. If $d = 1$, f is a one-dimensional (1D) signal; if $d = 2$, f is a two-dimensional (2D) image. For simplicity, we use the word "image" for all dimensions unless otherwise specified. The integer N is the number of the color channels of the images. In general, $N = 1$ means f is a grey level image, $N = 3$ means f is a color image (with RGB color values). In the scope of this dissertation we will focus on the analysis of grey level images, i.e., $N = 1$, f is a scalar value function.

Given an original image g, which we refer to as a true image or a clean image, its corrupted version f can be generally categorized as two main types: noisy image $f = g + n$ and blurry/noisy image $f = Ag + n$. Here n is noise which is usually random and highly oscillatory data (cf. [Mey01]). A is an integral operator, which in most case is a convolution kernel. We will mainly focus on the restoration of noisy images.

In the discrete case, f is an array (1D) or a matrix (2D) defined on a grid. In numerical experiments, the grid size can be set as h or simply $h = 1$ by rescaling the parameters (e.g., λ and t in the time-dependent Euler-Lagrange equation of the ROF model). In all the numerical results presented in this dissertation, we use $h = 1$ if not otherwise specified.

One important measurement characterizing the quality of images is the *signal-to-noise-ratio* (SNR). It has many different definition in the literatures. Given a clean image g and a noisy image f, one popular definition of SNR which is used in this dissertation is

$$SNR(f) = 20\log_{10}\left(\frac{\|g-\bar{g}\|_{L^2}}{\|n-\bar{n}\|_{L^2}}\right), \tag{1.1}$$

where $n = f - g$ is the noise, $\bar{g}$ and $\bar{n}$ are the average of the values g and n respectively. $\|\cdot\|_{L^2}$ is the regular L^2 norm. In discrete case, if g is a $M \times N$ matrix, then $\bar{g} = \frac{\sum_{i,j} g_{i,j}}{M\times N}$, $\|g\|_{L^2} = \sqrt{\frac{\sum_{i,j}|g_{i,j}|^2}{M\times N}}$.

1.2 Total Variation Based Image Restoration

Given a noisy image f, we want to obtain a decomposition:

$$f = u + v,$$

where u recovers the true image and v is the residual which is expected to be noise.

This is an inverse problem and it has a very long history (cf., e.g., [Mey01, CS05]). One of the most successful and popular techniques for approximating the solution of this problem is due to Rudin-Osher-Fatemi (cf. [ROF92] and also their related work [OR90, RO94]), and is defined as follows:

$$u = \operatorname*{arg\,min}_{u \in BV(\Omega)} \left\{ |u|_{BV} + \lambda \|f - u\|_{L^2}^2 \right\} \tag{1.2}$$

for some scale parameter $\lambda > 0$, where $BV(\Omega)$ denotes the space of functions with bounded variation on Ω and $|\cdot|_{BV}$ denotes the BV seminorm, formally given by

$$|u|_{BV} = \int_\Omega |\nabla u| \, dxdy, \tag{1.3}$$

which is also referred to as the *total variation* (TV) of u. This variational problem is called the ROF model. It has been used and analyzed by several authors in several different contexts (cf. [AV94, CL97, CK97, CKP99, MO00, Nik00, Rin00, Vog02, SC03, HM04]). Also, in [CL97] and subsequently in [Nik00, Nik97, Nik04, Rin00, CEP05a, CEP05b] the "staircasing" effect of this model was analyzed. No completely satisfying remedy has yet been found, e.g. see the results in Figures 2.2 (the part $x \in [0, 120]$) in Chapter 2. In spite of this phenomenon, the ROF model is still quite popular. (In the one-dimensional case, (1.3) becomes $|u|_{BV} = \int_I |u_x| \, dx$ where $\Omega = I \subset \mathbb{R}$. For simplicity, in the following of this dissertation we will drop dx and $dxdy$, i.e., $\int_\Omega \equiv \int_\Omega dx$ for $\Omega \subset \mathbb{R}$ and $\int_\Omega \equiv \int_\Omega dxdy$ for $\Omega \subset \mathbb{R}^2$, etc., unless otherwise specified.)

The use of the BV seminorm is essential since it allows us to recover images with edges. It is well-known that this would be impossible if the first term in (1.2) were replaced by $J_p(u) := \int_\Omega |\nabla u|^p$ for any $p > 1$, which might seem more attractive at a first glance due to differentiability and strict convexity. The main reason for this effect is that for $p > 1$ the derivative of J_p corresponds to a nondegenerate elliptic differential operator of second order and thus has a smoothing effect in the optimality condition, whereas for total variation the operator is degenerate and only affects the level lines of the image.

The ideal result of the minimization procedure (1.2) would be to decompose f into the true signal u and the additive noise v. In practice, this is not fully attainable. We must expect to find some signal in v, and some smoothing of textures in u. The concept "texture" is imprecise so far and the decomposition depends on the scale parameter λ. Large λ corresponds to very little noise removal, and hence u is close to f. Small λ yields a blurry, oversmoothed u. These statements can be quantified, as discussed below.

In his book [Mey01], Meyer did some very interesting analysis on the ROF model. He began by characterizing textures which he defines as "highly oscillatory patterns in image processing" as elements of the dual space of $BV(\Omega)$. This can be motivated by using the rigorous definition of the BV seminorm

$$\int_\Omega |\nabla u| = |u|_{BV} = \sup_{|\vec{\xi}|_\infty \le 1,\, \vec{\xi} \in C_c^1(\Omega)^2} \int_\Omega u(\nabla \cdot \vec{\xi}). \tag{1.4}$$

Here: $\vec{\xi} = (\xi_1, \xi_2), |\vec{\xi}| = \sqrt{\xi_1^2 + \xi_2^2}$. Defining the space G as the distributional closure of the set

$$\left\{ w = \nabla \cdot \vec{\xi} = \partial_x \xi_1 + \partial_y \xi_2 \;\middle|\; \vec{\xi} \in C_c^1(\Omega)^2 \right\},$$

equipped with the norm $\|w\|_* = \inf_{\vec{\xi}} \sup_{x,y} |\vec{\xi}(x,y)|$, Meyer showed that elements of this dual space G can be regarded as textures. He also showed that the space G arises implicitly in the ROF model as follows: For $f = u + v$, with u defined by

(1.2), we have

$$\|f\|_* < \frac{1}{2\lambda} \implies u = 0,\, v = f, \tag{1.5}$$

$$\|f\|_* \geq \frac{1}{2\lambda} \implies \|v\|_* = \frac{1}{2\lambda},\, \int_\Omega uv = |u|_{BV} \|v\|_*. \tag{1.6}$$

The Euler-Lagrange equation arising in the ROF minimization (1.2) is:

$$-\frac{1}{2\lambda} \nabla \cdot \frac{\nabla u}{|\nabla u|} = f - u = v. \tag{1.7}$$

Of course the expression on the left in (1.7) needs to be defined when $|\nabla u| = 0$. This is easily done – see [Mey01].

We see that the term v, which was usually thrown away and which represents noise, is an element of G with $*-$norm $\leq \frac{1}{2\lambda}$. This expression is (formally) $-\frac{1}{2\lambda}$ times the curvature of the level contour of $u(x, y)$ at each point. Moreover, by (1.6), it does have $*-$norm $\frac{1}{2\lambda}$ if $\|f\|_* \geq \frac{1}{2\lambda}$, as was shown in [Mey01].

The following interesting example of the effect of ROF minimization was analyzed in [Mey01, page 36]: Let $f(x, y) = \alpha \chi_R(x, y) \; : \mathbb{R}^2 \mapsto \mathbb{R}$ where $\chi_R(x, y) \equiv 1$ if $\sqrt{x^2 + y^2} \leq R$, $\chi_R(x, y) \equiv 0$ otherwise. Meyer showed (a) $\|f\|_* = \frac{(\alpha R)}{2}$ and (b) the $f = u + v$ decomposition is as follows: If $\alpha \lambda R \geq 1$, then

$$u = \left(\alpha - \frac{1}{\lambda R}\right) \chi_R, \quad v = \frac{1}{\lambda R} \chi_R.$$

Notice that v is independent of α, which is quite surprising. If $\alpha \lambda R \leq 1$, then $u = 0$, $v = f$. Clearly the ROF $u + v$ decomposition is defective in this case. The function v is certainly not noise.

Meyer then suggested a modified variational problem:

$$u = \underset{u \in BV(\Omega)}{\arg\min} \left\{ |u|_{BV} + \lambda \|f - u\|_* \right\}. \tag{1.8}$$

Here we can think of a decomposition, $f = u + v$ where u is a cartoon, or primal sketch, and v is texture plus noise. This model is difficult to minimize using the usual Euler-Lagrange equation approach due to the nonsmoothness of both terms involved in the functional. However, it can be solved effectively as the minimization of a smooth function subject to constraints, for example, as a second-order cone program [GY04].

Vese and Osher [VO03] approximated Meyer's model by

$$(u, \vec{\zeta}) = \underset{(u,\vec{\zeta})}{\operatorname{argmin}} \left\{ |u|_{BV} + \lambda \|f - u - \nabla \cdot \vec{\zeta}\|_{L^2}^2 + \mu \left(\int_\Omega |\vec{\zeta}|^p \right)^{\frac{1}{p}} \right\}, \tag{1.9}$$

with $p \geq 1$ and $\lambda, \mu > 0$. As $\lambda, p \to \infty$ (1.9) approaches Meyer's model. We call the variational problem (1.9) the Vese-Osher model or the VO model. The results displayed in [VO03] were quite good, especially in separating texture from cartoon. Analytical results were also obtained in [VO03] (following Meyer's approach in [Mey01]):

$$f = u + v + w, \text{ with } \ v = \nabla \cdot \vec{\zeta}$$

$$\text{If } u = 0, \quad \text{then} \quad \|f - \nabla \cdot \vec{\zeta}\|_* \leq \frac{1}{2\lambda}; \tag{1.10}$$

$$\text{If } \vec{\zeta} = 0, \quad \text{then} \quad \|\nabla(u - f)\|_q \leq \frac{\mu}{2\lambda}, \text{ where } q = \frac{p}{p-1}; \tag{1.11}$$

$$\text{Both } u = 0, \vec{\zeta} = 0, \quad \Longleftrightarrow \quad \|f\|_* \leq \frac{1}{2\lambda}, \ \|\nabla f\|_q \leq \frac{\mu}{2\lambda}. \tag{1.12}$$

Yet another approximation to (1.9) was later constructed by Osher, Solé and

Vese [OSV03]

$$u = \underset{u \in BV(\Omega)}{\arg\min} \left\{ |u|_{BV} + \lambda \|\nabla \Delta^{-1}(f-u)\|_{L^2}^2 \right\}, \tag{1.13}$$

see [OSV03] for details. We call this variational problem the Osher-Solé-Vese model or the OSV model. The $(L^2)^2$ fitting term used in the ROF model is replaced by an $(H^{-1})^2$ fitting term. This is also an $f = u + v$ model. Compared to the BV+$(L^2)^2$ decomposition of the ROF model, we call this decomposition as BV+$(H^{-1})^2$. The resulting Euler-Lagrange equation is equivalent to:

$$\frac{1}{2\lambda}\Delta\left(\nabla \cdot \frac{\nabla u}{|\nabla u|}\right) = f - u = v$$

which is easy to solve, e.g. by using gradient descent method [OSV03]. This time v is $\frac{1}{2\lambda}$ times the Laplacian of the curvature of the level contours of u, the cartoon version of f.

Following [OSV03], we can easily show for this model that

$$\|\Delta^{-1} f\|_* \leq \frac{1}{2\lambda} \iff u = 0, v = f; \tag{1.14}$$

$$\|\Delta^{-1} f\|_* > \frac{1}{2\lambda} \iff \|\Delta^{-1} v\|_* = \frac{1}{2\lambda}, \int_\Omega (-\Delta^{-1} v) u = |u|_{BV} \|\Delta^{-1} v\|_*. \tag{1.15}$$

It has been found experimentally by looking at the error term v found for optimal choice of parameters that this model does a somewhat better job at denoising images than the ROF model (although there is more computational effort involved), but does not do as well in separating cartoon from texture as the Vese-Osher model [VO03]. See also [AC05] for an explanation of this phenomenon.

Additional work on a cartoon/texture decomposition was done, e.g., in [CE05, YGO05] using BV+L^1 decomposition, in [LV05a] using BV+BMO decomposition, in [LV05b] using BV+$H^{-s}(s > 0)$ decomposition, in [AAB05] using duality, in [SED05] using a combination of sparse representations and TV regularization, and in [KO05] using a saddle point formulation which is based on (1.8). One of the many reasons to separate cartoon from texture is to improve image inpainting algorithms. See [BVS03] for a successful approach to this and [BSC00] for a pioneering paper on this subject.

Using duality, Chambolle [Cha04] constructed an algorithm solving for v directly in a way that simplifies the calculations needed to solve (1.2), (1.9) and (1.13). Duality was also used in [CGM00, Car02] to solve (1.2). This will be discussed in Chapter 2.

We note that for each choice of the scale parameter λ there is a σ such that problem (1.2) is equivalent to the constrained minimization problem

$$u = \underset{u \in BV(\Omega)}{\arg\min} \left\{ |u|_{BV} \quad s.t. \quad \|f - u\|_{L^2}^2 = \sigma^2 \right\}. \tag{1.16}$$

Often one has a reasonable estimate of σ, whereas it is difficult to know how to choose λ in (1.2), which corresponds to the Lagrange multiplier for the noise constraint in (1.16). The original ROF paper [ROF92] used a projected gradient method to solve (1.16), in which λ is dynamically updated so that the constraint $\|f - u\|_{L^2}^2 = \sigma^2$ is kept throughout the evolution of the pseudo time-dependent Euler-Lagrange equation. However, in most models developed after the ROF model, which include the Vese-Osher model and the Osher-Solé-Vese model discussed above, λ is a fixed value constant. The scale-dependent property of ROF model was analyzed, e.g., in [SC03]. In this dissertation λ is also set as a fixed constant, which is independent of the spatial variable x and the time variable t. We will see that the results using the new procedures described in the Chapter 2 and Chapter 3 are invariably much better than the constrained denoising of ROF. The error is much smaller and the edges are sharper with the new models.

For example, a numerical experiment of such comparison is presented in Figure 2.2 in Chapter 2.

1.3 Contributions and Organizations of This Dissertation

As discussed above, for a (denoising) decomposition $f = u + v$ from the ROF model and many other variational methods, one will find some signal in the removed residual part v, and some smoothing of textures in u. This problem is referred to as the signal loss defect or contrast degradation (cf., e.g., [SC03]). The main work of this dissertation is to develop some techniques to solve this problem. Besides image restoration, these techniques can be generalized to many other inverse problems.

In Chapter 2 we design an iterative regularization procedure to improve ROF restoration and its generalizations. Instead of stopping after recovering the minimizer u in (1.2), we call this solution u_1 and use it to compute u_2, u_3, etc. This is done using the Bregman distance [Bre67] which will be defined in the context in Section 2.3.1. If we call $D^p(u, v)$ the Bregman distance between u and v associated to the functional J and its subgradient $p \in \partial J(v)$, the algorithm designed to improve (1.2) is:

$$u_k = \arg\min_{u \in BV(\Omega)} \left\{ D^{p_{k-1}}(u, u_{k-1}) + \lambda \|f - u\|_{L^2}^2 \right\}, \tag{1.17}$$

$p_{k-1} \in \partial J(u_{k-1})$ is unique or otherwise uniquely selected by our algorithm. The sequence $\{u_k\}$ is shown monotonically converges to f, the noisy image. However as k increases, for λ sufficiently small the values u_k also monotonically get closer to the true noise free image g, in the sense of the Bregman distance, until:

$$\|u_{\bar{k}} - f\|_{L^2} < \tau \|g - f\|_{L^2},$$

for any $\tau > 1$. The ideal situation is to take λ small and $\bar{k}$ large so that $\bar{k}\lambda$ converges to a critical time $\bar{t}$ at which the estimate above is satisfied. These results are generalized and made precise in Section 2.3.

Iterative procedures involving Bregman distance have been used before in signal processing algorithms , e.g. in [Cet89, Cet91]. There, and in all the other applications that we are aware of, the goal was to accelerate the computation of the solution to a fixed problem, e.g. to solve the ROF minimization equation (1.2). The approach probably closest to the iterative method is the one in [Cet89], where each iteration step consists in computing

$$u_k = \operatorname*{argmin}_u D^{p_{k-1}}(u, u_{k-1}) \qquad \text{subject to } \|Ku - f\|_{L^2} \leq \epsilon$$

for some $\epsilon > 0$. The difference however, is that for increasing iteration, the residual $\|Ku - f\|_{L^2}$ will, in general, not decrease further during the iteration and hence, the iteration procedure rather yields a smoothing of the solution than a closer approximation of the data f. Here the apparently novel idea is to replace the variational problem (1.2) by a sequence (1.17) so as to obtain an improved restoration, or indeed improved solution to a wide class of inverse problems. Another new aspect of the approach is that an iteration with a Bregman distance (in the generalized sense) is used, which is corresponding to a nondifferentiable functional, the total variation.

We note that previously in [SG01, TNV04] the authors constructed a sequence of approximations $\{u_k\}$ using ROF with a quite different approach, used more to decompose images than to restore them. This will be commented on in Section 2.3.5.

We will also show how the iterative regularization procedure can be used for other image restoration tasks, e.g. restoring blurry and noisy images, thus improving the results of [RO94]. The decomposition in this case becomes

$$f = Au + v$$

where A is a given compact operator, often a convolution using, e.g. a Gaussian kernel. If A is not known, this becomes a blind deconvolution problem. See [CW98] for an interesting approach to blind deconvolution, also minimizing functionals involving the BV seminorm. In [HMO05] He, Marquina and Osher generalized the iterative regularization idea to blind deconvolution and obtained impressive results.

The iterative regularization procedure yields a sequence of convex variational problems, evolving towards the noisy image. In Chapter 3, by reinterpreting the scale parameter λ as a pseudo time-step Δt tending to zero, while the number of iteration steps tends to infinity, we develop an inverse scale space flow. We obtain similar properties for this new flow as for the iterative regularization procedure. Specifically, when a discrepancy principle is used as the stopping criterion, the error between the reconstruction and the noise-free image decreases until termination, even if only the noisy image is available and a bound on the variance of the noise is known.

The inverse flow is computed directly for one-dimensional signals, yielding high quality restorations. In higher spatial dimensions, we introduce a relaxation technique using two evolution equations, which we call the relaxed inverse scale space flow. These equations allow fast, accurate, efficient and straightforward implementation. We investigate the properties of these new types of flows and show their excellent denoising capabilities, wherein noise can be well removed with minimal loss of contrast of larger objects.

Another important technique for image restoration is wavelet shrinkage (cf., e.g.,[Don95, DJ94, DJ95]). The relation between the total variation regularization and wavelet shrinkage techniques has been studied by several authors, e.g., in [CDL98, SWB04, SED05, DT05]. In the last part of this dissertation we extend the iterative regularization and inverse scale space methods to wavelet-based image restoration and obtain promising results. This is in Chapter 4.

The rest of this dissertation is organized as follows. Chapter 2 introduces the iterative regularization method which has been presented in part in [OBG05]. Chapter 3 presents the inverse scale space flow, along with its relaxation version – relaxed inverse scale space flow which has been presented in part in [BOX05, BGO06]. Both chapters focus on the variational based image restoration. In Chapter 4 we extend these two methods to wavelet based denoising which has been presented in part in [XO06].

CHAPTER 2

Iterative Regularization Method

2.1 Introduction

In this chapter an iterative regularization procedure for inverse problems based on the use of Bregman distances [Bre67] is introduced, with particular focus on problems arising in image processing. The idea is motivated by the problem of restoring noisy and blurry images via variational methods, by using total variation regularization. Rigorous convergence results and effective stopping criterion for the general procedure are obtained. The numerical results for denoising appear to give significant improvement over standard models and preliminary results for deblurring/denoising are very encouraging.

2.2 Using Geometry and Iterative Regularization

The work in this chapter has several immediate antecedents. In [TWB03] Tasdizen, Whitaker, Burchard and Osher processed deformable surfaces via the level set method [OS88]. The idea used was to

(a) first process the unit normals to a given initial surface;

(b) deform the surface so as to simultaneously process it and fit it to the previously computed surface.

The results were visually very pleasing, but no detailed theoretical analysis has been obtained yet.

In [LOT04] Lysaker, Osher and Tai borrowed the basic idea discussed above and applied it to images as follows: (for implementation details and formal analysis, see [LOT04]).

- *Step 1:* Given a noisy f, smooth the unit normal vectors $\vec{n}_0 = \frac{\nabla f}{|\nabla f|}$ by using a one-harmonic map as in [VO02]

$$\vec{n}_1 = \underset{|\vec{n}|=1}{\arg\min} \left\{ \int_\Omega |\nabla \vec{n}| + \mu \int_\Omega |\vec{n} - \vec{n}_0|^2 \right\}, \tag{2.1}$$

  where the scale parameter $\mu > 0$.

- *Step 2:* Replace the ROF minimization (1.2) by

$$u_2 = \underset{u \in BV(\Omega)}{\arg\min} \left\{ \int_\Omega (|\nabla u| - \vec{n}_1 \cdot \nabla u) + \lambda \int_\Omega (f - u)^2 \right\}.$$

This minimization procedure attempts to match normals as well as grey level values.

Unlike all the other methods discussed in this dissertation, the minimization problem in step 1 is not convex and it does not produce an image u_1 satisfying $\vec{n}_1 = \frac{\nabla u_1}{|\nabla u_1|}$. Because of this nonconvexity, we decided here to compute $\vec{n}_1$ by using ROF itself:

- *Step 1:* First solve the ROF model to obtain

$$u_1 = \underset{u \in BV(\Omega)}{\arg\min} \left\{ \int_\Omega |\nabla u| + \lambda \int_\Omega (f - u)^2 \right\}.$$

  Then define $\vec{n}_1 = \frac{\nabla u_1}{|\nabla u_1|}$.

- *Step 2:* Perform a correction step to obtain

$$u_2 = \operatorname*{arg\,min}_{u \in BV(\Omega)} \left\{ \int_\Omega (|\nabla u| - \vec{n}_1 \cdot \nabla u) + \lambda \int_\Omega (f-u)^2 \right\}.$$

Then we make the following obvious, but crucial, observation:

$$-\int_\Omega \vec{n}_1 \cdot \nabla u = \int_\Omega u \nabla \cdot \vec{n}_1 = \int_\Omega u \left(\nabla \cdot \frac{\nabla u_1}{|\nabla u_1|} \right).$$

But, from the Euler-Lagrange equations for ROF, we have:

$$\nabla \cdot \frac{\nabla u_1}{|\nabla u_1|} = -2\lambda(f - u_1) = -2\lambda v_1$$

(recall $f = u_1 + v_1$), and hence $-\int_\Omega \vec{n}_1 \cdot \nabla u = -\int_\Omega 2\lambda u v_1$.

We can thus rewrite Step 2 as

$$\begin{aligned} u_2 &= \operatorname*{arg\,min}_{u \in BV(\Omega)} \left\{ \int_\Omega |\nabla u| + \lambda \int_\Omega [(f-u)^2 - 2uv_1] \right\} \\ &= \operatorname*{arg\,min}_{u \in BV(\Omega)} \left\{ \int_\Omega |\nabla u| + \lambda \int_\Omega (f + v_1 - u)^2 - \lambda \int_\Omega (v_1^2 + 2v_1 f) \right\}. \end{aligned}$$

Since the last integral above is independent of u, we have

$$u_2 = \operatorname*{arg\,min}_{u \in BV(\Omega)} \left\{ \int_\Omega |\nabla u| + \lambda \int_\Omega (f + v_1 - u)^2 \right\}.$$

Remarkably, we are led to the concept that v_1, the "noise" computed by the ROF procedure should be *added* back to f, the original noisy image, and the sum then processed by the ROF minimization procedure.

2.2.1 Iterative Regularization: Total Variation Denoising

Clearly the above regularization process can be repeated. Moreover, the first step can be put into this iterative framework by choosing initial values $u_0 = 0$ and $v_0 = 0$. We shall give precise reasons why this is a good procedure, using the concept of Bregman distance [Bre67, CT93] from convex programming, in the next section. Specifically, we are proposing the following *iterative regularization* procedure:

Algorithm 2.1 (Iterative Regularization on ROF Denoising).

- *Initialize* $v_0 = 0$*;*
- *For k=0, 1, 2,* $\cdots$ *: compute* u_{k+1} *as a minimizer of the modified ROF model with new image* $f + v_k$*, i.e.,*

$$u_{k+1} = \underset{u \in BV(\Omega)}{arg\min} \left\{ |u|_{BV} + \lambda \|f + v_k - u\|_{L^2}^2 \right\} \tag{2.2}$$

  and update

$$v_{k+1} = v_k + f - u_{k+1}. \tag{2.3}$$

We certainly need a stopping criterion, which gives some information for which k we would obtain an approximation as close as possible to the true noise-free image g. In the next section we shall show that the *discrepancy principle* is a reasonable stopping rule; it consists in stopping the iterative procedure the first time the residual $\|u_k - f\|_{L^2}$ is of the same size as the noise level σ. We will prove that some distance measure between the iterate u_k and the true image g decreases monotonically until the stopping index $\bar{k}$ is reached and that the regularization procedure enjoys the usual *semiconvergence* properties of iterative regularization methods; i.e., the reconstructed image $u_{\bar{k}}$ obtained at the stopping index converges to the true noise-free image as the noise level tends to zero (in a stronger topology than the one of $L^2(\Omega)$). Note that if we do not stop the iteration properly, the iterates would just converge to the noisy image f in $L^2(\Omega)$, and the total variation of the iterates could become unbounded, which is clearly undesirable.

These facts indicate that, for denoising f, a good strategy is to proceed iteratively until the result gets noisier, say, until u_{k+1} is more noisy than u_k. Of course, if we happen to have a good estimate of σ, we can use the discrepancy principle.

It is interesting to further understand how the iterative procedure (2.2), (2.3) works. If we consider why u_2 might contain more signal than u_1 we have

$$u_2 = f + v_1 - v_2 = u_1 + 2v_1 - v_2.$$

This implies that for u_2 to be less noisy than u_1, we need $2v_1 - v_2$ to have more signal than noise. This is indeed the case if the stopping index is greater than one.

It is also clear that the results depend on λ. If λ is very large, we may approximate the noisy image too much, and the stopping index may be satisfied already after the first step. In such a case we may expect a bad reconstruction. If λ is small we oversmooth initially and can make sure that the stopping index is not satisfied after one step. The numerical results at the end of this chapter will confirm that the images u_k, $k = 1, 2, \cdots$, become less blurry and noisy until the stopping index is reached. Later they eventually become noisy, converging to the original noisy image f. In the numerical experiments we also found out that if λ is sufficiently small, a further decrease does not have a large impact on the final reconstruction. Roughly speaking, by dividing λ by two, the number of iterations needed until the stopping index is reached doubles, and the final reconstruction is almost the same. This fact induces the conjecture that there exists a limiting flow of images on which this iterative procedure can be interpreted as an implicit time discretization with time step λ. If this is the case, then the dependence of the results on λ is somehow one-sided, i.e., only too large large values of λ will create bad reconstructions. This is true. In Chapter 3, λ will be interpreted as timestep and the discrete iterative regularization procedure will be developed to a continuous inverse scale flow.

Example. It is instructive to see what this procedure does to the the specific clean image mentioned in Section 1.2:

$$f = \alpha\chi_R = \begin{cases} \alpha, & \text{if } \sqrt{x^2+y^2} \leq R; \\ 0, & \text{if } \sqrt{x^2+y^2} > R. \end{cases} \tag{2.4}$$

If $\alpha\lambda R \geq 1$, Meyer's result [Mey01] gives us

$$f = u_1 + v_1 = (\alpha - \frac{1}{\lambda R})\chi_R + \frac{1}{\lambda R}\chi_R.$$

Then

$$f + v_1 = \left(\alpha + \frac{1}{\lambda R}\right)\chi_R = \alpha\chi_R + \frac{1}{\lambda R}\chi_R = u_2 + v_2.$$

This follows because we merely replace α by $\alpha + \frac{1}{\lambda R}$ in the equation above. So $u_2 = f$, as do all the $u_k,\ k \geq 2$. The objection that ROF degrades clean images by shrinking extrema is no longer valid. Figure 2.1 illustrates this result.

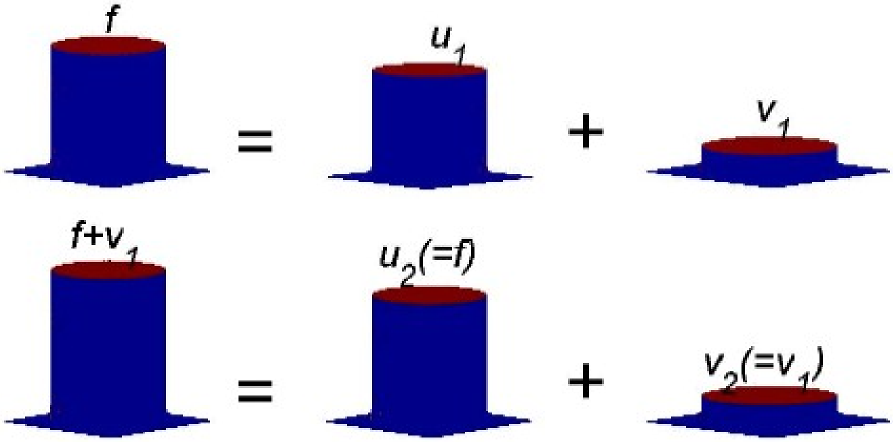


Figure 2.1: (Example) ROF with a two-step iterative regularization on a characteristic function. ROF degrades clean image, and this is resolved by the iterative regularization.

If $\alpha\lambda R < 1$, we have

$$\begin{aligned} f = u_1 + v_1 &= 0 + \alpha\chi_R, \\ f + v_1 &= 2\alpha\chi_R. \end{aligned}$$

Let n be the smallest integer for which $n\alpha\lambda R \geq 1$. We have $u_{n-1} = 0, v_{n-1} = (n-1)\alpha\chi_R$. But $u_n = \left(n\alpha - \frac{1}{\lambda R}\right)\chi_R$, $v_n = \frac{1}{\lambda R}\chi_R$. Finally, $u_{n+1} = f$, as do all u_k for $k \geq n+1$. This illustrates the strongly nonlinear nature of this iterative procedure. We go from a sequence of "restored" images, all of which are totally black, to the true result in two steps.

The above results also apply to the radially symmetric piecewise constant image f in (2.4) if radially symmetric noise that is not too large is added to it. This follows from an analysis of the ROF model by Strong and Chan [SC03]. Strong and Chan present numerical results that show that their analytical results predict quite well the actual performance of ROF even on digital images with no radial symmetry.

Chambolle [Cha04] has shown that the problem dual to the TV regularization (restoration) problem (1.2) is:

$$v = \underset{p\in\mathcal{K}}{\operatorname{argmin}}\left\{ \|p - f\|_{L^2}^2 \right\}, \tag{2.5}$$

where

$$\mathcal{K} := \mathrm{cl}\left\{ \frac{1}{2\lambda}\nabla\cdot\vec{\xi}\;\middle|\;\vec{\xi}\in\mathcal{C}_c^1(\Omega,\mathbb{R}^2), |\vec{\xi}(x)|\le 1, \forall x\in\Omega \right\}, \tag{2.6}$$

with closure taken in the space G, i.e., v is a projection of f onto the convex set $\mathcal{K}$. A simple alternative proof of this in the finite dimensional case can be found in [GY04]. This minimization problem determines the "noise" v in f, whereas the minimization problem (1.2) determines the "signal" $u = f - v$ in f. The dual version of the iterative regularization procedure (2.2) and (2.3) becomes:

Algorithm 2.2 (Dual Form of Iterative Regularized ROF Denoising)**.**

- *Initialize:* $v_0 = 0$*;*
- *For k=0,1,2,$\cdots$: compute* v_{k+1} *as the minimizer of the modified dual problem, i.e.,*

$$v_{k+1} = \underset{p\in\mathcal{K}}{\mathit{argmin}}\left\{ \|p - (f + v_k)\|_{L^2}^2 \right\}. \tag{2.7}$$

Note that

$$u_{k+1} = (f + v_k) - v_{k+1}. \tag{2.8}$$

Although this procedure will not be used in this dissertation, it is included here for its simplicity and elegance.

Note that had the dual iterates v_k and the update (2.3) for them not been introduced, the expression (2.2) for u_{k+1}, in terms of only the primal iterates u_k, would have had the much more complicated form

$$u_{k+1} = \underset{u \in BV(\Omega)}{\arg\min} \left\{ |u|_{BV} + \lambda \left\| (k+1)f - \sum_{j=0}^{k} u_j - u \right\|_{L^2}^2 \right\}, \text{ for } k = 0, 1, \cdots . \quad (2.9)$$

where $u_0 = 0$.

2.2.2 Iterative Regularization: General Case

The above regularization procedure can be generalized to other inverse problems and other regularization models, as to be outlined in the following and detailed in the succeeding chapters. Specifically Algorithm 2.1 can be generalized to regularization models of the form

$$\min_u \left\{ J(u) + H(u, f) \right\}, \quad (2.10)$$

where J is a convex nonnegative regularization functional (for total variation regularization we have $J(u) = \int_\Omega |\nabla u|$) and the fitting functional H is convex nonnegative with respect to u for fixed f. As usual for convex functionals (cf. [ET99]) we shall denote the subdifferential of J at a point u by

$$\partial J(u) := \left\{ p \in BV(\Omega)^* \;\middle|\; J(v) \geq J(u) + \langle p, v - u \rangle, \ \forall\, v \in BV(\Omega) \right\}. \quad (2.11)$$

After initializing $u_0 = 0$ and $p_0 = 0 \in \partial J(u_0)$, the iterative procedure is given by the sequence of variational problems

$$u_k = \underset{u}{\arg\min} \left\{ J(u) - \langle u, p_{k-1} \rangle + H(u, f) \right\} \quad (2.12)$$

for $k = 1, 2, \cdots$, where $\langle \cdot, \cdot \rangle$ denotes the standard duality product and p_{k-1} is a subgradient of J at u_{k-1}.

As particular examples we may consider the following:

- the Vese-Osher model (1.9), which also minimizes over $\vec{\zeta}$ at each step. The k^{th} step yields the decomposition $f + w_k = u_{k+1} + \nabla \cdot \vec{\zeta}_{k+1} + w_{k+1}$, with $w_0 = 0$, via the minimization problem (1.9), with f replaced by $f + w_k$ for $k \geq 0$.

- The Osher-Solé-Vese model (1.13), where we merely decompose $f = u_1 + v_1$ and iterate via $f + v_k = u_{k+1} + v_{k+1}$ for $k \geq 1$.

In principle, the iteration procedure can be written down for arbitrary functionals H and J, but the well-definedness of the algorithm is not obvious since one needs the existence of u_k as the minimizer of a variational problem and the existence of an associated subgradient p_k for the next step. This will introduce some conditions on J and H that will be discussed in further detail below.

2.3 Analysis of the Iterative Regularization Procedure

In the following a detailed analysis will be provided for the most important case of functionals which are interested in this dissertation, namely

$$J(u) := |u|_{BV} \tag{2.13}$$

and

$$H(u, f) := \lambda \|f - Ku\|_{L^2}^2, \tag{2.14}$$

with $K : L^2(\Omega) \to \mathcal{H}$ being a bounded linear operator whose kernel does not include the space of continuous functions, and $\mathcal{H}$ being some Hilbert space. In this case it is easy to see that $N(u) := J(u) + \sqrt{H(u,0)}$ is an equivalent norm on $BV(\Omega)$. The case of more general J and H will be discussed in Section 2.3.4. For quadratic H we can use Fréchet-derivatives instead of subgradients, they are given by

$$\partial_u H(\cdot, f) = 2\lambda K^*(Ku - f),$$

where K^* denotes the adjoint of K.

Note that due to the definition of $H(\cdot, f)$ on the larger space $L^2(\Omega)$, its gradients can be considered as elements of this space, too, while the gradients of J are in the larger space $BV(\Omega)^*$, in general. This will have some interesting implications on the regularity of subgradients of $J(u_k)$ we obtain through the iterative minimization procedure. Moreover, note that we can extend J to a convex functional on $L^2(\Omega)$ by setting $J(u) = \infty$ for $u \in L^2(\Omega)\backslash BV(\Omega)$. The identity

$$\partial_u(J + H(\cdot, f)) = \partial J + \partial_u H(\cdot, f)$$

holds (in $BV(\Omega)^*$) for any $f \in L^2(\Omega)$. A detail proof of this assertion is referred to [ET99, Proposition 5.6].

The general iterative regularization procedure can be formulated as follows:

Algorithm 2.3 (General Iterative Regularization)**.** *Let $u_0 = 0$, $p_0 = 0$, and for $k = 1, 2, \cdots,$*

- *Compute u_k as a minimizer of the convex functional*

$$Q_k(u) := J(u) - J(u_{k-1}) - \langle p_{k-1}, u - u_{k-1}\rangle + H(u, f), \tag{2.15}$$

  where $\langle \cdot, \cdot \rangle$ denotes the usual duality product.

- *Compute $p_k = p_{k-1} - \partial u H(u_k, f) \in \partial J(u_k)$.*

Note that in principle one could also start with different initial values that satisfy $p_0 \in \partial J(u_0)$. Since for $u_0 \neq 0$ an analytic expression for the subgradient is not known, one would have to solve another complicated optimization problem to determine p_0, which seems to be not desirable from a practical standpoint.

2.3.1 Iterative Regularization and Bregman Distances

Before considering the well-definedness of the above Algorithm 2.3, the connection to Bregman distances will be established first. For $p(v) \in \partial J(v)$, we define the (nonnegative) quantity

$$D^{p(v)}(u,v) \equiv D_J^{p(v)}(u,v) \equiv J(u) - J(v) - \langle p(v), u - v \rangle, \tag{2.16}$$

which is known as a generalized Bregman distance associated with $J(\cdot)$ (cf. [Bre67, CT93, Kiw97] for an extension to nonsmooth functions). For simplicity, we will drop the dependence on $J(\cdot)$ from the notation $D_J^{p(v)}(u,v)$ in the following. Moreover, if not otherwise specified, $D_J^p(u,v) \equiv D_J^{p(v)}(u,v)$.

For a continuously differentiable functional there is a unique element in the subdifferential and consequently a unique Bregman distance. In this case the distance is just the difference at the point u between $J(\cdot)$ and the first-order Taylor series approximation to $J(\cdot)$ at v. Moreover, if $J(u)$ is strictly convex, $D^p(u,v)$ is also strictly convex in u for each v, and as a consequence $D^p(u,v) = 0$ if and only if $u = v$.

Even for a continuously differentiable and strictly convex functional, the quantity $D^p(u,v)$ is not a distance in the usual (metric) sense, since in general, $D^{p(v)}(u,v) \neq D^{p(u)}(v,u)$ and the triangle inequality does not hold. However, it is a measure of closeness in the sense that $D^p(u,v) \geq 0$ and $D^p(u,v) = 0$ if

$u = v$ (if and only if for strictly convex functionals). For the case of a nonsmooth and not strictly convex functional like the total variation, it is not clear if one can introduce a Bregman distance for arbitrary u and v, since $\partial J(v)$ might be empty or multivalued. However, one can consider a multivalued version of the Bregman distance in this case, i.e., as the set including all $D^p(u,v)$ for all $p \in \partial J(v)$. As we shall prove below, this issue is not important for our purpose, since the iterative regularization algorithm automatically selects a unique subgradient.

Here we give some examples of Bregman distance:

- $J(u) = \frac{1}{2}\|u\|_{L^2}^2 = \frac{1}{2}\int_\Omega |u|^2$, then $p = \partial J(v) = v$,

$$\begin{aligned} D^p(u,v) &= \frac{1}{2}\int_\Omega |u|^2 - \frac{1}{2}\int_\Omega |v|^2 - \langle u - v, v\rangle \\ &= \frac{1}{2}\int_\Omega |u-v|^2 = J(u-v) = J(v-u) \\ &= D^{p(u)}(v,u). \end{aligned}$$

- $J(u) = \frac{1}{2}\|\nabla u\|_{L^2}^2 = \frac{1}{2}\int_\Omega |\nabla u|^2$, assuming u, v satisfy Neumann boundary conditions on boundaries, then $p = \partial J(v) = -\Delta v$,

$$\begin{aligned} D^p(u,v) &= \frac{1}{2}\int_\Omega |\nabla u|^2 - \frac{1}{2}\int_\Omega |\nabla v|^2 - \langle u - v, -\Delta v\rangle \\ &= \frac{1}{2}\int_\Omega |\nabla u|^2 - \frac{1}{2}\int_\Omega |\nabla v|^2 - \langle \nabla u - \nabla v, \nabla v\rangle \\ &= \frac{1}{2}\int_\Omega |\nabla u - \nabla v|^2 = J(u-v) = J(v-u) \\ &= D^{p(u)}(v,u). \end{aligned}$$

- Total variation $J(u) = |u|_{BV} = \int_\Omega |\nabla u|$, assuming u, v satisfy Neumann boundary conditions on boundaries, then $p = \partial J(v) = -\nabla \cdot \frac{\nabla u}{|\nabla u|}$,

$$\begin{aligned} D^p(u,v) &= \int_\Omega |\nabla u| - \int_\Omega |\nabla v| - \left\langle u - v, -\nabla \cdot \frac{\nabla u}{|\nabla u|} \right\rangle \\ &= \int_\Omega |\nabla u| - \int_\Omega |\nabla v| - \left\langle \nabla u - \nabla v, \frac{\nabla v}{|\nabla v|} \right\rangle \\ &= \left\langle \nabla u, \frac{\nabla u}{|\nabla u|} - \frac{\nabla v}{|\nabla v|} \right\rangle. \end{aligned}$$

Note that in this case $D^{p(v)}(u,v) \not\equiv D^{p(u)}(v,u)$. In Section 3.2.2 we will revisit this Bregman distance and show that if $D^p(u,v) = 0$ for $J(u) = |u|_{BV}$, then u and v are the same up to a contrast change.

Remark. *It is worthy to point out here that for total variation based image restoration, $J(u) = |u|_{BV}$, the Bregman distance $D^p(u,v)$ defined above characterizes the distance between the two normal vectors $\frac{\nabla u}{|\nabla u|}$ and $\frac{\nabla v}{|\nabla v|}$. Thus in the iterative regularization procedures, we are not only smoothing the grey levels of the image, but also fitting the normal vectors of restored images to a "smooth" one. In this sense we are able to expect refined and better results in the iterates than the original models.*

As we shall see below, we shall obtain convergence of the reconstructions in the weak-$*$ topology of $BV(\Omega)$ (and by compact embedding also in $L^2(\Omega)$), which is the same kind of convergence one obtains for the reconstructions of the ROF model (cf. [AV94]). From this viewpoint one may consider the Bregman distance only as an auxiliary term used in the convergence analysis. However, monotone decrease of some Bregman distances between the true image and the computed reconstruction will also be obtained. This may be interpreted as an additional indicator of the quality of the reconstruction, though the meaning of the Bregman distance associated with the total variation is difficult to interpret.

However, at least for some cases the convergence of Bregman distances can be used to interpret the convergence speed of discontinuities (cf. [BO04]).

2.3.2 Well-Definedness of the Iterates

In the following it will be shown that the iterative procedure in Algorithm 2.3 is well-defined, i.e., that Q_k has a minimizer u_k and that one may find a suitable subgradient p_k. The latter will be obtained from the optimality condition for the minimization of Q_k, which yields an interesting decomposition of f involving "noise" at levels k and $k-1$ and signal at each level k:

Proposition 2.3.1. *Assume that J and H are given by* (2.13) *and* (2.14)*, respectively, and let $u_0 = 0$ and $p_0 := 0 \in \partial J(u_0)$. Then, for each $k \in \mathbb{N}$ there exists a minimizer u_k of Q_k and there exists a subgradient $p_k \in \partial J(u_k)$ and $q_k = \partial_u H(u_k, f) = 2\lambda K^*(u_k - f)$ such that*

$$p_k + q_k = p_{k-1}. \tag{2.17}$$

If, in addition, K has no nullspace, then the minimizer u_k is unique.

Proof. We prove the above result by induction. For $k = 1$, we have $Q_1(u) = J(u) + H(u, f)$ and the existence of minimizers as well as the optimality condition $p_1 + q_1 = p_0 = 0$ is well-known [AV94]. Moreover, with $r_1 := 2\lambda(f - Ku_1) \in \mathcal{H}$ we have $p_1 = K^* r_1$.

Now we proceed from $k-1$ to k, and assume that $p_{k-1} = K^* r_{k-1}$ for $r_{k-1} \in \mathcal{H}$. Under the above assumptions, the functional

$$Q_k : u \mapsto J(u) - J(u_{k-1}) - \langle p_{k-1}, u - u_{k-1} \rangle + H(u, f)$$

is weak-* lower semicontinuous (due to convexity and local boundedness, cf. [ET99]) and it is bounded below by $H(u, f)$ due to the properties of subgradients. Moreover, we can estimate

$$\begin{aligned} Q_k(u) &= J(u) - J(u_{k-1}) - \langle r_{k-1}, f - Ku_{k-1} \rangle + \\ &\quad \lambda \| Ku - f - \frac{1}{2\lambda} r_{k-1} \|_{L^2}^2 - \frac{1}{4\lambda} \| r_{k-1} \|_{L^2}^2 \\ &\geq J(u) - J(u_{k-1}) - \langle r_{k-1}, f - Ku_{k-1} \rangle - \frac{1}{4\lambda} \| r_{k-1} \|_{L^2}^2 . \end{aligned}$$

Since only the first term on the right-hand side of this inequality is not constant, boundedness of $Q_k(u)$ implies boundedness of $J(u)$ and consequently boundedness of $N(u)$. This shows that the level sets of Q_k are bounded in the norm of $BV(\Omega)$, and therefore they are weak-* compact. Hence, there exists a minimizer of Q_k due to the fundamental theorem of optimization. Moreover, if K has no nullspace, the strict convexity of $H(\cdot, f)$ and convexity of the other terms implies the strict convexity of Q_k, and therefore the minimizer is unique. Since

$$\partial(-\langle p_{k-1}, \cdot \rangle) = \{-p_{k-1}\},$$

the optimality conditions for this problem imply

$$p_{k-1} \in \partial J(u_k) + \partial_u H(u_k, f),$$

which yields the existence of $p_k \in \partial J(u_k)$ and $q_k = \partial_u H(u_k, f) = 2\lambda K^*(Ku_k - f)$ satisfying (2.17). With $r_k := r_{k-1} - 2\lambda(Ku_k - f) \in L^2(\Omega)$ and $p_k := K^* r_k$ we obtain (2.17). $\square$

Note that as a result of (2.17) we obtain that

$$p_k = -\sum_{j=1}^{k} q_j = 2\lambda \sum_{j=1}^{k} K^*(f - Ku_j),$$

i.e., the subgradient p_k is equal to the adjoint applied the sum of residuals $f - Ku_j$. Moreover, the iterative algorithm 2.3 constructs a sequence of minimizers u_k such that there exists $p_k \in L^2(\Omega) \cap \partial J(u_k)$ (for smoothing K we even has p_k in the image of K^*), which can be thought of as a regularity property of u_k, respective of its level sets. This corresponds to results of Meyer [Mey01] for the ROF-model showing that the indicator function of a ball may be a solution, but not the indicator function of a square. In the same way we could show that the indicator function of a square (or more generally a function whose level sets are squares) cannot arise as an iterate in the regularization procedure. However, the method may still converge to such solutions as $k \to \infty$.

We again consider some special cases:

- *Denoising:* If

$$H(u, f) = \lambda \|f - u\|_{L^2} = \lambda \int_\Omega (f - u)^2, \tag{2.18}$$

 i.e., K is the identity, we have $\partial_u H(u, f) = 2\lambda(u - f)$ and hence,

$$p_k + 2\lambda(u_k - f) = p_{k-1}, \; k = 1, 2, \cdots, \quad p_0 = 0.$$

 If we set $p_k \equiv 2\lambda v_k$, we obtain the usual decomposition (2.3):

$$f + v_{k-1} = u_k + v_k.$$

- *Deblurring:* If

$$H(u,f) = \lambda \int_\Omega (Au - f)^2 \tag{2.19}$$

for $A : L^2(\Omega) \to L^2(\Omega)$ being a compact linear operator (typically a convolution operator), we have $\partial H_u(u,f) = 2\lambda(A^*(Au - f))$, where A^* is the L^2-adjoint operator of A, and hence,

$$p_k + 2\lambda A^*(Au_k - f) = p_{k-1}.$$

Notice that since $p_0 = 0 = A^*0$ we may conclude inductively that $p_k \in \mathcal{R}(A^*)$, and hence there exist v_k with $p_k = 2\lambda A^* v_k$. Hence, we can alternatively write an update formula for v_k given by

$$f + v_{k-1} = Au_k + v_k.$$

Finally one can see that the Osher-Sole-Vese model (1.13) can also be interpreted as deblurring, with the compact operator $A = \nabla \Delta^{-1}$ and transformed output $\tilde{f} = Af$.

- *Binary Denoising:* Another interesting example uses

$$H(u,f) = \lambda \int_\Omega |f - u| \tag{2.20}$$

(cf. [CE05]), which is not everywhere differentiable. In this case the subgradient of H with respect to u at u_k consists of all functions $q \in L^\infty(\Omega)$ such that

$$q(x) \in \lambda \text{sign}[u_k(x) - f(x)] \quad \text{a.e.} x \in \Omega$$

with the generalized sign function

$$\operatorname{sign}[t] = \begin{cases} \{1\} & \text{if } t > 0, \\ \{-1\} & \text{if } t < 0, \\ [-1, 1] & \text{if } t = 0. \end{cases}$$

Setting $p_k = \lambda v_k$ we have the decomposition

$$v_k \in \{v_{k-1}\} + \operatorname{sign}[f - u_k] = \begin{cases} \{v_{k-1}(x) + 1\} & \text{a.e. where } f(x) > u_k(x), \\ \{v_{k-1}(x) - 1\} & \text{a.e. where } f(x) < u_k(x), \\ \{v_{k-1}(x)\} + [-1, 1] & \text{a.e. where } f(x) = u_k(x). \end{cases}$$

This is of interest in the study of binary (black and white) images [CE05], where the noise structure is also binary, which is well reflected by the fact that the noise is changed by a unit amount if $f(x) \neq u_k(x)$.

In this dissertation we will mainly focus on the denoising case. For the generalizations on the other two cases, one can refer to the work of He et. al. in [HMO05] (for blind deconvolution) and [HBO05].

2.3.3 Convergence Analysis

Now some convergence properties of the iterative regularization process will be studied. The analysis below is motivated by that of Hanke [Han97] who analyzed Levenberg-Marquardt methods for ill-posed problems (also related to nonstationary iterative Tikhonov regularization, cf. [HG98, GS00] and inverse scale space methods cf. [SG01]), which turns out to be a special case of the iterative regularization strategy when using a quadratic regularization functional $J(u) = \|u\|_{L^2}$ for some Hilbert space norm.

First, two important monotonicity properties of the residual and of the Bregman distance are shown below:

Proposition 2.3.2. *Under the above assumptions, the sequence $H(u_k, f)$ obtained from the iterates of Algorithm 2.3 is monotonically non-increasing, we even have*

$$H(u_k, f) \leq H(u_k, f) + D^{p_{k-1}}(u_k, u_{k-1}) \leq H(u_{k-1}, f). \tag{2.21}$$

Moreover, let u be such that $J(u) < \infty$, then we have

$$D^{p_k}(u, u_k) + D^{p_{k-1}}(u_k, u_{k-1}) + H(u_k, f) \leq H(u, f) + D^{p_{k-1}}(u, u_{k-1}). \tag{2.22}$$

Proof. From the definition of subgradient and because u_k minimizes $Q_k(u)$ we have

$$\begin{aligned} H(u_k, f) &\leq H(u_k, f) + J(u_k) - J(u_{k-1}) - < u_k - u_{k-1}, p_{k-1} > \\ &= Q_k(u_k) \leq Q_k(u_{k-1}) = H(u_{k-1}, f), \end{aligned}$$

which implies (2.21).

Next we use the following interesting identity for Bregman distances, which seems to have first been pointed out in [CT93]:

$$
\begin{aligned}
& D^{p_k}(u,u_k) - D^{p_{k-1}}(u,u_{k-1}) + D^{p_{k-1}}(u_k,u_{k-1}) \\
= \quad & J(u) - J(u_k) + < u_k - u, p_k > \\
& -J(u) + J(u_{k-1}) - < u_{k-1} - u, p_{k-1} > \\
& +J(u_k) - J(u_{k-1}) + < u_{k-1} - u_k, p_{k-1} > \\
= \quad & < u_k - u, p_k - p_{k-1} > .
\end{aligned}
$$

Replacing $p_k - p_{k-1}$ by $-q_k$ using equation (2.17) and using the fact that q_k is a subgradient of $H(\cdot, f)$ at u_k, we obtain

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

which is equivalent to (2.22). □

If there exists a minimizer $\tilde{u}$ of $H(\cdot, f)$ with $J(\tilde{u}) < \infty$, then we obtain in particular from the choice $u = \tilde{u}$ in (2.22),

$$
\begin{aligned}
D^{p_k}(\tilde{u},u_k) \quad &\leq \quad D^{p_k}(\tilde{u},u_k) + D^{p_{k-1}}(u_k,u_{k-1}) \\
&\leq \quad D^{p_k}(\tilde{u},u_k) + D^{p_{k-1}}(u_k,u_{k-1}) + H(u_k,f) - H(\tilde{u},f) \\
&\leq \quad D^{p_{k-1}}(\tilde{u},u_{k-1}). \qquad (2.23)
\end{aligned}
$$

This result allows us to conclude a general convergence theorem:

Theorem 2.3.3 (Exact Data)**.** *Assume that there exists a minimizer* $\tilde{u} \in BV(\Omega)$

of $H(\cdot, f)$ such that $J(\tilde{u}) < \infty$. Then

$$H(u_k, f) \leq H(\tilde{u}, f) + \frac{J(\tilde{u})}{k} \tag{2.24}$$

and in particular u_k is a minimizing sequence.

Moreover, u_k has a weak- convergent subsequence in $BV(\Omega)$, and the limit of each weak-* convergent subsequence is a solution of $Ku = f$. If $\tilde{u}$ is the unique solution of $Ku = f$, then $u_k \to \tilde{u}$ in the weak-* topology in $BV(\Omega)$.*

Proof. We now sum (2.22) arriving at

$$D^{p_k}(\tilde{u}, u_k) + \sum_{\nu=1}^{k} [D^{p_{\nu-1}}(u_\nu, u_{\nu-1}) + H(u_\nu, f) - H(\tilde{u}, f)] \leq D^0(\tilde{u}, u_0) = J(\tilde{u}). \tag{2.25}$$

From $D^{p_{\nu-1}}(u_\nu, u_{\nu-1}) \geq 0$ and the monotonicity of $H(u_\nu, f)$ due to (2.21), we further conclude

$$D^{p_k}(\tilde{u}, u_k) + k\,[H(u_k, f) - H(\tilde{u}, f)] \leq J(\tilde{u}),$$

and the nonnegativity of the first term implies (2.24).

For $f = K\tilde{u}$, (2.24) implies (together with the monotonicity of $\|Ku_k - f\|_{L^2}^2$)

$$k\lambda\|Ku_k - f\|_{L^2}^2 \leq \lambda \sum_{\nu=1}^{k} \|Ku_\nu - f\|_{L^2}^2 \leq J(\tilde{u}).$$

From (2.24) and (2.17) we obtain

$$
\begin{aligned}
J(\tilde{u}) &\geq \sum_{\nu=1}^{k} D^{p_{\nu-1}}(u_\nu, u_{\nu-1}) = J(u_k) - \sum_{\nu=1}^{k} \langle p_{\nu-1}, u_\nu - u_{\nu-1} \rangle \\
&= J(u_k) - J(u_0) - \langle p_{k-1}, u_k - \tilde{u} \rangle + \sum_{\nu=1}^{k-1} \langle p_\nu - p_{\nu-1}, u_\nu - \tilde{u} \rangle \\
&= J(u_k) - \sum_{\nu=1}^{k-1} \langle q_\nu, u_k - \tilde{u} \rangle - \sum_{\nu=1}^{k-1} \langle q_\nu, u_\nu - \tilde{u} \rangle.
\end{aligned}
$$

Since $q_\nu = 2\lambda K^*(Ku_\nu - f)$ in this case, we may further estimate

$$
\begin{aligned}
J(\tilde{u}) &\geq J(u_k) - 2\lambda \sum_{\nu=1}^{k-1} \langle Ku_\nu - f, Ku_k - f \rangle - 2\lambda \sum_{\nu=1}^{k-1} \|Ku_\nu - f\|^2 \\
&\geq J(u_k) - k\lambda \|Ku_k - f\|^2 - 3\lambda \sum_{\nu=1}^{k-1} \|Ku_\nu - f\|^2 \\
&\geq J(u_k) - 4J(\tilde{u}).
\end{aligned}
$$

Thus, $J(u_k) \leq 5J(\tilde{u})$, and by equivalence of norms we obtain that

$$
|u_k|_{BV} \leq C(J(u_k) + \|Ku_k\|),
$$

whose right-hand side is uniformly bounded. The further assertions then follow by standard weak-* convergence techniques. □

The above result is a typical convergence result for exact data. In the special case of denoising it would mean that $f = \tilde{u}$ is of bounded variation, i.e., it does not include any noise that is not of bounded variation. For the specific models of denoising and deblurring considered above, this yields a rate of convergence:

Corollary 2.3.4. *Under the assumptions of Theorem 2.3.3, the following results* *hold:*

- *For denoising, i.e., for H given by* (2.18)*, we have*

$$\|f - u_k\|_{L^2} \leq \sqrt{\frac{J(f)}{\lambda k}} = \mathcal{O}\left(k^{-1/2}\right) \tag{2.26}$$

  if $J(f) < \infty$.

- *For deblurring, i.e., for H given by* (2.19)*, we have*

$$\|f - Au_k\|_{L^2} \leq \sqrt{\frac{J(\tilde{u})}{\lambda k}} = \mathcal{O}\left(k^{-1/2}\right) \tag{2.27}$$

  if $f = A\tilde{u}$ and $J(\tilde{u}) < \infty$.

- *For binary denoising, i.e., for H given by* (2.20)*, we have*

$$\|f - u_k\|_{L^1} \leq \frac{J(f)}{\lambda k} = \mathcal{O}\left(k^{-1}\right) \tag{2.28}$$

  if $J(f) < \infty$.

Next, we consider the noisy case; i.e., we suppose that g is the true noise-free image and that $\tilde{u}$ is a minimizer of $H(\cdot, g)$ with $H(\tilde{u}, g) = 0$, which satisfies

$$H(\tilde{u}, f) \leq \delta. \tag{2.29}$$

The positive number δ can be considered as the noise level (or rather as an estimate for the noise level, which is easier to obtain in practice). For example, in ROF model, $H(\tilde{u}, f) = \lambda\|u - f\|_{L^2}^2$ and then $\delta = \lambda\sigma^2$, where $\sigma = \|u - f\|_{L^2}$ is the one defined in (1.16). The meaning of δ will be specified for the special cases below.

In medical imaging, for example, one often has a very good estimate of the noise induced by the imaging apparatus obtained by imaging known objects (phantoms). In general, a procedure having been found satisfactory to estimate δ is to restrict the image to a square region which is "quiet" and contains no edges and compute the standard deviation of this restriction of the image.

Theorem 2.3.5 (Noisy Data)**.** *Let $\tilde{u}$, f and g be such that $\tilde{u}$ is a minimizer of $H(\cdot, g)$ and such that* (2.29) *holds. Then, as long as $H(u_k, f) > \delta$ (i.e., the residual lies above the noise level), the Bregman distance between u_k and $\tilde{u}$ is decreasing, more precisely,*

$$D^{p_k}(\tilde{u}, u_k) \leq D^{p_k}(\tilde{u}, u_k) + D^{p_{k-1}}(u_k, u_{k-1}) < D^{p_{k-1}}(\tilde{u}, u_{k-1}).$$

Proof. From (2.22) we obtain by inserting (2.29)

$$D^{p_k}(\tilde{u}, u_k) + D^{p_{k-1}}(u_k, u_{k-1}) + H(u_k, f) \leq \delta + D^{p_{k-1}}(\tilde{u}, u_{k-1}).$$

Thus, due to the non-negativity of $D^p(\cdot, \cdot)$, for $H(u_k, f) > \delta$ we may conclude the decrease of $D^{p_k}(\tilde{u}, u_k)$. □

Note that, due to Theorem 2.3.5, we can obtain that if g is the noise-free image and $\tilde{u}$ is the true solution, iterations actually approach the true solution until the residual in the iteration drops below the noise level.

The result of Theorem 2.3.5 yields a natural stopping rule, the so-called *generalized discrepancy principle* (cf. [EHN96]), which consists in stopping the iteration at the index $\bar{k} = \bar{k}(\delta, f)$ given by

$$\bar{k} = \max\{\ k \in \mathbb{N} \mid H(u_k, f) \geq \tau\delta\ \}, \tag{2.30}$$

where $\tau > 1$. Note that due to the monotone decrease of $H(u_k, f)$, which is guaranteed by (2.21), the stopping index $\bar{k}$ is well-defined. We also mention that the choice $\tau = 1$ that would seem obvious with respect to the noise estimate is too severe in order to guarantee the boundedness of $J(u_{\bar{k}})$ and the semiconvergence of the regularization method as we shall see below, but this statement is also true for other iterative regularization methods (cf. [EHN96]).

If we sum the inequality in the proof of Theorem 2.3.5, we obtain

$$kH(u_k, f) \leq D^{p_k}(\tilde{u}, u_k) + \sum_{\nu=1}^{k} H(u_\nu, f) \leq \delta k + J(\tilde{u}),$$

i.e.,

$$H(u_k, f) \leq \delta + \frac{J(\tilde{u})}{k}.$$

As a consequence, $\bar{k}(\delta)$ is finite for $\tau > 1$ and, since $H(u_{\bar{k}(\delta)+1}, f) \leq \tau\delta$ we have

$$\delta(\bar{k}(\delta) + 1) \leq \frac{J(\tilde{u})}{\tau - 1}. \tag{2.31}$$

Theorem 2.3.6 (Semi-Convergence for Noisy Data)**.** *Let the assumptions of Theorem 2.3.5 be satisfied and let the stopping index $\bar{k}$ be chosen according to (2.30). Moreover, let $K\tilde{u} = f$. Then, $J(u_{\bar{k}(\delta)})$ is uniformly bounded in δ and hence, as $\delta \to 0$ there exists a weak-* convergent subsequence $(u_{\bar{k}(\delta_\ell)})$ in $BV(\Omega)$. If the set $\{\bar{k}(\delta)\}_{\delta \in \mathbb{R}^+}$ is unbounded, the limit of each weak-* convergent subsequence is a solution of $Ku = g$.*

Proof. By analogous reasoning, as in the proof of Theorem 2.3.3, we can derive an estimate of the form

$$J(u_k) \leq C(J(\tilde{u}) + k\delta)$$

for $k \leq \bar{k}(\delta)$ and some positive constant C. From (2.31) we further obtain

$$J(u_{\bar{k}(\delta)}) \leq \frac{\tau C}{\tau - 1},$$

and hence, $J(u_{\bar{k}(\delta)})$ is bounded. The existence of converging subsequences then follows from standard weak-* convergence techniques. In order to show that a weak-* limit u satisfies $Ku = g$, we use again the estimate

$$H(u_{\bar{k}(\delta)}, f) \leq \delta + \frac{J(\tilde{u})}{k_*(\delta)}$$

derived above. If $\bar{k}(\delta_\ell) \to \infty$ for some subsequence δ_ℓ, then clearly $H(u_{\bar{k}(\delta)}, f) \to 0$ and from the lower semicontinuity of H in this case we obtain $H(u, g) = 0$ for the limit, i.e., $Ku = g$ for the special H we consider. □

We again consider this relation for the special cases:

- *Denoising:* for H given by (2.18) we obviously have $\tilde{u} = g$ and hence, (2.29) becomes

  $$H(g, f) = \lambda \int_\Omega (f - g)^2 \leq \delta.$$

  Thus, $\sigma = \sqrt{\frac{\delta}{\lambda}}$ is an estimate for the variance of the noise, which can be obtained from statistical tests in typical applications. The stopping rule consists in terminating the iteration when the residual $\|u - f\|_{L^2}$ drops below this variance estimate σ. For $k \leq \bar{k}$ we actually have the stronger estimate

  $$D^{p_k}(g, u_k) + D^{p_{k-1}}(u_k, u_{k-1}) + \lambda \left(1 - \frac{1}{\tau}\right) \|u_k - f\|_{L^2}^2 \leq D^{p_{k-1}}(g, u_{k-1}).$$

- *Deblurring:* for H given by (2.19) we have $A\tilde{u} = g$ and hence, (2.29) is again

$$H(\tilde{u}, f) = \lambda \int_{\Omega} (f - g)^2 \leq \delta,$$

and $\sigma = \sqrt{\frac{\delta}{\lambda}}$ is an estimate for the variance of the noise in the output image. For $k \leq \bar{k}$ we have

$$D^{p_k}(\tilde{u}, u_k) + D^{p_{k-1}}(u_k, u_{k-1}) + \lambda \left(1 - \frac{1}{\tau}\right) \|Au_k - f\|_{L^2}^2 \leq D^{p_{k-1}}(\tilde{u}, u_{k-1}).$$

Note that in the particular case of the Osher, Solé, Vese model [OSV03] in equation (1.13) we have $A = \nabla\Delta^{-1}$ and $f = \nabla\Delta^{-1} f_0$, where f_0 is the actual noisy image we obtain, and therefore the noise estimate is

$$\lambda \|\nabla\Delta^{-1}(f_0 - \tilde{u})\|_{L^2}^2 \leq \delta,$$

i.e., an estimate of the variance of the noise in the H^{-1}-norm is needed.

- *Binary Denoising:* for H given by (2.20) we again have $\tilde{u} = g$ and hence, if we square (2.29) then

$$H(g, f)^2 = \lambda^2 \left(\int_{\Omega} |f - g|\right)^2 \leq \delta^2.$$

Because $(\int_{\Omega} |f - g|)^2 \leq (\int_{\Omega} 1^2)(\int_{\Omega} |f - g|^2)$, if we assume Ω is bounded and $\int_{\Omega} 1 = 1$, then in this case $\sigma = \frac{\delta}{\lambda}$ is a (rough) estimate for the variance of the noise, the stopping rule is the same as for denoising, we stop when the residual drops below the variance estimate. In this case we obtain for $k \leq \bar{k}$

$$D^{p_k}(g, u_k) + D^{p_{k-1}}(u_k, u_{k-1}) + \lambda \left(1 - \frac{1}{\tau}\right) \|u_k - f\|_{L^2} \leq D^{p_{k-1}}(g, u_{k-1}).$$

2.3.4 Further Generalizations

In the following some possible generalizations of the above procedure with respect to the fitting functional H, the regularization functional J and additional constraints will be discussed.

We start with different regularization functionals J. The above analysis is not restricted to the space $BV(\Omega)$ and J being the BV seminorm. One can easily generalize the results to other locally bounded, convex, and nonnegative regularization functionals J defined on a Banach space $\mathcal{U} \subset L^2(\Omega)$. The conditions needed on J are that:

- The level sets $\{ u \in \mathcal{U} \mid J(u) \leq M \}$ are compact in $L^2(\Omega)$ (or any stronger topology than the one of $L^2(\Omega)$) for all $M \in \mathbb{R}$ and nonempty for $M > M_0 > 0$.
- J can be extended to a weakly lower semicontinuous functional from $L^2(\Omega)$ to $\mathbb{R} \cup \{+\infty\}$.

Under these conditions, then by similar reasoning as above, there exists a minimizer of the functional Q_k, which is the minimal property we need for the well-definedness of the iterative procedure. If in addition $J + H(\cdot, f)$ is strictly convex, then this minimizer is unique and we obtain a unique iterate u_k. From standard optimality theory for convex problems (cf. [ET99]), we may also conclude the decomposition (2.17) and the regularity $p_k \in L^2(\Omega) \subset \mathcal{U}^*$. The convergence analysis with the same stopping rule can be carried out as above, with the modification that the weak-* topology in BV has to be replaced by the topology in which the level sets of J are compact.

Possible generalizations of the regularization functional include:

- *Anisotropic Total Variation:* In order to obtain different minimizers like indicator functions of squares as minimizers, one can use anisotropic regularization functionals of the form

$$J(u) = \int_\Omega G(\nabla u)$$

with $G : \mathbb{R}^2 \to \mathbb{R}^+$ being a continuous one-homogeneous function (cf. [EO04]). An example of particular interest is $G(\nabla u) = |u_x| + |u_y|$. Of course, we can also use functions, which are not one-homogeneous like $G(\nabla u) = \|\nabla u\|_{L^2}^2$, thus including standard Tikhonov-type regularization techniques.

- *Approximations of Total Variation:* In several instances one rather minimizes the smooth approximation

$$J_\epsilon(u) = \int_\Omega \sqrt{|\nabla u|^2 + \epsilon^2},$$

for some small constant $\epsilon > 0$ (cf. e.g. [DS96]). Such an approximation simplifies numerical computations due to the differentiability of J_ϵ and may help to avoid the staircasing effect in some cases. The analysis can be carried out in the same way as above, and due to the strict convexity of J_ϵ for $\epsilon > 0$ one even obtains that the Bregman distance is a strict distance. In the numerical experiments of this chapter this regularized J_ϵ is used, instead of using $J(u) = |u|_{BV}$.

- *Bounded variation norms:* Instead of taking the seminorm in $BV(\Omega)$, one might also use a full norm for the regularization, i.e.,

$$J(u) = |u|_{BV} + \rho \|u\|_{L^2}^2 \tag{2.32}$$

for constant $\rho > 0$. In this case, the Bregman distance $D^p(u,v)$ is bounded below by $\rho\|u-v\|^2_{L^2}$, and hence convergence of the Bregman distance implies L^2-convergence, which is interesting in particular for deblurring and for more general fitting functionals as outlined below.

- *Derivatives of bounded variation:* another obvious generalization considered by several authors (cf. [CMM00, HS06]) is to use the bounded variation of ∇u, i.e.,

$$J(u) = \int_\Omega |D^2 u|,$$

where $D^2 u$ denotes the Hessian of u, or even more general functionals of the form

$$J(u) = \int_\Omega \varphi(u, \nabla u, D^2 u),$$

with convex $\varphi : \mathbb{R} \times \mathbb{R}^2 \times \mathbb{R}^{2\times 2} \to \mathbb{R}_+$. The analysis can be carried out in the Banach space $\mathcal{U} = BV^2(\Omega)$ of functions with second order bounded variation.

- *Maximum-Entropy Regularization:* A classical regularization functional in the reconstruction of probability distributions is the entropy (cf. [Egg93, EL93])

$$J(u) = -\ \mathrm{Entropy}(u) = \int_\Omega (u \ln u - u).$$

In this case the Bregman distance is the so-called *Kullback-Leibler divergence*

$$D^p(u,v) = \int_\Omega \Big(u \ln \frac{u}{v} - u + v\Big),$$

which is well-known in information theory and statistics. The analysis can be carried out in spaces of Radon measures.

- *Finite-dimensional approximations:* by analogous reasoning one can consider the discrete version of all the models introduced above and obtains the same type of convergence results.

For generalizations with respect to the fitting functional, the situation is more delicate. In general, even under rather strong assumptions on H, the compactness of level sets of the functional Q_k is not guaranteed, so that the iterates in Algorithm 2.3 are possibly not well-defined. Moreover, we do not know any argument showing that the total variation of u_k remains bounded (even for exact data), so that the convergence analysis cannot be carried out as above.

Finally, generalizations to additional constraints would be of interest in practice. The iterative procedure then consists in minimizing Q_k subject to the additional constraints. This is of importance e.g. for nonnegativity constraints or for for multiplicative noise, where one wants to choose

$$H(u, f) = \int_\Omega \left(\frac{f}{u}\right)^2$$

subject to the constraint (cf. [RLO03])

$$C(u) = -1 + \int_\Omega \frac{f}{u} = 0.$$

If the constraint set is not empty, the analysis of well-definedness of the iterates is of similar difficulty as in the unconstrained case, but the convergence analysis cannot be carried over easily to additional constraints, in particular the update formula (2.17) must involve additional terms corresponding to Lagrange multipliers of the constraints. Since preliminary numerical experiments demonstrate the success of the iterative regularization procedure also for multiplicative denoising, such an analysis seems to be an important task for future research.

2.3.5 Related Work

In interesting earlier work, [TNV04], the authors propose an iterative procedure also based on the ROF model. They also generate a sequence u_k which converges to the given image f. It is interesting to compare the two approaches.

To recall: the approach described in Algorithm 2.3 is to compute u_k as a minimizer of the convex functional

$$Q_k(u) = \lambda \int_\Omega (u - f)^2 + J(u) - J(u_{k-1}) - \langle p_{k-1}, u - u_{k-1} \rangle$$

for $k = 1, 2, \cdots$, with $u_0 = 0, p_0 = 0$ and compute $p_k \in \{p_{k-1} - 2\lambda(u_k - f)\} \cap \partial J(u_k)$.

The Tadmor-Nezzar-Vese (TNV) approach is (in our language): set $u_0 = 0$ and compute u_k as a minimizer of the convex functional

$$\tilde{Q}_k(u) = \lambda 2^{k-1} \int_\Omega (u - f)^2 + J(u - u_{k-1})$$

for $k = 1, 2, \cdots$.

For $J(u)$ homogeneous of degree one, as in the ROF model, this can be rewritten as: minimize

$$\tilde{\tilde{Q}}_k(u) = \lambda \int_\Omega (u - f)^2 + J\left(\frac{u - u_{k-1}}{2^{k-1}}\right).$$

Thus we see the differences: (1) The TNV algorithm uses a hierarchical decomposition where the difference in total variation between u and the previous iterate is computed. (2) A dyadic sequence of scales, $\lambda 2^{k-1}$, is used to obtain convergence.

The differences in performance can also be seen. If we define $f = \alpha x_R$ for $\alpha\lambda R \geq 1$ as in (2.4), the new algorithm recovers $u_k \equiv f$ for all $k \geq 2$. The TNV algorithm finds

$$u_k = \left(\alpha - \frac{1}{\lambda 2^{k-1} R}\right) \chi_R, \qquad k = 1, 2, \cdots$$

Also, the new algorithm has a denoising aspect to it. Theorem 2.3.5 indicates that the iterative refinement sequence u_k has the property that Bregman distance between u_k and $\tilde{u}$, the true restored solution, decreases until the discrepancy principle is satisfied. There is no such result in [TNV04]. Finally, we mention that a similar approach as in [TNV04], but without proofs, can also found in the earlier paper [SG01].

2.4 Numerical Examples

In this section some numerical results using the iterative regularization procedure will be presented. These examples concentrate on total variation denoising (and thus use Algorithm 2.1. Some of the results were previously presented in [OBG05].

Some notation and formula used here are: g is a clean signal (one-dimensional, 1D) or image (two-dimensional, 2D) and it is supposed to be unknown, the given noisy data is $f = g + n$ (noisy) or $f = Ag + n$ (blurry and noisy), where $n \sim \mathcal{N}(0, \sigma^2)$ is Gaussian noise, $\sigma = \|n\|_{L^2}$ is the variance of noise and it can also be considered as the noise level. The grid size is $h = 1$. The decomposition in the restored results is $f = u + v$ (denoising) or $f = Au + v$ (deblurring and denoising), where u is the restored image and v is considered as a residual. Note that the knowledge of the noise level is useful in the experiments only as a stopping criterion via the discrepancy principle. The results are otherwise independent of the noise level. Variational methods applied to image processing often use noise estimates as constraints.

All solutions to the variational problem (1.2) were obtained using gradient descent in a standard fashion, see e.g. [ROF92]. This amounts to solving a parabolic equation to steady state with normal derivative zero imposed at the boundaries. The only nontrivial difficulty comes when $|\nabla u| \approx 0$. This is fixed, as is usual, by perturbing $J(u) = \int_\Omega |\nabla u|$ to $J_\epsilon(u) = \int_\Omega \sqrt{|\nabla u|^2 + \epsilon^2}$, where ϵ is a small positive number, e.g. [DS96]. Without confusion, the subindex ϵ will be dropt from J_ϵ in the left of this section. In the examples below a very small value $\epsilon = 10^{-6}$ is used[1]. The initial guess for first step was the noisy data f. For succeeding iterates in the Bregman procedure (2.2) and (2.3), f will be replaced by $f + v_{k-1}$ and the ROF procedure (1.2) will be proceeded again, with the initial guess replaced by $f + v_{k-1}$ or the previous iterate u_{k-1}.

The stopping rule for determining the "optimal" $u_{\bar{k}}$ among the iterates $\{u_k\}$ is described in Section 2.3.3, and can be simply restated as follows:

$$\bar{k} = \min_k \{\|u_k - f\|_{L^2} \leq \sigma\} \quad \text{for denoising;} \tag{2.33}$$

$$\bar{k} = \min_k \{\|Au_k - f\|_{L^2} \leq \sigma\} \quad \text{for deblurring and denoising.} \tag{2.34}$$

The first comparison is taken between the results of the iterative procedure and the constrained denoising used in the original ROF paper [ROF92].

Example 1 (1D signal): In the first example a one-dimensional noisy signal is considered. Figure 2.2a displays the original signal g. Figure 2.2b displays the noisy signal $f = g + n$, where the variance of noise is $\sigma\|n\|_{L^2} = 9.45$. Figure 2.2c shows the restored u obtained using ROF with the constraint $\|u - f\|_{L^2} = \sigma$

[1]Note that in [OBG05] we set $J_\epsilon(u) = \int_\Omega \sqrt{|\nabla u|^2 + \epsilon}$ so the ϵ there is 10^{-12}.

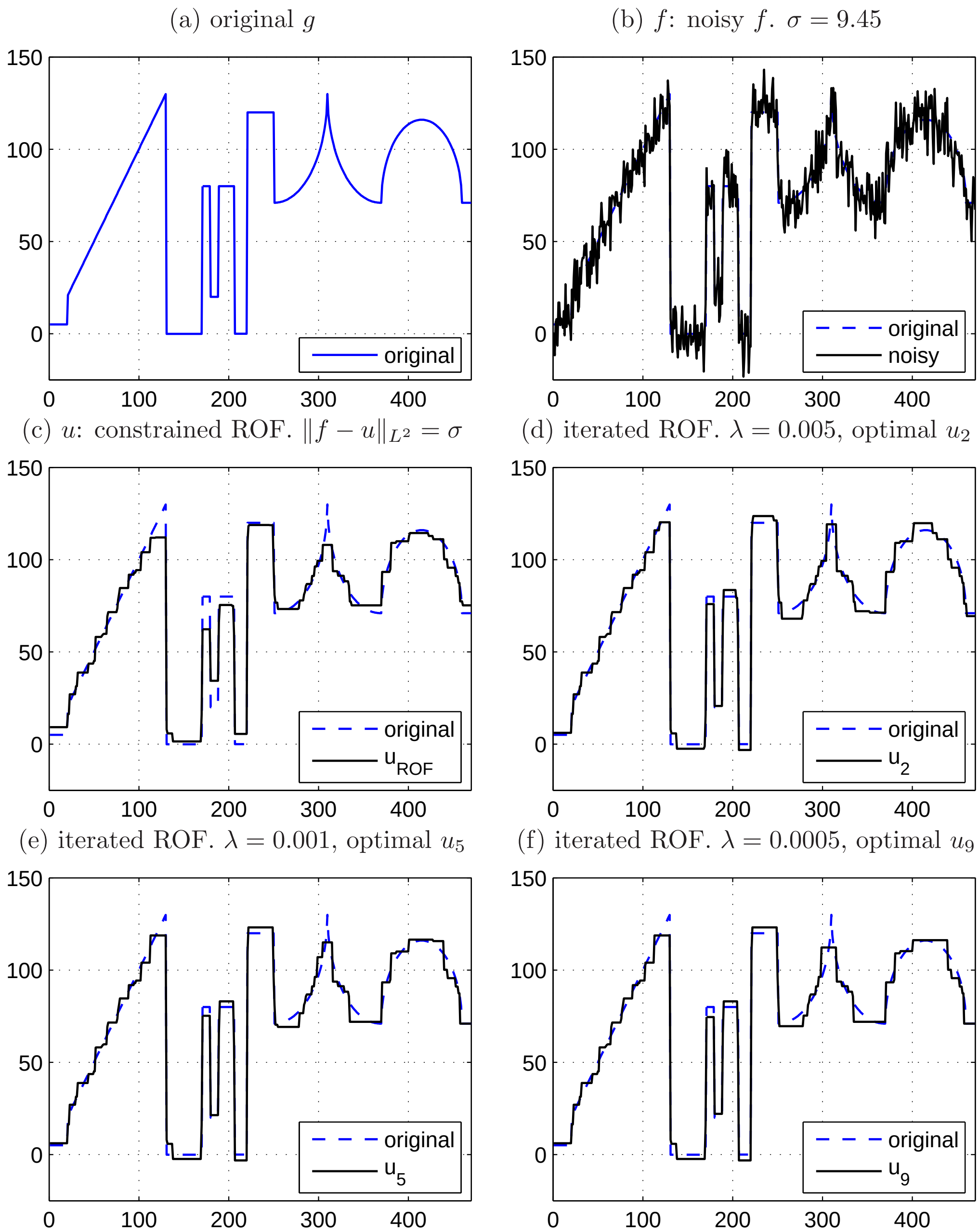


Figure 2.2: (1D signal) Denoising comparison: constrained ROF vs. ROF with iterative regularization.

[ROF92]. One could see the typical loss of accuracy in regions where there are narrow peaks and valleys. In Figures 2.2d, 2.2e and 2.2f, the optimal results $u_{\bar{k}}$ of the iterative regularization procedure with $\lambda = 0.005, 0.001$, and 0.0005 are presented respectively. All three results are more accurate than the result obtained with the one-step ROF minimization subject to a constraint on the L^2 norm of the removed noise, especially near local minima and maxima of f. The results also confirm numerically the assertion that using a smaller λ (i.e., initially over-smoothing) requires more regularization steps to get the optimal restoration. Therefore $\bar{k} = \bar{k}(\lambda, \sigma)$ does not only depend on the noise level σ, but also on the choice of λ.

Example 2 (1D signal): Another one-dimensional signal is considered. The first experiment is due to the "restoration" of a clean signal. From Meyer's analysis (1.5) and (1.6), no matter how large λ is, the residual $v = f - u$ of the original ROF procedure is either f or satisfies $\|v\|_* = \frac{1}{2\lambda}$. In another word, it will not be zero and thus we can not completely recover $u = f$ for clean signal or image by one-step ROF procedure. In Figure 2.3a a result u_1 with $\lambda = 0.005$ is shown, which corresponds to the one-step ROF restoration. Figure 2.3b shows a second step result from the iterative regularization procedure. The improvement is obvious. Next, for the noisy signal ($SNR = 13.5$) in Figure 2.3c containing Gaussian noise whose $\sigma = 10$, the iterative regularization results u_1, u_2, u_3 are displayed in Figures 2.3d, 2.3e, 2.3f respectively. $\lambda = 0.001$. From u_1 ($SNR = 10.4$) to u_2 ($SNR = 22.9$), the results improved considerably. Succeeding u_k become noisy again for $k \geq 3$ ($SNR = 22.7$). One could conclude that $\bar{k} = 2$ for this example.

Example 3 (Fingerprint): From now on some two-dimensional images will be tested. The first one is a noisy fingerprint image f as shown in Figure 2.4b,

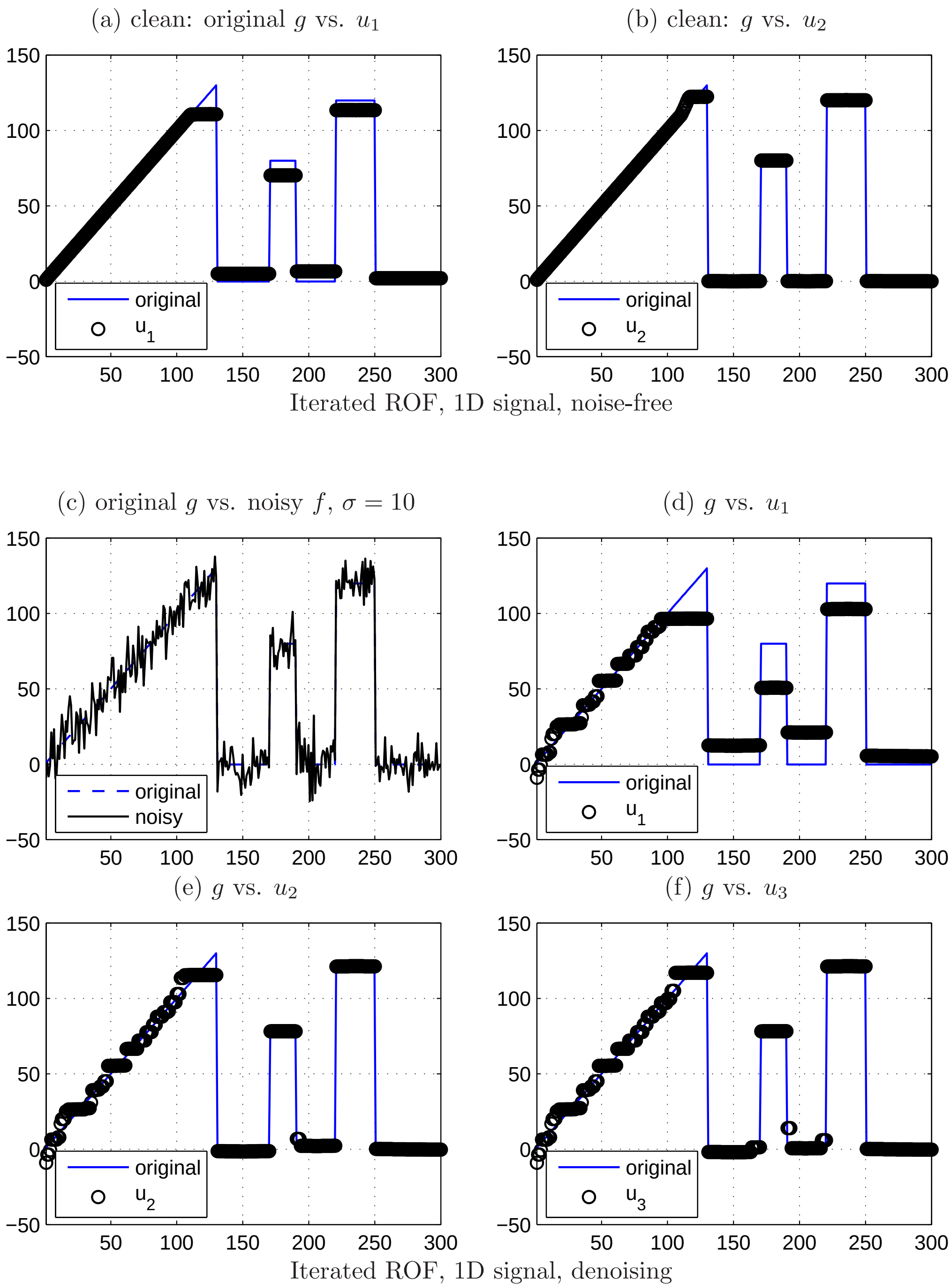


Figure 2.3: (1D signal) ROF with iterative regularization on 1D signal (noise-free & noisy) .

which is corrupted by the original clean image g in Figure 2.4a with a Gaussian noise n shown in Figure 2.4c, $\sigma = \|n\|_{L^2} = 10$.

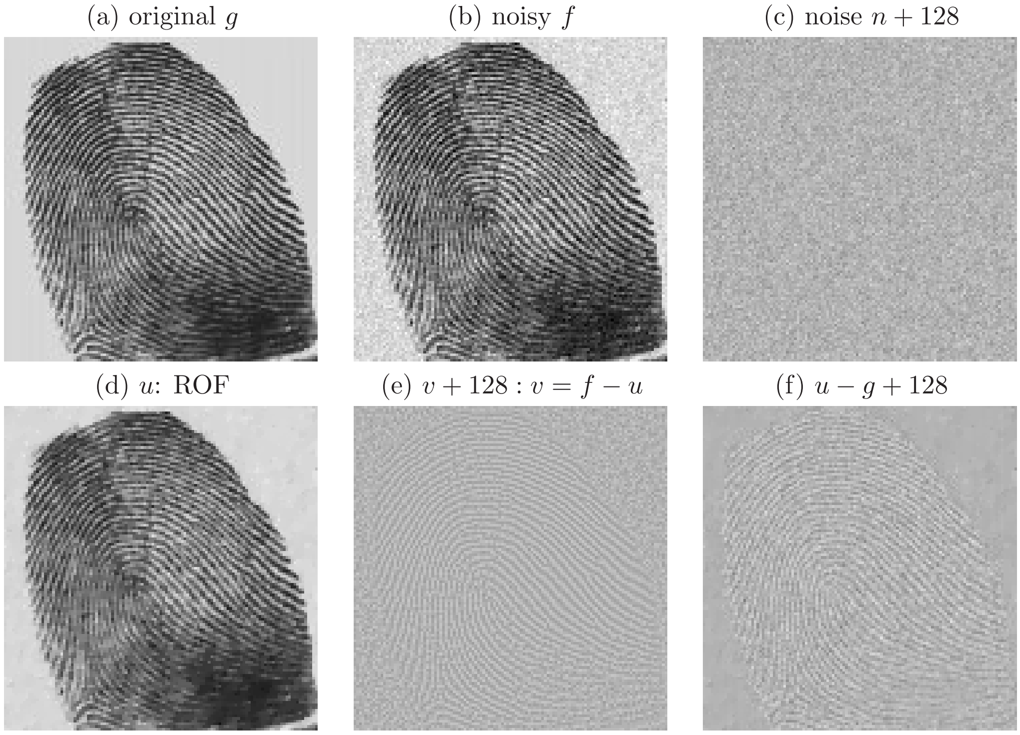


Figure 2.4: (Fingerprint) ROF restoration. Gaussian noise, $\sigma = \|f-g\|_{L^2} = 10.0$. $\lambda = 0.085$, $\|f-u\|_{L^2} = 10.2$, visible signal can be found in the residual $v = f-u$.

First, the ROF model (1.2) is applied with $\lambda = 0.085$. This produced a restored image u with $\|f - u\|_{L^2} = 10.2 \approx \sigma$. Some visible signal can be found in the removed noise component $v = f - u$. This is a common problem for the ROF model. The "signal loss" can also be observed by checking the error $u - g$. Figure 2.4d-e show the results u, $f - u + 128$ and $u - g + 128$. The shift by adding 128 is used to make the small values of n, $f - u$ and $u - g$ more visible. This technique will also be applied to the visualization of all noise n and residual v or v_k shown in this dissertation if not specifically specified. Next, ROF with iterative regularization is applied to denoise f. $\lambda = 0.013$. Notice that this time

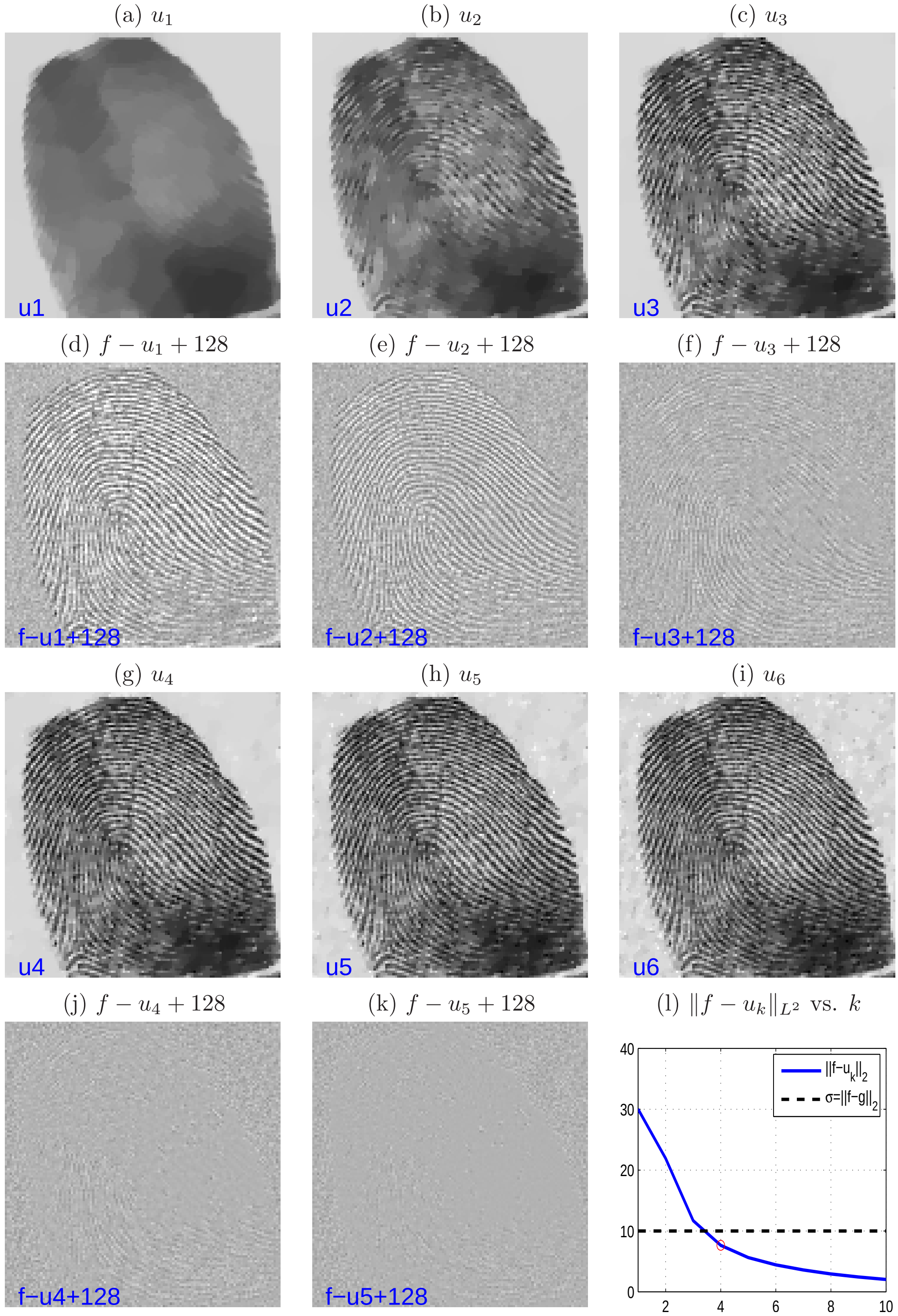


Figure 2.5: (Fingerprint) ROF with iterative regularization. Noisy image data from Figure 2.4b. $\lambda = 0.013$. Optimal restoration obtained when $\|f - u_k\|_{L^2}$ drops below σ at $\bar{k} = 4$. Noise returns in succeeding $u_5, u_6, \ldots$.

the value of λ is much smaller than the one used for ROF. As indicated before this small λ will over-smooth u_1. But u_k improves steadily as k increases, with u_4 in Figure 2.5g giving the best restoration. Figure 2.5j shows the residual $f - u_4$, which contains only very little visible signal. Figure 2.5l is a plot of $\|f - u_k\|_{L^2}$ as a function of the iterate k. It shows that $\|f - u_k\|_{L^2}$ decreases monotonically with k, first dropping below σ at the optimal iterate $\bar{k} = 4$, hence validating Theorem 2.3.5.

This example also shows visually why the iterative regularization procedure can be considered as a multiscale restoration: the largest scale (carton and shape) is restored first in the first step u_1, and the finer scales are restored in the succeeding iterations, with bigger k u_k restores finer scales.

(a) original g (b) noisy f (c) noise $n + 128$

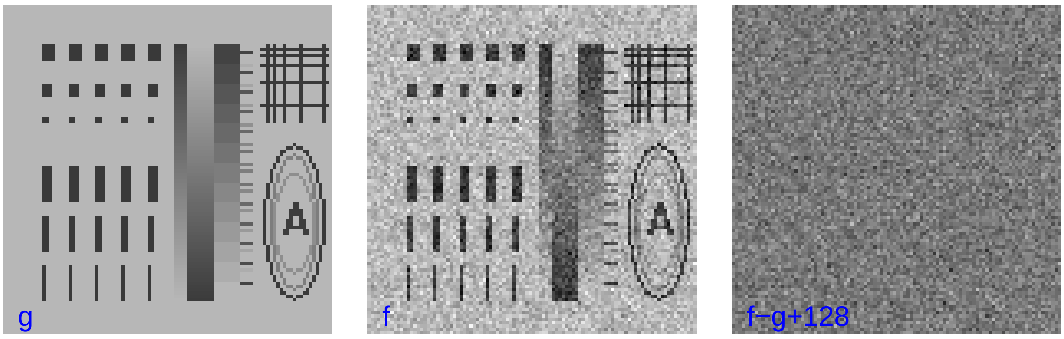
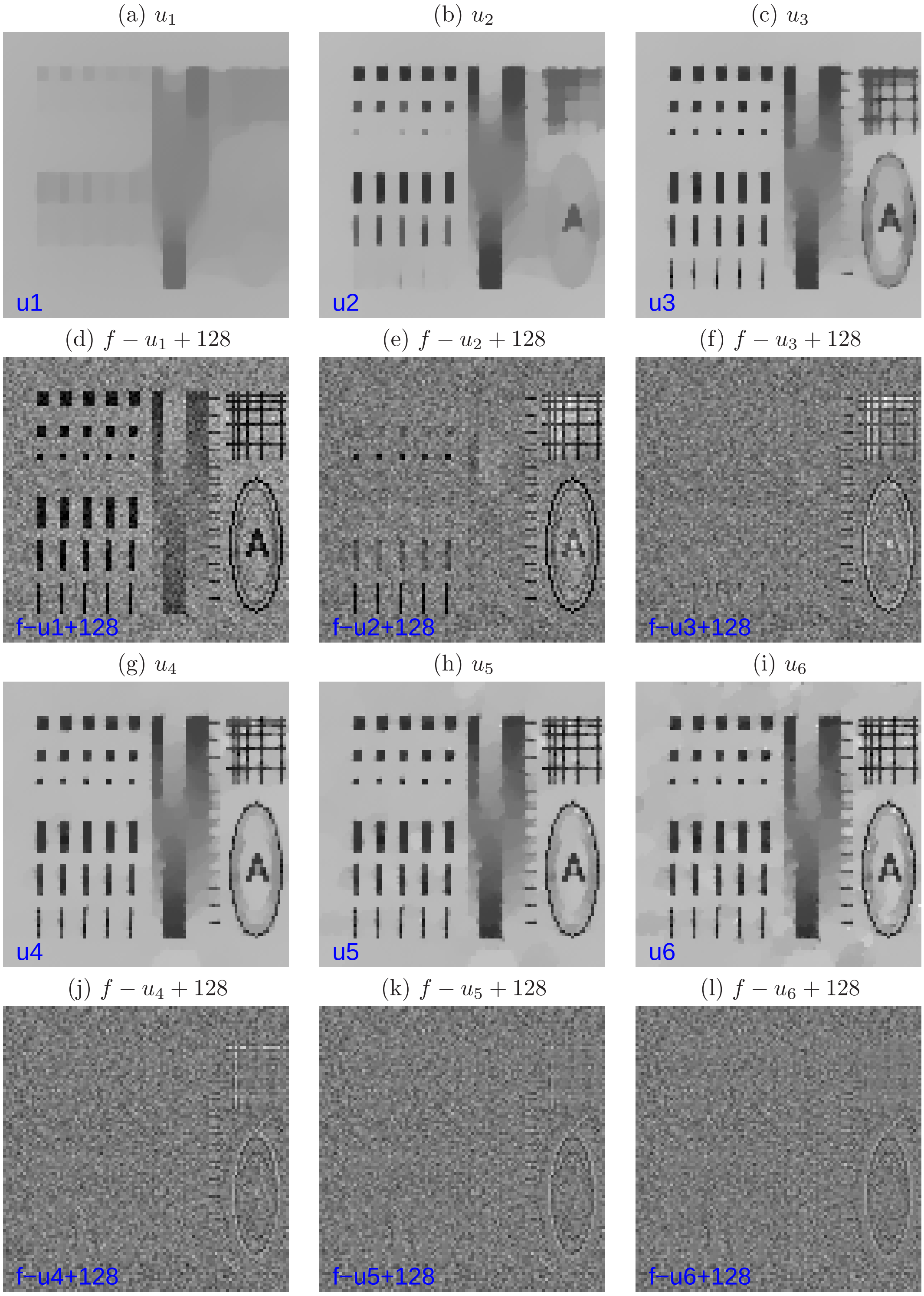


Figure 2.6: (Synthetic shape image) From left to right: original image g; noisy image f; Gaussian noise $n = f - g$, $n + 128$, $\sigma = 40$.

Example 4 (Synthetic shape image): In this example, a synthetic image containing various shapes and patterns is used to test the iterative regularization procedure. Figure 2.6 shows the original image g, noisy image f and the Gaussian noise $n = f - g$ (in the plot $n + 128$ is used, as was explained before). $\sigma = 40$, $SNR(f) = 6.5$. Figure 2.7 shows a sequence of results u_k and $f - u_k + 128$, which comes from the iterative regularized ROF procedure with $\lambda = 0.002$. As can be



(a) u_1 (b) u_2 (c) u_3

(d) $f - u_1 + 128$ (e) $f - u_2 + 128$ (f) $f - u_3 + 128$

(g) u_4 (h) u_5 (i) u_6

(j) $f - u_4 + 128$ (k) $f - u_5 + 128$ (l) $f - u_6 + 128$

Figure 2.7: (Synthetic shape image) ROF with iterative regularization. Data from Figure 2.6, $\sigma = 40$, $\lambda = 0.002$. Optimal restoration obtained when $\|f - u_k\|_{L^2}$ drops below σ at $\bar{k} = 5$. $SNR(u_4) = 11.7$, $SNR(u_5) = 12.5$, $SNR(u_6) = 12.2$.

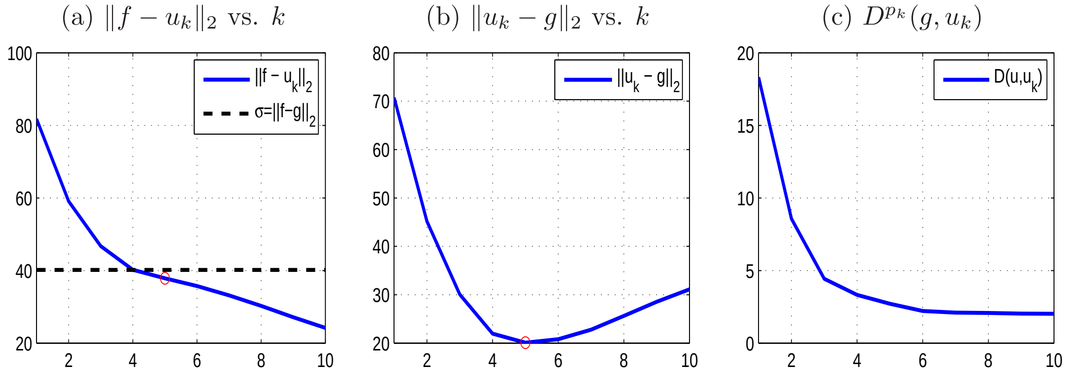


Figure 2.8: (Synthetic shape image) From left to right: $\|f - u_k\|_2$ vs. k; $\|u_k - g\|_2$ vs. k; Bregman distance $D^{p_k}(g, u_k)$. Data u_k from Figure 2.7.

seen from the figure the restored results u_k improve until the 5th step, at which point the inequality $\|f - u_5\|_{L^2} < \sigma$ first becomes satisfied. Then the succeeding $u_k(k \geq 6)$ become noisier, again validating Theorem 2.3.5. Figure 2.8 shows some quantities related with these results: $\|f - u_k\|_{L^2}$ vs. k, the error $\|u_k - g\|_{L^2}$ vs. k and the Bregman distance $D^{p_k}(g, u_k)$. One could see that: $\|f - u_k\|_{L^2}$ is monotonically decreasing w.r.t. k; $\|g - u_k\|_{L^2}$ also attains its minimum at the optimal point $\bar{k}$ (This is not generally true for all of the numerical examples here. Often the smallest true L^2 error occurs for one or two more more regularization iterations); $D^{p_k}(u, u_k)$ is decreasing for $k \leq \bar{k}$, as predicted by Theorem 2.3.5. It is interesting that this quantity sometimes continues to decrease well after noise has returned to the iterate u_k, see Figure 2.8c.

Next, to illustrate the relationship between λ and the optimal step $\bar{k}(\lambda, \sigma)$, another two different λ values are used to denoise the same noisy image in Figure 2.6b. Figure 2.9 displays the results for $\lambda = 0.004$. The restoration u_3 is the best, $\bar{k}(0.004) = 3$. Figure 2.10 presents the results for $\lambda = 0.006$ and shows that u_2 is the best, $\bar{k}(0.006) = 2$. Moreover, as previously stated $\bar{k}(0.002) = 5$. This verifies that $\bar{k}(\lambda)$ monotonically decreases as λ increases, or equivalently, as the

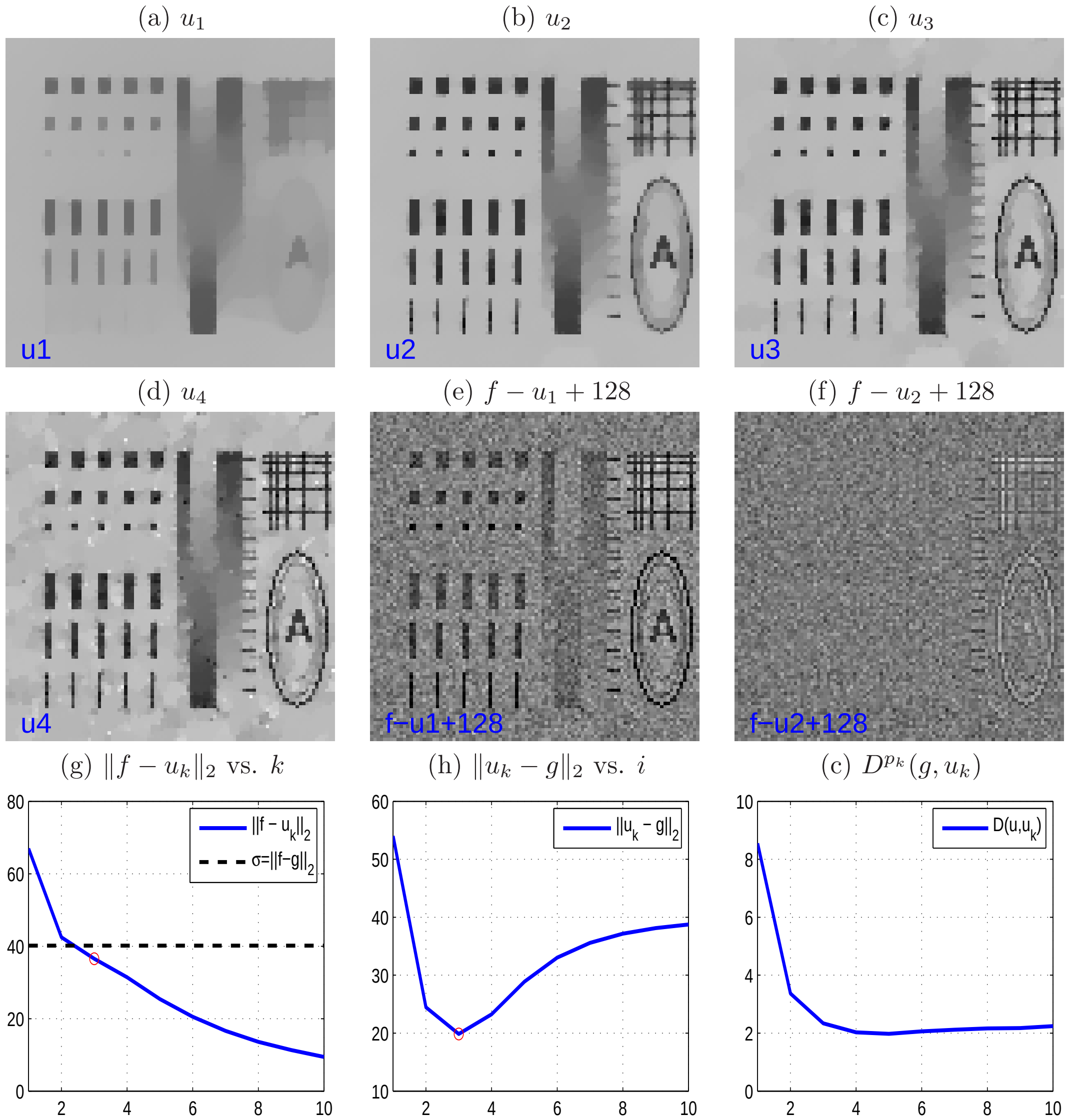


Figure 2.9: (Synthetic shape image) ROF with iterative regularization. Data from Figure 2.6. $\lambda = 0.004$. Optimal restoration obtained when $\|f - u_k\|_{L^2}$ drops below σ at $\bar{k} = 3$. $SNR(u_1) = 3.9$, $SNR(u_2) = 10.8$, $SNR(u_3) = 12.6$, $SNR(u_4) = 11.2$.

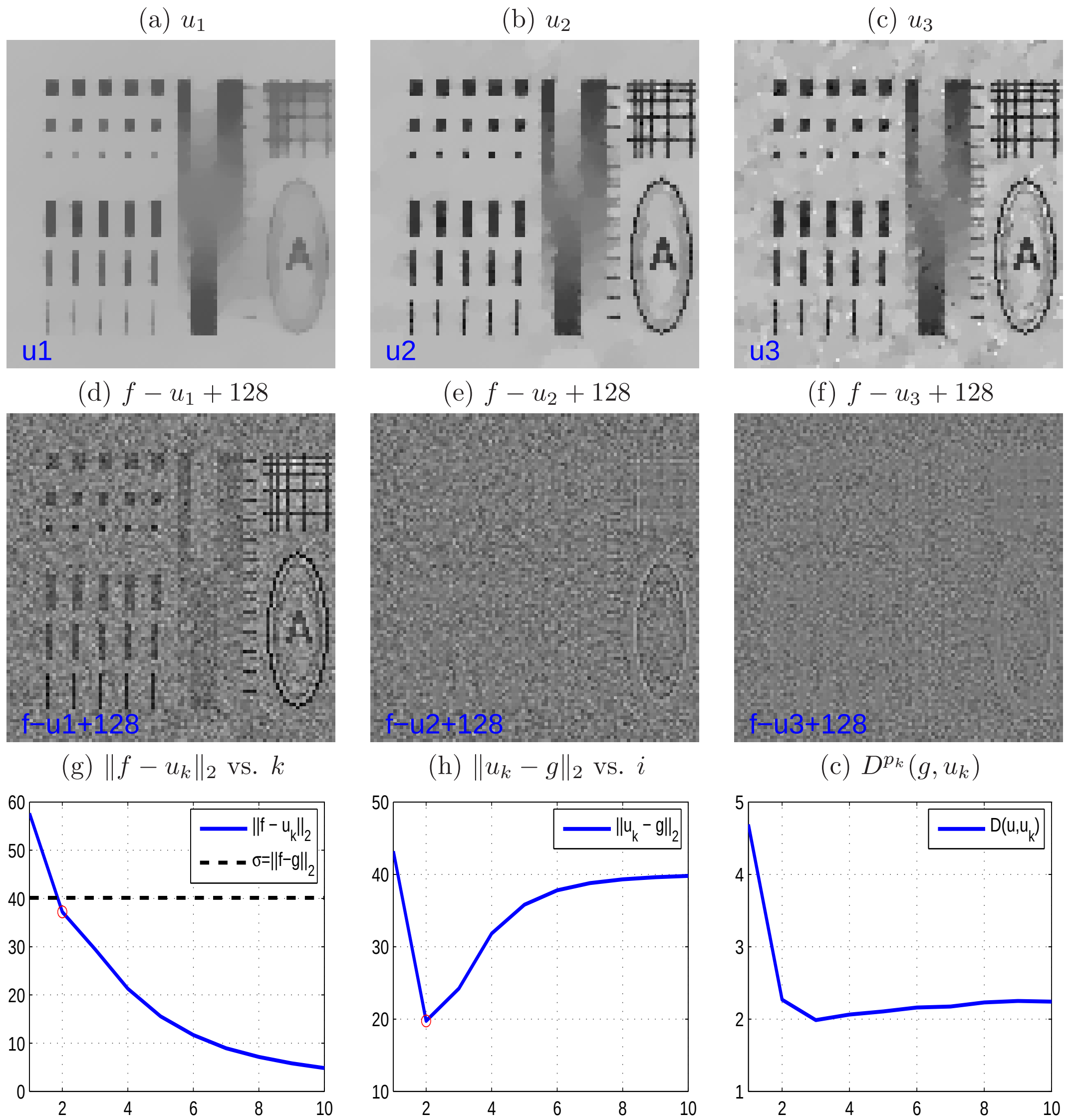


Figure 2.10: (Synthetic shape image) ROF with iterative regularization. Data from Figure 2.6. $\lambda = 0.006$. Optimal restoration obtained when $\|f - u_k\|_{L^2}$ drops below σ at $\bar{k} = 2$. $SNR(u_1) = 5.8$, $SNR(u_2) = 12.6$, $SNR(u_3) = 10.9$.

amount of initial over-smoothing decreases.

Example 5 (Textured image): In this example denoising a textured image is considered. The noisy data is the one same as in Lysaker-Osher-Tai [LOT04], $\sigma = 17.2$, $SNR = 9.4$. As can be seen in Figure 2.11, for $\lambda = 0.02$, the results improved considerably from u_1 to u_2, and got noisy again from u_3. u_2 gave the best restoration in this example.

Example 6 (Satellite): This example denoises a satellite image. The data is shown in the first row of Figure 2.12. $\sigma = 40$, $SNR = 5.1$. For $\lambda = 0.0055$, u_2 yields the best restoration, with almost all signal restored and very little visible signal in the residual. In u_3 and succeeding $u_k, k > 3$, some noise comes back. This is displayed in Figure 2.12.

Example 7 (Satellite deblurring): As mentioned before, the iterative procedure can be generalized to many cases such as deblurring. In [HMO05] He, Marquina and Osher developed the Bregman based iterative blind deconvolution. Thanks to their generous help one deblurring+denosing example is presented in here: a clean satellite image g is blurred by a Gaussian kernel $(A * g)$ and added with Gaussian noise, $f = A * g + n$, $\sigma = 10$, and then proceeded with iterative regularized blind deconvolution. Figure 2.13 shows the data and results. $\lambda = 0.1$. With iterative regularization, u_2 recovers more signal than u_1, especially small details. The restored image u_2 has the least noise, but u_3 appears to be sharper. Succeeding iterations ($k \geq 4$) become noisier. For more details about the iterative regularized blind deconvolution, please refer to [HMO05].

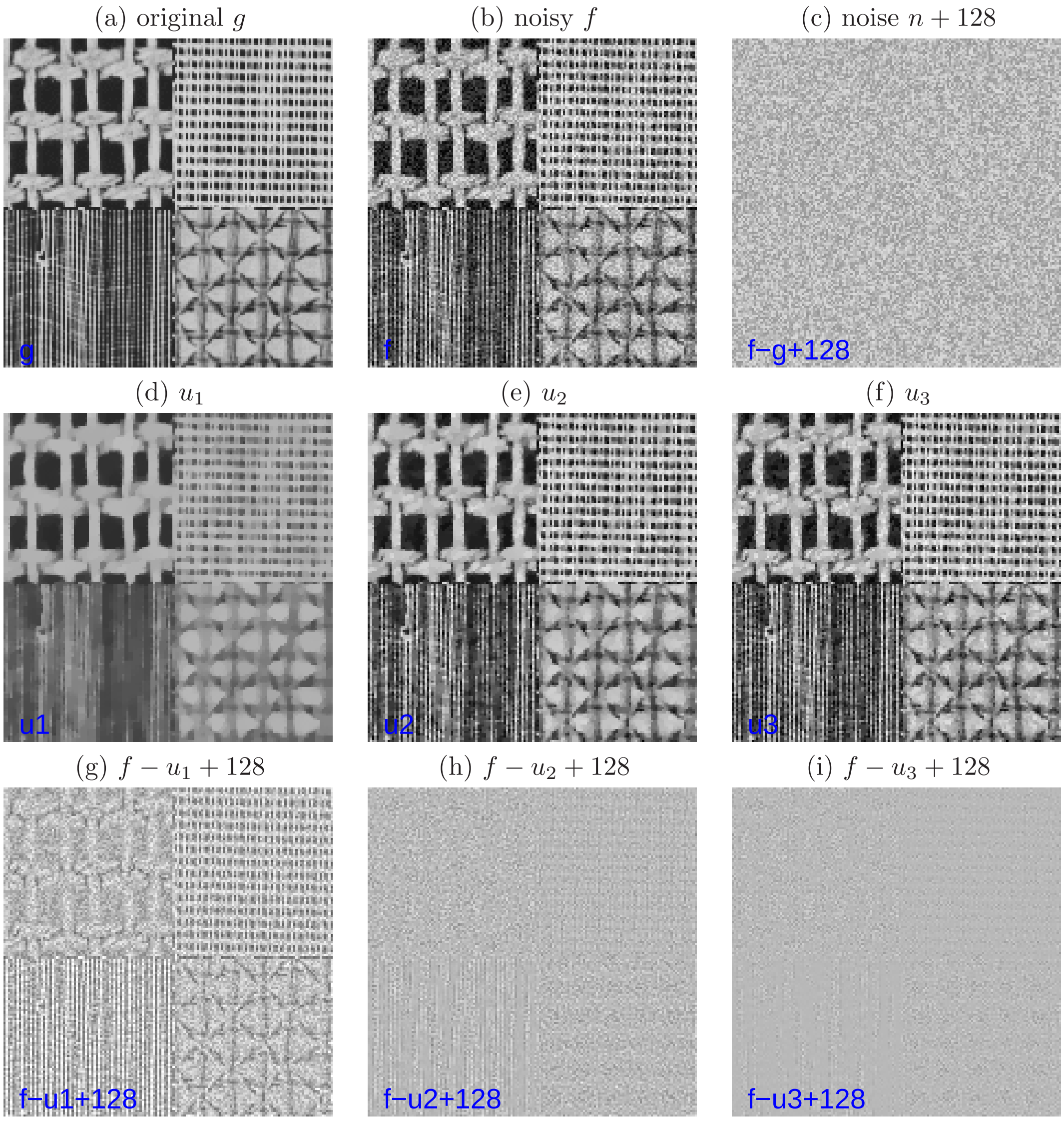


Figure 2.11: (Textured image) ROF with iterative regularization. Data from [LOT04]. $\lambda = 0.02$. Optimal restoration obtained at u_2. $SNR(u_1) = 5.9$, $SNR(u_2) = 10.7$, $SNR(u_3) = 10.4$.

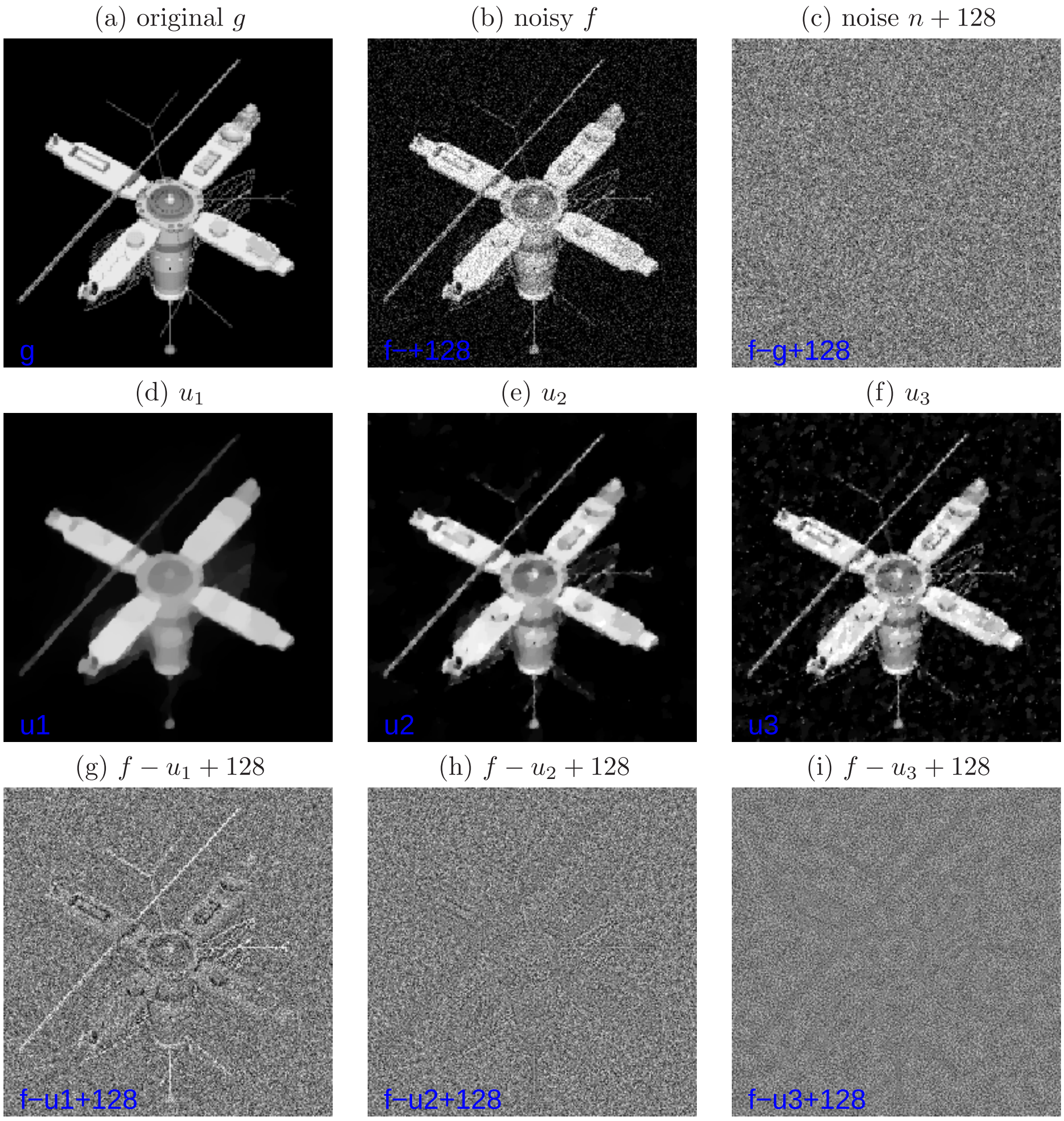


Figure 2.12: (Satellite image) ROF with iterative regularization. $\lambda = 0.0055$. Optimal restoration obtained at u_2. $SNR(u_1) = 12.6$, $SNR(u_2) = 15.1$, $SNR(u_3) = 10.5$.

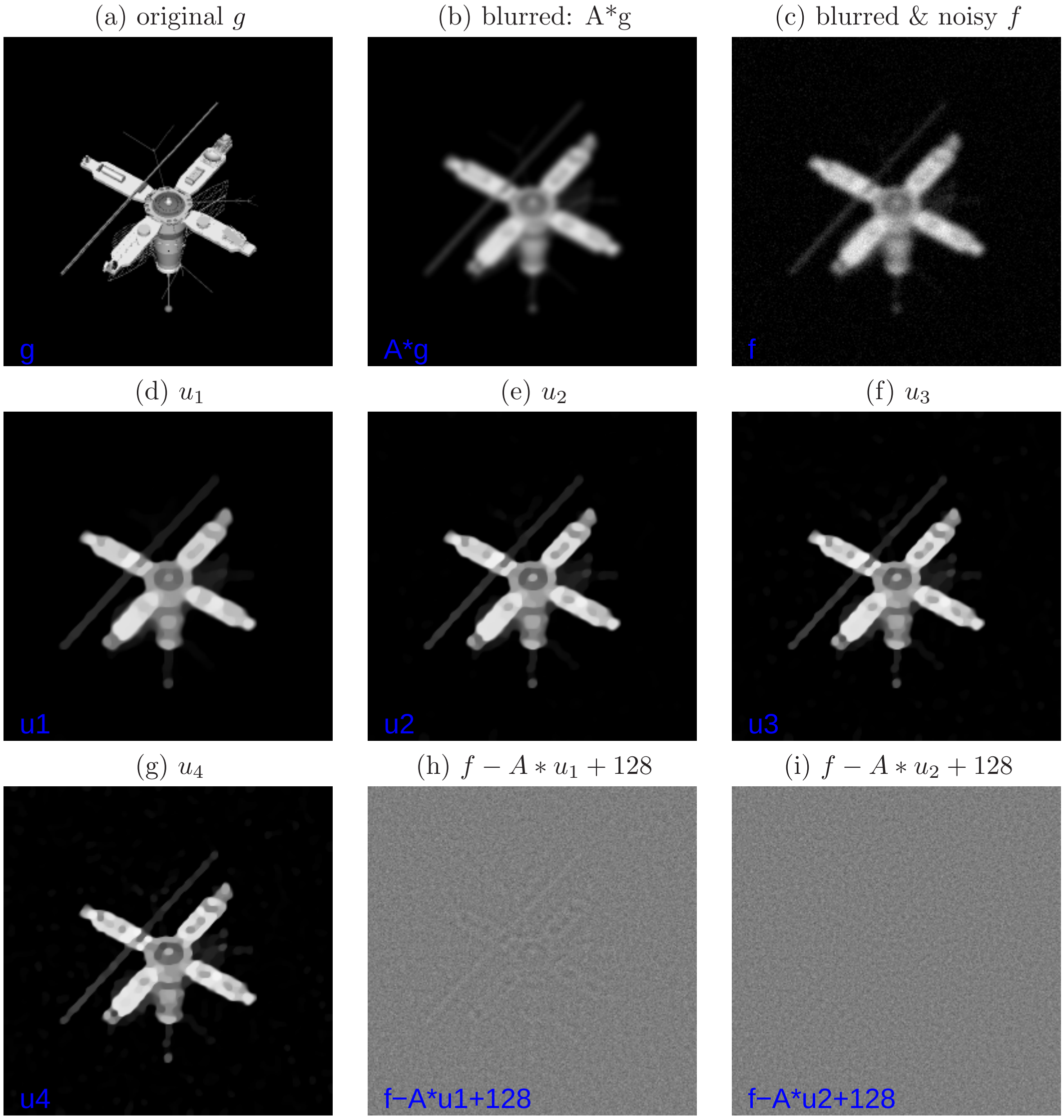


Figure 2.13: (Satellite image, deblurring+denoising) ROF with iterative regularization. $\lambda = 0.1$.

2.5 Discussion and Conclusion

In this chapter we introduced a new iterative regularization procedure for inverse problems based on the use of generalized Bregman distances, with particular focus on problems arising in total variation based image restoration. We obtained rigorous convergence analysis and effective stopping criterion. The numerical results for denoising appear to give significant improvement over standard models. The iterative regularization procedure can be generalized to many types of variational models and inverse problems.

One important parameter of the iterative regularization procedure is the scale parameter λ. This was briefly discussed in Section 2.2.1. As we have seen in the numerical experiments, smaller λ corresponds to more iterations to reach the stopping point and the optimal results seem to be better than that of using larger λ and fewer iterations. But its corresponding computation cost is more expensive. In next chapter we will reinterpret λ as a pseudo time-step which tends to zero and generalize the iterations to a time-continuous inverse scale space flow. In most cases, for example, the relaxed inverse scale space of ROF model has about the same computational complexity as using standard gradient descent methods to solve the original ROF (1.2). We will detail these in Chapter 3.

Acknowledgments

Useful and stimulating discussions are acknowledged to Antonin Chambolle (Ecole Polytechnique, Paris), Michael Hintermüller, and Wolfgang Ring (University Graz). The original data of 1D signal in Example 1 and synthetic shape image in Example 4 are acknowledged to Wotao Yin (Columbia University). The noisy data of textured image in Example 5 is acknowledged to Marius Lysaker

(Simula Research Laboratory AS, Norway) and Xue-Cheng Tai (University of Bergen). Numerical experiment of satellite image blind deconvolution in Example 7 is acknowledged to Lin He (UCLA) and Antonio Marquina (Universidad de Valencia).

CHAPTER 3

Nonlinear Inverse Scale Space Methods

3.1 Introduction

In this chapter we generalize the iterative regularization method (IRM) described in previous chapter to a time-continuous inverse scale space (ISS) formulation.

Since the noise in images is usually expected to be a small scale feature, particular attention has been paid to methods separating scales, in particular those smoothing small scale features faster than large scale ones, so-called *scale space methods*.

Scale space methods are obtained for example by nonlinear diffusion filters [PM90] of the form

$$\frac{\partial u}{\partial t} = \mathrm{div}(\gamma(|\nabla u|^2)\nabla u), \tag{3.1}$$

in $\Omega \times \mathbb{R}_+$ with $u(x,0) = f(x)$, where $f : \Omega \to \mathbb{R}$ denotes the given image intensity (Ω being a bounded open subset in $\mathbb{R}^2$) and $u : \Omega \times \mathbb{R}_+ \to \mathbb{R}$ the flow of smoothed images. The diffusion coefficient involves a positive and monotone function γ. For such methods it can be shown that small scales are smoothed faster than large ones, so if the method is stopped at a suitable final time, we may expect that noise is smoothed while large-scale features are preserved to some extent. For some examples of linear and nonlinear scale-spaces see [Wit83, PM90, ALM92, Wei99, KMS00, GSZ04] and the references therein. Diffusion filters can be related to regularization theory (cf. [SW00]) with certain regularization functionals, but theoretical foundations of choosing optimal stopping times are still missing (see [MN03, GSZ05] for two recent studies concerning the stopping time problem).

Inverse scale space methods have been introduced in [SG01], which are based on a different paradigm. Instead of starting with the noisy image and gradually smoothing it, inverse scale space methods start with the image $u(x, 0) = 0$ and approach the noisy image f (which will be normalized to have mean zero) as time increases, with large scales converging 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. The inverse scale space method can also be related to regularization theory, in particular iterated Tikhonov regularization (cf. [GS00, SG01]) with the same regularization functionals as for diffusion filters. The construction of inverse scale space methods in [SG01] worked well for quadratic regularization functionals, which led to an interesting, but linear evolution equation, but did not yield convincing results for other important functionals, in particular for the total variation. In this chapter we present a different version of constructing inverse scale space (ISS) methods as the limit of an iterative regularization procedure described in previous chapter and demonstrate its applicability to image restoration. With the new approach we are able to easily implement nonlinear inverse scale space methods for the total variation functional, and, in contrast to diffusion filters, a rigorously justified and simple stopping criterion will be obtained for the methods.

3.2 Inverse Scale Space Methods

In the following we generalize the concept of inverse scale space theory introduced in [GS00, SG01] in the context of Tikhonov regularization for the case

$$J(u) = \frac{1}{2} \int_\Omega |\nabla u|^2. \tag{3.2}$$

We shall derive general inverse scale space methods as a limit of the iterative regularization procedure for $\lambda \to 0$, with particular emphasis on the functional

$$J(u) = \int_\Omega \sqrt{|\nabla u|^2 + \epsilon^2}, \tag{3.3}$$

$\epsilon > 0$ is a small constant.

Instead of (1.2), we replace the scale parameter λ by $\frac{\lambda}{2}$ and rewrite the ROF minimization as

$$u = \underset{u \in BV(\Omega)}{\arg\min} \left\{ |u|_{BV} + \frac{\lambda}{2} \|f - u\|_{L^2}^2 \right\}. \tag{3.4}$$

Note that λ is a notation for a constant parameter. This change will facilitate the discussion in this chapter and will be used from now on.

Now we recall from previous chapter that for a special $\lambda > 0$ the iterative refinement procedure for ROF constructs sequences u_k of primal and p_k of dual variables such that $u_0 = p_0 = 0$,

$$\begin{aligned} u_k &= \underset{u \in BV(\Omega)}{\arg\min} \left\{ D^{p_{k-1}}(u, u_{k-1}) + \frac{\lambda}{2} \|f - u\|_{L^2}^2 \right\}, \\ p_k &\in \partial J(u_k). \end{aligned}$$

From the Euler-Lagrange equation

$$p_k - p_{k-1} + \lambda(u_k - f) = 0$$

we are led to the dual iteration

$$\frac{p_k - p_{k-1}}{\lambda} = f - u_k, \quad k = 1, 2, \cdots$$

for the updates. We now reinterpret $\lambda = \Delta t$ as a time step and the difference quotient on the left-hand side as an approximation of a time derivative. Setting $t_k = k\Delta t$, $p(t_k) = p_k$, and $u(t_k) = u_k$, we have $p_{k-1} = p(t_{k-1}) = p(t_k - \Delta t)$ and thus

$$\frac{p(t_k) - p(t_k - \Delta t)}{\Delta t} = f - u(t_k).$$

For $\Delta t \downarrow 0$ (dropping the subindex k) we arrive at the differential equation

$$\frac{\partial p}{\partial t}(t) = f - u(t), \qquad p(t) \in \partial J(u(t)), \tag{3.5}$$

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

In order to obtain well-posedness also if J is only the total variation seminorm (or any other functional vanishing on constant functions), we shall always assume in the following that the image f is scaled such that $\int_\Omega f = 0$. At this point we mention that all the inverse scale space methods and arguments discussed below can be generalized in a straightforward way for $\int_\Omega f \neq 0$, with the only difference that the initial value has to be chosen as the constant $u = \frac{1}{|\Omega|}\int_\Omega f$.

If the flow $u(t)$ according to (3.5) exists and is well behaved, it is an inverse scale space method in the sense of [GS00]. This means that the flow starts at $u(0) = 0$ and incorporates finer and finer scales (with the concept of scale depending on the functional J) finally converging again to the image f as $t \to \infty$, i.e. $\lim_{t\to\infty} u(t) = f$. Through (3.5) the image $u(t)$ flows from the smoothest possible image ($u(0) = 0$) to the noisy image f. Our goal is to use the flow to denoise the image, and therefore we shall use a finite stopping time for the flow. As we shall see below, we can use a simple stopping criterion related to the fitting term $\|u(t) - f\|_{L^2}$ only.

The inverse scale space (ISS) approach should give more accurate results than the iterative regularization method (IRM) because we can compute the stopping time more accurately in the ISS approach due to the continuous evolution. This is borne out to some degree by our results in Section 3.6, but the differences are small. (Of course both methods are significantly better than solving the standard variational problem (3.4), even with the best choice of the parameter λ). We mention that with careful choice of parameter, the IRM corresponds to an implicit Euler discretization of the ISS so that similarities are not surprising. However, under such conditions the IRM requires many solutions of (3.4), with a modified f, which creates quite a high computational effort. The ISS approach gives a chance to create a much faster algorithm, which is true in particular after a relaxation we shall introduce below. The complexity can be reduced to that of using a simple forward Euler time integration of two evolution equations, and their structure is such that there is no severe time step restriction for stability. Of course, using the relaxation we only solve a (reliable) approximation to ISS instead the true evolution equation, which slightly reduces the accuracy but in a controllable way (via a relaxation parameter). From these reasons and from the detailed elaborations below it seems clear that the ISS can yield some superior properties compared to IRM, and due to its scale space interpretation it is more appealing for a wide community in image processing.

3.2.1 Behaviour for Quadratic Regularization

We start by briefly reviewing the results obtained in [GS00] for the quadratic regularization (3.2). In this case we obtain from the variation of the functional J the boundary value problem

$$\begin{aligned} -\Delta u &= p \qquad \text{in } \Omega, \\ \frac{\partial u}{\partial n} &= 0 \qquad \text{on } \partial\Omega, \\ \int_\Omega u &= 0 = \int_\Omega f. \end{aligned}$$

Given p with $\int_\Omega p = 0$, there exists a unique solution u.

A simple manipulation (and the fact that $\frac{\partial f}{\partial t} = 0$) leads us to the equation

$$\frac{\partial}{\partial t}(u - f) = \Delta^{-1}(u - f) = -A(u - f),$$

with the notation $A := -(\Delta)^{-1}$. Thus, the function $w = u - f$ satisfies an integro-differential equation (the integral kernel corresponds to the Green's function of $-\Delta$), whose solution is given by

$$u(t) - f = w(t) = e^{-tA} w(0) = -e^{-tA} f.$$

It is well-known that A is a positive definite operator and thus, $e^{-tA} f$ decays to zero. As a consequence, the difference $u(t) - f = -e^{-tA} f$ decays exponentially as $t \to \infty$. Also, as a consequence of the results of the following section, for any function g for which $\|\nabla g\|_{L^2} < \infty$ and $\int_\Omega g = 0$, then the error $\|\nabla(u(t) - g)\|_{L^2}$ decreases as long as $\|u(t) - f\|_{L^2} > \|g - f\|_{L^2}$. This indicates that the inverse scale space procedure is a good alternative to the classical Wiener filter or diffusion filtering via the heat equation

$$\partial_t u(t) = \Delta u(t), \qquad u(0) = f.$$

3.2.2 General Convex Regularization

We consider the case of general convex functionals $J : \mathcal{U} \to \mathbb{R}$ on a Banach space $\mathcal{U}$ (the digital image in $\mathbb{R}^d$ is then interpreted as the discretization on a grid). If J is continuously differentiable, we can compute the implicitly defined primal variable $u = u(p)$ as the solution of $J'(u(p)) = p$. Note that if J is smooth and strictly convex, the Hessian $H = J''$ is positive definite, and hence, the existence of a solution is guaranteed under a standard condition like $J(0) = 0$ by the inverse function theorem.

A possibility to invert the equation for u is the use of the dual functional (or *convex conjugate*, cf. [ET99]), defined by

$$J^*(p) := \sup_u \Big\{ \langle u, p \rangle - J(u) \Big\}. \tag{3.6}$$

Then one can easily show that $p \in \partial_u J(u)$ is equivalent to $u \in \partial_p J^*(p)$ and we obtain an explicit relation for $u(p)$ provided we can compute the dual functional J^*.

Under the above conditions, we can obtain some important estimates for the inverse scale space flow (3.5) associated to J. We start by computing the time-derivative of the fitting functional and the (partial) time derivative of u:

$$\frac{1}{2}\frac{d}{dt}\|u(t) - f\|_{L^2}^2 = \langle u(t) - f, \partial_t u(t) \rangle$$

$$\partial_t u(t) = \frac{d}{dt}(\partial_p J^*(p(t))) = H^*(p(t))\partial_t p(t) = -H^*(p(t))(u(t) - f),$$

where we used the notation $H^* = \partial^2_{pp} J^*$ for the Hessian of the dual functional. If J^* is strictly convex, then there exists a constant $a > 0$ such that

$$\langle \varphi, H^*(q)\varphi \rangle \geq a \|\varphi\|^2_{L^2}$$

for all $\varphi, q \in \mathcal{U}^*$. Hence, combining the above estimates we deduce

$$\frac{1}{2}\frac{d}{dt}\|u - f\|^2_{L^2} = -\langle u(t) - f, H^*(p(t))(u(t) - f)\rangle \leq -a\|u - f\|^2_{L^2}$$

and using Gronwall's inequality we have

$$\|u(t) - f\|_{L^2} \leq e^{-a(t-s)}\|u(s) - f\|_{L^2} \leq e^{-at}\|f\|_{L^2}$$

if $t > s$. Thus, as $t \to \infty$ we obtain convergence $u(t) \to f$ with exponential decay of the error in the L^2-norm.

Note that for the above L^2-estimates, we do not need severe assumptions on f, so that the estimate holds for a clean image as well as for a noisy version used in the algorithm. If we assume that f is a clean image and $J(f) < \infty$, then we can also obtain a decay estimate on the error in the Bregman distance via

$$\begin{aligned} \frac{d}{dt} D^p(f, u(t)) &= \frac{d}{dt}\big[J(f) - J(u(t)) - \langle f - u(t), p(t)\rangle\big] \\ &= -\langle f - u(t), \partial_t p(t)\rangle = -\|u(t) - f\|^2_{L^2}. \end{aligned}$$

We can also have the following convergence of $p(t)$ to $q \in \partial J(f)$ if we assume the stronger condition $q \in L^2$ (a so-called *source condition*, cf. [BO04]). From

(3.5) we proceed formally to

$$\begin{aligned}\frac{1}{2}\frac{d}{dt}\|p(t)-q\|_{L^2}^2 &= \langle \partial_t p(t), p(t)-q\rangle = \langle f-u, p(t)-q\rangle \\ &= -D^p(f,u) - D^q(u,f).\end{aligned}$$

So for strictly convex smooth J and for $J(f)<\infty$ we have that the subgradient of $u(t)$ monotonically goes to the subgradient of f in L^2. Also, there are subsequences in t going to infinity for which both $D^p(f,u(t))$ and $D^q(u(t),f)$ converge to zero. Moreover, we can integrate the last inequality from time zero to t, which gives

$$\frac{1}{2}\left(\|p(t)-q\|_{L^2}^2 - \|p(0)-q\|_{L^2}^2\right) + \int_0^t \left[\, D^p(f,u(s)) + D^q(u(s),f)\,\right]\, ds = 0.$$

Since $p(0)=0$, $\|p(t)-q\|_{L^2}^2$ and $D^q(u(s),f)$ are nonnegative, and $D^{p(u(s))}(f,u(s)) \geq D^{p(u(t))}(f,u(t))$ for $s\leq t$, we obtain

$$D^p(f,u(t)) \leq \frac{\|q\|_{L^2}^2}{2t}.$$

All results so far give information about the convergence of u to the clean image f (with a finite value $J(f)$) only. In a more practical situation, f is the noisy version of an image g to be restored, and we might even have $J(f)=\infty$, while $J(g)<\infty$. In this case we can state the following proposition:

Proposition 3.2.1. *For the above conditions, the Bregman distance $D^p(g,u(t))$ is decreasing with time at least as long as $\|f-u(t)\|_{L^2} > \sigma$, where $\|f-g\|_{L^2}\leq\sigma$.*

Proof. As in the case of the clean image we directly compute

$$\begin{aligned}\frac{d}{dt}D^p(g,u) &= \langle -\partial_t p(t), g-u(t)\rangle = -\langle f-u(t), g-u(t)\rangle \\ &= -\|f-u(t)\|_{L^2}^2 - \langle f-u(t), g-f\rangle \\ &\le -\frac{\|f-u(t)\|_{L^2}^2}{2} + \frac{\|f-g\|_{L^2}^2}{2}.\end{aligned}$$

The last term on the right-hand side is negative if $\|f-u(t)\|_{L^2} > \|f-g\|_{L^2}$. $\square$

This means that $u(t)$ approaches any "noise free" image g in the sense of Bregman distance, as long as the residual (the L^2 difference between $u(t)$ and f) is larger than the difference between the noisy image f and g. The left-hand side, namely the residual $\|f-u(t)\|_{L^2}$ can be monitored during the iteration, it only involves the known noisy image f and the computed restoration $u(t)$. The right-hand side is not known for the "real" image g to be restored, since g itself is unknown. However, in typical imaging situations, an estimate for the noise variance is known, which yields a bound of the form $\|f-g\|_{L^2} \le \sigma$. The above estimate guarantees that the distance $D^p(g,u)$ is decreasing at least as long as $\|f-u(t)\|_{L^2} > \sigma$, and one could terminate the inverse scale space flow for the minimal $\bar{k}$ such that $\|f-u(\bar{k})\|_{L^2} = \sigma$. This stopping criterion is well-known in the theory of iterative regularization of inverse problems as the so-called *discrepancy principle* (cf. [EHN96, Pla96] for a detailed discussion). *This is a key justification for our denoising approach.*

We emphasize this result because the Bregman distance is stronger than L^2 for the regularization we are considering here, which is significant for denoising. For example, if $J(u) = \frac{1}{2}\int_\Omega u^2 = \frac{1}{2}\|u\|_{L^2}^2$ then the inverse scale space equation is

$$\partial_t u = f - u, \quad u(0) = 0,$$

and $D(g,u) = \frac{1}{2}\|g-u\|_{L^2}^2$. Clearly, for any L^2 function g, we have

$$\frac{d}{dt}\|g-u\|_{L^2}^2 \leq \frac{1}{2}\left(\|f-u\|_{L^2}^2 - \|f-g\|_{L^2}^2\right),$$

and $\|g-u\|_{L^2}$ decreases until $\|f-u(t)\|_{L^2} < \|f-g\|_{L^2}$. This does not imply any sort of regularization or denoising! If, on the other hand, $J(u) = \frac{1}{2}\int_\Omega |\nabla u|^2$, then we have

$$\frac{d}{dt}\|\nabla(g-u)\|_{L^2}^2 \leq \frac{1}{2}\left(\|f-u\|_{L^2}^2 - \|f-g\|_{L^2}^2\right),$$

and we do have a regularization effect for $\int_\Omega |\nabla g|^2 < \infty$.

For the total variation functional

$$J(u) = \int_\Omega |\nabla u|,$$

then, formally,

$$D^p(g,u) = \int_\Omega \left(|\nabla g| - \frac{\nabla g \cdot \nabla u}{|\nabla u|}\right) = \left\langle \nabla g, \frac{\nabla g}{|\nabla g|} - \frac{\nabla u}{|\nabla u|}\right\rangle$$

(ignoring the case of $|\nabla u| = 0$) and this diminishes as the normal to the level curves $\{u = c\}$ line up with those of $\{g = c\}$. Although $D^p(g,u)$ can vanish for g not identical to u, it is fairly easy to show that $D^p(g,u) = 0$ implies that $g = R(u)$, R being a non-decreasing function. This means that g and u are the same up to a contrast change. For a discussion of this kind of morphological equivalence, see [AGL93]. The proof can be outlined as follows: $D^p(g,u) = 0$ implies $\nabla g = |\nabla g|\frac{\nabla u}{|\nabla u|}$. When taking the curl of this equation, the resulting linear partial differential equation for u has the general solution $u = F(|\nabla g|/|\nabla u|)$, which means that $|\nabla g| = |\nabla u| r(u)$ for some nonnegative function r. The solution to this eikonal equation is $g = R(u)$, where $R' = r$.

3.2.3 Conservation and Scaling Properties

So far we have mainly used dissipation properties to analyze the convergence behaviour of the inverse scale space approach. Some interesting insights can also be gained by investigating conserved quantities and scaling properties of the flow.

A natural quantity to be conserved in image processing is the mean value of the image. Here we assume that $\int_\Omega f = 0$, and of course a natural regularization functional for denoising should satisfy the invariance

$$J(v) = J(v + c), \qquad \forall\ v \in \mathcal{U}, c \in \mathbb{R}.$$

Then, for $v = u + 1$ the subgradient p satisfies

$$\int_\Omega p = \langle p, 1 \rangle = \langle p, v - u \rangle \leq J(v) - J(u) = J(u+1) - J(u) = 0.$$

Similarly for $v = u - 1$ we have $\int_\Omega p \geq 0$. Consequently

$$0 = \frac{d}{dt} \int_\Omega p = \int_\Omega \partial_t p = \int_\Omega (f - u) = -\int_\Omega u,$$

i.e., u has mean zero.

Another interesting property concerns the scaling of solutions. We consider the case $\tilde{f} = \alpha f$ for some $\alpha \in \mathbb{R}$, and $\tilde{u}$ (with subgradient $\tilde{p}$) denotes the solution of (3.5) with f replaced by $\tilde{f}$. Then, depending on the scaling properties of the regularization functional J, we obtain interesting rescalings of the flow.

Proposition 3.2.2. *Under the above conditions and notations, there is the following connections between u, p and $\tilde{u}$, $\tilde{p}$:*

- *If J is quadratic (such as (3.2)), i.e., $J(\alpha v) = \alpha^2 J(v)$ for all $v \in \mathcal{U}$, then $\tilde{u} = \alpha u$, $\tilde{p} = \alpha p$.*

- *If J is positively one-homogeneous (such as the total variation), i.e., $J(\alpha v) = \alpha J(v)$ for all $v \in \mathcal{U}$, then $\tilde{u}(t) = \alpha u(\alpha^{-1} t)$, $\tilde{p}(t) = p(\alpha^{-1} t)$.*

- *If J is the regularized total variation, i.e. $J(v) = \int_\Omega \sqrt{|\nabla u|^2 + \epsilon^2}$, then $\tilde{u}(t) = \alpha u(\alpha^{-1} t)$, $\tilde{p}(t) = p(\alpha^{-1} t)$ for $\tilde{\epsilon} = \alpha \epsilon$.*

Proof. For all cases we have

$$\partial_t \tilde{p} = \tilde{f} - \tilde{u} = \alpha(f - \alpha^{-1} \tilde{u}), \qquad \tilde{p} \in \partial J(\tilde{u}).$$

In the quadratic case we have $\partial J(\tilde{u}) = \alpha \partial J(\alpha^{-1} \tilde{u})$ and hence

$$\partial_t (\alpha^{-1} \tilde{p}) = f - (\alpha^{-1} \tilde{u}), \qquad \alpha^{-1} \tilde{p} \in \partial J(\alpha^{-1} \tilde{u}).$$

This means $\alpha^{-1} \tilde{u}$ is a solution of (3.5) with subgradient $\alpha^{-1} \tilde{p}$, and by uniqueness $\tilde{u} = \alpha u$, $\tilde{p} = \alpha p$.

If J is a positively homogeneous of degree one functional like the total variation, then $\partial J(\alpha u) = \partial J(u)$. Hence, with the notation as above, we obtain (with time variable $\tilde{t}$)

$$\alpha^{-1} \partial_{\tilde{t}} \tilde{p} = f - (\alpha^{-1} \tilde{u}), \qquad \tilde{p} \in \partial J(\alpha^{-1} \tilde{u}).$$

After an additional time rescaling $t = \alpha \tilde{t}$ we obtain again that $\alpha^{-1} \tilde{u}(t)$ is a solution of (3.5) with subgradient $\tilde{p}(t)$. Again, by uniqueness, we obtain that for a solution u of (3.5) with image f, the rescaling $\alpha u(\alpha^{-1} t)$ is a solution with image $\tilde{f} = \alpha f$.

In the case of the regularized total variation the proof is analogous, noticing that

$$\int_\Omega \sqrt{|\nabla \tilde{u}|^2 + \tilde{\epsilon}^2} = \int_\Omega \alpha \sqrt{|\nabla u|^2 + \epsilon^2}.$$

□

Proposition 3.2.2 yields a certain scaling invariance of the algorithm, in any case the inverse scale space evolution with some image f can be obtained from the evolution with rescaled image. In the case of regularized total variation it confirms the rather obvious fact that the regularization parameter ϵ should be scaled with the image.

3.2.4 Comparison to ROF Scale Space

In contrast to the evolution (3.5) generating $u(t)$ we would like to show why a different obvious inverse scale space, namely the one generated by varying the penalty parameter in (3.4), is a less appealing alternative. Note that the improvement seen with respect to (3.4) seen in numerical experiments is one of the major motivations for investigating (3.5). Let us consider the inverse scale space defined by $w(t)$ which satisfies:

$$w(t) = \arg\min_{u \in BV(\Omega)} \left\{ J(u) + \frac{t}{2} \|u - f\|_{L^2}^2 \right\}, \quad t \geq 0.$$

For the sake of simplicity we assume that J is twice differentiable. Clearly $w(0) = 0$, $w(\infty) = f$ if we have the usual hypotheses on $J(u)$ and $H(u, f)$ and consider the familiar class of examples. The Euler-Lagrange equation is

$$p(t) + t(w(t) - f) = 0, \qquad p(t) = J'(w(t)).$$

Differentiating in time yields

$$\partial_t p(t) + t\partial_t w(t) = f - w(t). \qquad w(0) = 0,\, p(0) = 0. \tag{3.7}$$

We claim that this evolution equation is not as useful as our inverse scale space equation (3.5) (which it resembles). To show this, we first examine the convergence to f:

$$\frac{d}{dt}\frac{1}{2}\|w - f\|_{L^2}^2 = \langle \partial_t w, w - f\rangle = -\langle (J''(w) + t)^{-1}(w - f), w - f\rangle.$$

This means that $\|w - f\|_{L^2}$ decays to zero, but only at a slow algebraic rate $\frac{1}{t}$, not exponentially. A more serious drawback comes from the relation

$$\frac{d}{dt}D^{p(w)}(g, w) = -\langle g - w, f - w\rangle + \langle g - w, t\partial_t w\rangle,$$

and an analogous reasoning as in Proposition 3.2.1 is not apparent due to the second term. In fact, for the quadratic case (3.2) we have

$$\begin{aligned}
\frac{d}{dt}\|\nabla(w - g)\|_{L^2}^2 &= -\langle\, w - g, w - f\,\rangle + t\langle\, w - g, (-\Delta + tI)^{-1}(w - f)\,\rangle \\
&= -\langle\, w - g, -\Delta(-\Delta + tI)^{-1}(w - f)\,\rangle, \\
&= -\langle\, w - f, [I - (I - t^{-1}\Delta)^{-1}](w - f)\,\rangle
\end{aligned}$$

and the discrepancy principle based on the L^2 distances of $f - g$ and $w(t) - g$ fails. For the ROF scale space the natural norms would be the ones generated by $I - (I - t^{-1}\Delta)^{-1}$, which is on the other hand not the right one to control the image noise. In particular for low-frequency components of $w - g$ and large t, the evolution is very slow, and this is the reason why the reconstruction obtained from (3.4) can lose a lot of the image variation if $\lambda < \infty$ (cf. [Mey01]).

3.3 Inverse Scale Space for Signals

In the following we discuss the numerical solution of (3.5) in spatial dimension one. $\Omega = I \subset \mathbb{R}$. We recall here that $p(t) \in \partial J(u(t))$ and $u(t) \in \partial J^*(p(t))$. We consider again the regularized total variation $J(u) = \int_I \sqrt{u_x^2 + \epsilon^2}\, dx$, which yields

$$\partial J(u(t)) = -\left(\frac{u_x(t)}{\sqrt{u_x(t)^2 + \epsilon^2}}\right)_x = p(t). \tag{3.8}$$

Note that since $\partial J(u + c) = \partial J(u)$, the solution of (3.8) is not unique if we take the standard assumption that u satisfies homogeneous Neumann boundary condition. In this case, the solvability condition is $\int_I p(x,t)\, dx = 0$ for all t and the conservation of mean value discussed above provides an additional property implying uniqueness, namely $\int_I u\, dx \equiv \int_I f\, dx = 0$.

For a fixed time t, we have to solve

$$-\left(\frac{u_x}{\sqrt{u_x^2 + \epsilon^2}}\right)_x = p \ \text{ in } I = (a, b), \quad \int_a^b u\, dx = 0, \tag{3.9}$$

If we denote $q := \dfrac{u_x}{\sqrt{u_x^2 + \epsilon^2}}$, then

$$q(x,t) = -\int_a^x p(s,t)\, ds = \int_x^b p(s,t)\, ds \tag{3.10}$$

and hence, $u_x = \epsilon \dfrac{q}{\sqrt{1 - q^2}}$. Therefore,

$$u(x,t) \ = \ \epsilon \int_a^x \frac{q(y,t)}{\sqrt{1 - q^2(y,t)}}\, dy + C \tag{3.11}$$

where C is a constant chosen to normalize $\int_a^b u(x)\,dx = 0$. We mention that the same formula for u can be obtained by duality arguments, since J^* can be explicitly calculated in spatial dimension one.

Thus, the inverse scale space flow can be computed by two simple integrations. If we compute first the value of p by explicit time discretization in $\partial_t p = f - u$, then we can directly integrate to obtain the value of u at the new time step.

3.4 Relaxed Inverse Scale Space Flow

In order to implement the process in any dimension we resort to a new kind of approximation. First, we would like write the general expressions of the discrete Bregman procedure and the direct inverse scale space having any convex smooth fidelity terms $H(u, f)$.

3.4.1 General fidelity term $H(u, f)$

We rewrite the general form of a convex variational problem (2.10) as

$$\min_u \left\{ J(u) + \lambda H(u, f) \right\}, \tag{3.12}$$

where $H(u, f)$ is usually a fidelity to a known image (or signal) f, (in the L^2 case studied before $H(u, f) = \frac{1}{2}\|f - u\|_{L^2}^2$). Note that the only difference between (2.10) and (3.12) is that we extract the constant parameter λ out of $H(u, f)$. This will facilitate our discussion in this chapter. We assume for simplicity that f is normalized beforehand to have a zero mean: $\int_\Omega f = 0$. We also assume $p(0) = 0 \in \partial J(0)$.

The general form of the series of iterative refined variational minimizations is

$$u_k = \underset{u}{\operatorname{argmin}} \Big\{ D^{p_{k-1}}(u, u_{k-1}) + \lambda H(u, f) \Big\} \tag{3.13}$$

where $u_0 = 0$ and $p_0 = 0$. Expanding $D(\cdot, \cdot)$ according to (2.16) and omitting constant parts which are not relevant to the minimization yields

$$u_k = \underset{u}{\operatorname{argmin}} \Big\{ J(u) - \langle u, p(u_{k-1}) \rangle + \lambda H(u, f) \Big\}. \tag{3.14}$$

The Euler-Lagrange equation of (3.14) is

$$p(u_k) - p(u_{k-1}) + \lambda \partial_u H(u_k, f) = 0.$$

We use $\partial_u H$ to denote that the variation is taken with respect to u. Assigning $p(u_k) = p(u(t))$, one can view the iterations in the limit $\lambda \to 0$ as a continuous process

$$\partial_t p = -\partial_u H(u(t), f), \qquad p \in \partial J(u) \tag{3.15}$$

with the initial conditions $u|_{t=0} = 0$, $p|_{t=0} = 0$.

Contrary to standard scale-spaces, where the flow begins with an input image f and simplifies over time towards zero, this flow begins with $u = 0$ and converges to f over time. The flow is such that large scale features appear much faster than fine scale ones. Therefore this was termed *nonlinear inverse scale space*. The process coincides with the linear inverse scale space of [SG01] (for the case $J(u) = \int_\Omega |\nabla u|^2$), which was formulated without explicit relations to Bregman iterations.

3.4.2 Relaxed Inverse scale space Method

The concise formulation of (3.15) is not straightforward to compute, as the relations between p and u are quite complicated in nonlinear cases. Here we present a relaxed version which aims at having a flow with qualitatively similar properties to that of (3.15) by using standard variational formulations, which are simple to compute.

Let us revisit the series of Bregman iterations stated in equation (3.14). Using the update formula for the decomposed"noise" $v_k = \frac{p_k}{\lambda}$ and the first order optimality conditions, we deduce

$$v_k = v_{k-1} - \partial_u H(u_k, f), \qquad k \geq 1, \quad v_0 = 0$$

and hence,

$$v_k = -\sum_{j=1}^{k} \partial_u H(u_j, f), \quad k \geq 1.$$

The iteration can then be rewritten via the sequence of equivalent variational problems

$$u_k = \operatorname*{argmin}_u \left\{ J(u) + \lambda H(u, f) + \lambda \sum_{j=0}^{k-1} \langle u, \partial_u H(u_j, f) \rangle \right\}. \tag{3.16}$$

or, coupled for u_k and v_k

$$\begin{aligned} u_k &= \operatorname*{argmin}_u \left\{ J(u) + \lambda (H(u, f) - \langle u, v_{k-1} \rangle) \right\}, \\ v_k &= v_{k-1} - \partial_u H(u_k, f), \end{aligned} \tag{3.17}$$

where $u_0 = 0, v_0 = 0,\ k = 0, 1, 2, \cdots$.

The standard way to solve these iterations for u_k and v_k (by an explicit scheme) is first to evolve a steepest descent flow for u_k, having a fixed v_k, based on the Euler-Lagrange equations:

$$\frac{\partial u}{\partial t} = -p + \lambda(-\partial_u H(u, f) + v_k), \quad p \in \partial J(u), \quad u|_{t=0} = u_{k-1}, \tag{3.18}$$

where $u_k = u(t \to \infty)$. Note that we assume some regularity in the sequence u_k (such that $\|u_k - u_{k-1}\|_{L^2} \leq const$) and therefore a good starting point for the time marching is u_{k-1}. After converging to a minimizer u_k it is easy to compute $\partial_u H(u_k, f)$ and update v_{k+1}. Then k is incremented by one and the process resumes, such that in each iteration (3.18) is evolved. Although in practice a finite stopping time is used, this process is computationally quite intensive.

Our observation is that the update for v_{k+1} in (3.17) can be viewed as an iterative descent in v_k for minimizing $H(u_k, f)$. This is an indirect minimization, which affects u_k by its coupling with v_k. Let us write the solution for v_{k+1} in the following (more complicated) manner:

$$\partial_\tau v = -\partial_u H(u_k, f), \quad v|_{\tau=0} = v_k, \tag{3.19}$$

where $v_{k+1} = v(\tau = 1)$. This extends the definition of the sequence v_k to a continuous formulation. (Note that for a fixed u_k and a unit stopping time the result is a simple linear interpolation between v_k and v_{k+1}). In the case of Bregman iterations, these flows are evolved iteratively, where in each time either u_k or v_k are being fixed while the dual variable is evolved.

We propose to approximate the sequences u_k, v_k as two continuous flows $u(t), v(t)$ by evolving both descent flows, similar to (3.18) and (3.19), simultaneously. Let us define the relation between the two time variables as $\tau = \alpha t$, and let $v(t) = v(\tau/\alpha)$. Replacing v_k in (3.18) by $v(t)$ and u_k in (3.19) by $u(t)$

yields the *relaxed inverse-scale space (RISS)* flow:

$$
\begin{aligned}
\partial_t u &= -p + \lambda(-\partial_u H(u,f) + v), \\
\partial_t v &= -\alpha \partial_u H(u,f),
\end{aligned}
\tag{3.20}
$$

with $p \in \partial J(u)$ and initial conditions $u|_{t=0} = v|_{t=0} = 0$.

3.4.2.1 Second order in time formulation

If J and H are smooth, the above flow can also be written as a single equation in u with second order derivative in the time domain. This can be done by taking the time derivative of the first equation in (3.20) and substituting for v_t by using the second equation, yielding the following evolution:

$$
\partial^2_{tt} u = -\partial_t p(u) - \lambda(\partial_t(\partial_u H(u,f)) - \alpha \partial_u H(u,f)), \tag{3.21}
$$

with initial conditions $u|_{t=0} = 0, u_t|_{t=0} = -\lambda \partial_u H(0,f)$, which can be written also as

$$
\partial^2_{tt} u = -(J''(u) + \lambda \partial^2_{uu} H(u,f)) \partial_t u - \lambda \alpha \partial_u H(u,f). \tag{3.22}
$$

3.4.2.2 Relation to the direct flow

In order to understand the relation to the original inverse scale space formulation (3.15), we consider the special case $H(u,f) = \frac{1}{2}\|u-f\|^2_{L^2}$, rescale time to $\hat{t} = \frac{t}{\alpha\lambda}$ and define $w(\hat{t}) = \lambda v(\alpha\lambda\hat{t})$. In this way we obtain

$$
\begin{aligned}
\alpha\lambda \partial_{\hat{t}} u &= -p + \lambda(f-u) + w, \\
\partial_{\hat{t}} w &= f - u.
\end{aligned}
\tag{3.23}
$$

If λ is very small (and α not too large) then the leading order term in the first equation is $p = w$, and hence the behaviour is close to the inverse scale space flow on this time scale.

One can observe the strong similarity of the flows numerically in the one-dimensional example presented in Section 3.6 (Figures 3.10 and 3.12).

3.4.3 Examples

Below are some examples of processes that can be evolved using different $J(u)$ and $H(u,f)$:

- *Linear model:* $J = \frac{1}{2}\int_\Omega |\nabla u|^2$, $H(u,f) = \frac{1}{2}\|u-f\|_{L^2}^2$:

$$\begin{aligned} u_t &= \Delta u + \lambda(f - u + v), \\ v_t &= \alpha(f-u). \end{aligned} \tag{3.24}$$

- *ROF model:* $J = \int_\Omega |\nabla u|$, $H(u,f) = \frac{1}{2}\|u-f\|_{L^2}^2$:

$$\begin{aligned} u_t &= \nabla \cdot \frac{\nabla u}{|\nabla u|} + \lambda(f - u + v), \\ v_t &= \alpha(f-u). \end{aligned} \tag{3.25}$$

- *TV-L^1 model (cf. [CE05]):* $J = \int_\Omega |\nabla u|$, $H(u,f) = \|u-f\|_{L^1}$:

$$\begin{aligned} u_t &= \nabla \cdot \frac{\nabla u}{|\nabla u|} + \lambda(\text{sign}(f - u) + v), \\ v_t &= \alpha\text{sign}(f-u). \end{aligned} \tag{3.26}$$

Remark. *Note that the L^1 distance function H is not strictly convex or smooth here and sign is just the notation for an element in the subgradient* *of H. One way to solve this problem numerically is by using approximation $H \approx \sqrt{|u-f|^2+\epsilon^2}$, $0<\epsilon \ll 1$ and then $\partial_u H(u,f) = \frac{u-f}{\sqrt{|u-f|^2+\epsilon^2}}$, which is well-defined and unique for all u.*

- *Deconvolution by ROF:* $J = \int_\Omega |\nabla u|$, $H = \frac{1}{2}\|f - Ku\|_{L^2}^2$, where $K : L^2(\Omega) \to \mathcal{H}$ is a real blurring kernel (see Section 2.3):

$$\begin{aligned} u_t &= \nabla \cdot \frac{\nabla u}{|\nabla u|} + \lambda \left(K^*(f-Ku)+v\right), \\ v_t &= \alpha K^*(f-Ku), \end{aligned} \tag{3.27}$$

where K^* denotes the adjoint of K.

3.5 Properties of the Relaxed Method

3.5.1 Linear Model

The linear case is naturally the easiest to analyze. We can write a closed form solution in the frequency domain and see how the relaxed flow approximates the direct flow.

It is easy to see that the steady state of these equations ($u_t = 0, v_t = 0$) is: $u = f$, $v = \frac{q}{\lambda}$. It remains to analyze the behaviour of the flow for suitable f, and to show that the solutions converge to this steady state, which we will do in the linear case below. A general convergence proof by Lie and Nordbotten [LN05], which can apply for general convex J and L^2 squared fidelity term, is discussed in the next section.

We examine the second order in time formulation (3.22). In the linear case the subgradient is unique and given by $p = -\Delta u$, and $\partial_u H(u,f) = u - f$. The flow can be written as:

$$\partial_{tt}u + (-\Delta + \lambda)\partial_t u + \lambda\alpha u = \lambda\alpha f, \tag{3.28}$$

where $u|_{t=0} = 0, u_t|_{t=0} = \lambda f$.

We rewrite the flow in the frequency domain (with variable ξ), which is obtained by taking the Fourier transform. The characteristic equation is $r^2 + (\lambda + |\xi|^2)r + \alpha\lambda = 0$, with the solutions

$$r_\pm = \frac{-(\lambda + |\xi|^2) \pm \sqrt{(\lambda + |\xi|^2)^2 - 4\alpha\lambda}}{2}. \tag{3.29}$$

Using the Taylor approximation $\sqrt{1+x} \approx 1 + \frac{x}{2}$, $x \ll 1$, one can approximate (for frequencies for which $|\xi|^4 \gg \alpha\lambda$)

$$r_\pm \approx \frac{-(\lambda + |\xi|^2)(1 \pm (1 - \frac{2\alpha\lambda}{(\lambda+|\xi|^2)^2}))}{2}, \tag{3.30}$$

obtaining two roots with different characteristic behavior: $r_+ \approx -(\lambda + |\xi|^2)$, $r_- \approx \frac{-\alpha\lambda}{\lambda+|\xi|^2}$. The Fourier transform of the solution is

$$U(\xi) = (c_+ e^{r_+ t} + c_- e^{r_- t} + 1)F(\xi) \tag{3.31}$$

where $c_+ = \dfrac{\lambda + r_-}{r_+ - r_-}$, $c_- = \dfrac{\lambda + r_+}{r_- - r_+}$.

We observe that the first part, containing r_+, corresponds to a Gaussian convolution, which decays rapidly with time. The second part, containing r_-, corresponds to the inverse scale space solution (with time rescaling by $\lambda\alpha$) which we actually want to solve. Our numerical results indicate that this kind of behavior extends to the nonlinear process.

From (3.29) we see that for both parts to have decaying exponential solutions (real valued $r_\pm$) we should require $\alpha \leq \frac{\lambda}{4}$. In the numerical experiments below we set $\alpha = \frac{\lambda}{4}$.

3.5.2 ROF Model

The ROF model is a natural choice for image regularization since the solution of the flow results in sharp and clean approximations of the input image f without introducing noise (or fine-scale textures in the case of decomposition) up to a very large time.

3.5.2.1 Convergence to Steady-State

In order to analyze the convergence behaviour we define the following energy function

$$e(t) = \frac{1}{2\lambda}\|u - f\|_{L^2}^2 + \frac{1}{2\alpha}\|v - \frac{q}{\lambda}\|_{L^2}^2, \tag{3.32}$$

where $q \in \partial J(f)$ is assumed to be an element of L^2 (this is again a source-condition on the data). Under this assumption, the following convergence property was elegantly proved by Lie and Nordbotten [LN05]:

Proposition 3.5.1. *Let $u(0)$, $v(0)$ be an initial value such that $e(0) < \infty$, and let $u(t)$, $v(t)$ be the solution of (3.25) with $\lambda > 0$, $\alpha > 0$. Then, the energy $e(t)$ decreases monotonically. Moreover, there exists at least a subsequence $t_k \to \infty$ such that*

$$\|f - u(t_k)\|_{L^2} \to 0, \quad D^{p(u)}(f, u(t_k)) \to 0, \quad D^q(u(t_k), f) \to 0. \tag{3.33}$$

Proof. We just compute the time derivative of the energy and insert the evolution law to obtain

$$\begin{aligned}
\frac{de(t)}{dt} &= \frac{1}{\lambda}\langle u - f, u_t\rangle + \frac{1}{\alpha}\langle v - \frac{q}{\lambda}, v_t\rangle \\
&= \frac{1}{\lambda}\langle u - f, -p(u) + \lambda(f - u + v)\rangle + \frac{1}{\alpha}\langle v - \frac{q}{\lambda}, \alpha(f - u)\rangle \\
&= -\|f - u\|_{L^2}^2 - \frac{1}{\lambda}\langle f - u, q - p(u)\rangle \\
&= -\|f - u\|_{L^2}^2 - \frac{1}{\lambda}(D^{p(u)}(f, u) + D^q(u, f)) \\
&\leq 0,
\end{aligned}$$

which implies the monotone decrease. Moreover, by integrating the last inequality with respect to time from 0 to t we have

$$\int_0^t \left(\|f - u(s)\|_{L^2}^2 + \frac{1}{\lambda}(D^{p(u)}(f, u(s)) + D^q(u(s), f)) \right) ds \leq e(0).$$

From the uniform bound for the integral we deduce the existence of a subsequence $t_k \to \infty$ such that

$$\|f - u(t_k)\|_{L^2}^2 + \frac{1}{\lambda}(D^{p(u)}(f, u(t_k)) + D^q(u(t_k), f)) \to 0,$$

and since the latter is the sum of three positive sequences, each of them converges to zero. $\square$

Note that we slightly changed the original proof from [LN05] by using the sum of two Bregman distances (just prior to the final inequality), and it is clear that the result holds for any convex functional J. By this proposition it is clear that $\{u = f,\ v = \frac{q}{\lambda}\}$ is the only steady-state solution of (3.25).

3.5.2.2 Initial conditions

We also note that the flow can be extended to arbitrary initial conditions, and we expect a similar behaviour for large time. Therefore, we would like to illustrate two examples of the flow when $u(0)$ is non-zero. In Figure 3.1 we present instances of the flow from three very different initial conditions $u(0)$, with $v(0) = 0$. In the top row we use the standard initial condition of the inverse scale space flow $u(0) = 0$ (the mean value of the original image is subtracted beforehand and added back after for visualization). The two other initial conditions are white Gaussian noise (with zero mean) in the second row and $u = f$ in the bottom row. All three flows converge to the input f. It appears that after some time (see third column $t = 80$) the evolution is fairly similar regardless of the initial condition. In Figure 3.2 the L^2 distance to the steady states of u and v and the joint energy e (Eq. (3.32)) are plotted as a function of time for all three cases of the above initial conditions. Note that $e(t)$ is monotone in all three cases. Naturally, for the case $u(0) = f$ we have that $\|u - f\|_{L^2}^2$ is not monotone.

Due to the multiscale interpretation of the flow we will however use zero initial value in most computational examples, and in particular in the case of noise it is much more reasonable to start with zero than with the noisy image.

3.5.2.3 The parameters α and λ

We have seen above that α corresponds to a time rescaling only, and both the relation to the original model and the convergence proof hold for any positive α. This allows more freedom in selecting the parameters but raises the issue of what values of α are preferred. The linear analysis shows that we have complex modes for $\alpha > \frac{\lambda}{4}$ which causes oscillations in the convergence. A similar phenomenon occurs for the TV-L^2 case in the analysis of the disk evolution (see

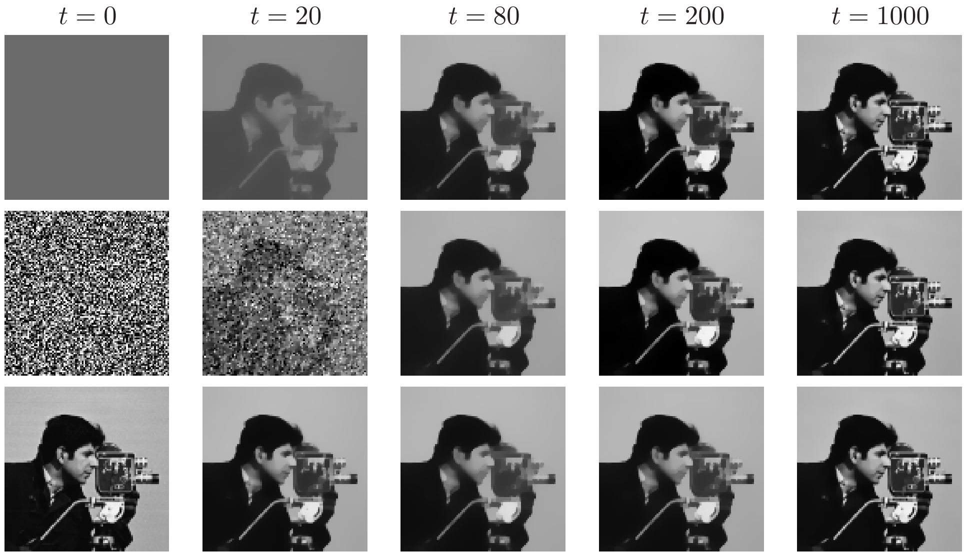


Figure 3.1: Evolution of u towards a clean image f with different initial conditions. Top $u(0) = 0$, second row $u(0) = n$ (white Gaussian noise), bottom: $u(0) = f$. [$v(0) = 0$, $\lambda = 0.02$, $\alpha = \lambda/4$].

Section 3.5.2.5), where we have the same bound on α for monotone convergence. We show in Figure 3.3 numerically that the value of α has a similar effect also when a much more complicated image is evolved, such as the Cameraman image. In Figure 3.4 a somewhat extreme example is shown where $\alpha = 16\lambda$ (that is, 64 times larger than the upper bound). Though eventually the flow converges (as seen in the corresponding plot in Figure 3.3), it is highly non-monotone. It is worth mentioning that even in this regime of α the oscillations are in the contrast of the entire image and the details within the image do not become oscillatory.

Note that the convergence to steady state is proved to be monotone only for u and v *jointly*. From our experiments, it appears that the distance $\|u - f\|^2_{L^2}$ is decreasing monotonically in most cases for zero initial conditions and $\alpha \leq \frac{\lambda}{4}$. We have been able to produce rare synthetic cases where $\|u - f\|^2_{L^2}$ is not monotone. This happens for very large features when λ is very large. However, from the

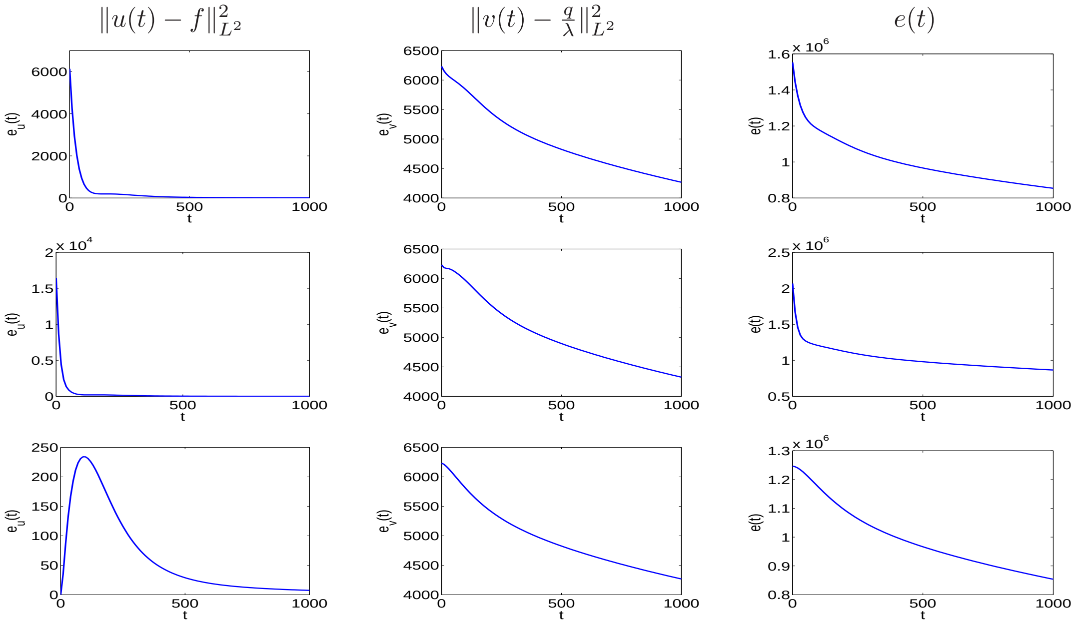


Figure 3.2: Distance from steady state as a function of time: u (left), v (middle) and the joint energy e (right) for different initial conditions. Top $u(0) = 0$, second row $u(0) = n$ (white Gaussian noise), bottom: $u(0) = f$.

relation to the inverse scale space flow above, a smaller choice of λ seems more reasonable anyway.

The parameter λ has a similar role as in the standard variational minimization in the sense that its value should be lower for noisier images or when larger features are considered textures in decomposition. In Figures 3.5 and 3.6 we show the denoising flow (u and the residual $f - u$, respectively) with various values of λ. When λ is too high (e.g. top row) small features, and consequently noise, get in too early. Very low values of λ, such as in the bottom row ($\lambda = 0.005$) produce very good results, though medium values can suffice for a good balance between performance and short evolution time.

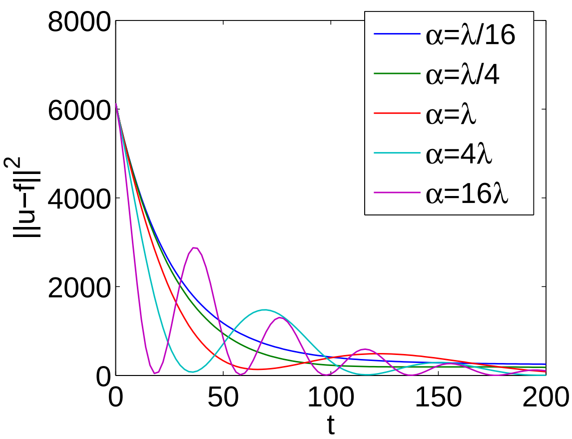


Figure 3.3: $\|u - f\|_2^2$ as a function of time for different values of α: $\alpha \in \{\frac{1}{16}, \frac{1}{4}, 1, 4, 16\}\lambda$. [Cameraman image, $u(0) = v(0) = 0$, $\lambda = 0.02$].

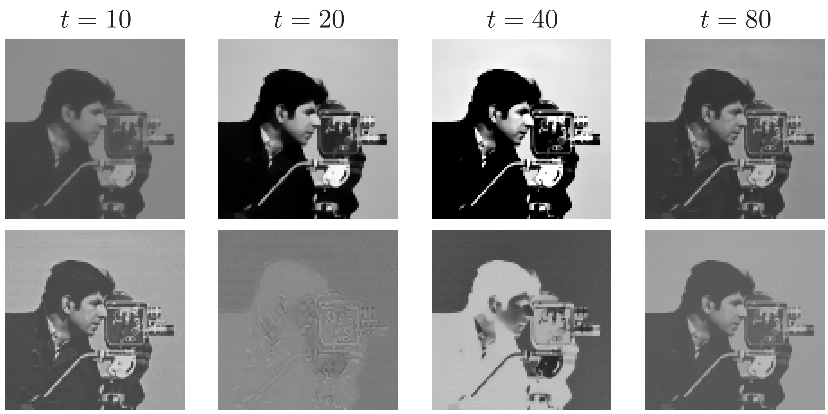


Figure 3.4: Example of an evolution with large α. $u(t)$ (top) and $f - u(t)$ (bottom) at times: $10, 20, 40, 80$, for $\alpha = 16\lambda$. [$u(0) = v(0) = 0$, $\lambda = 0.02$].

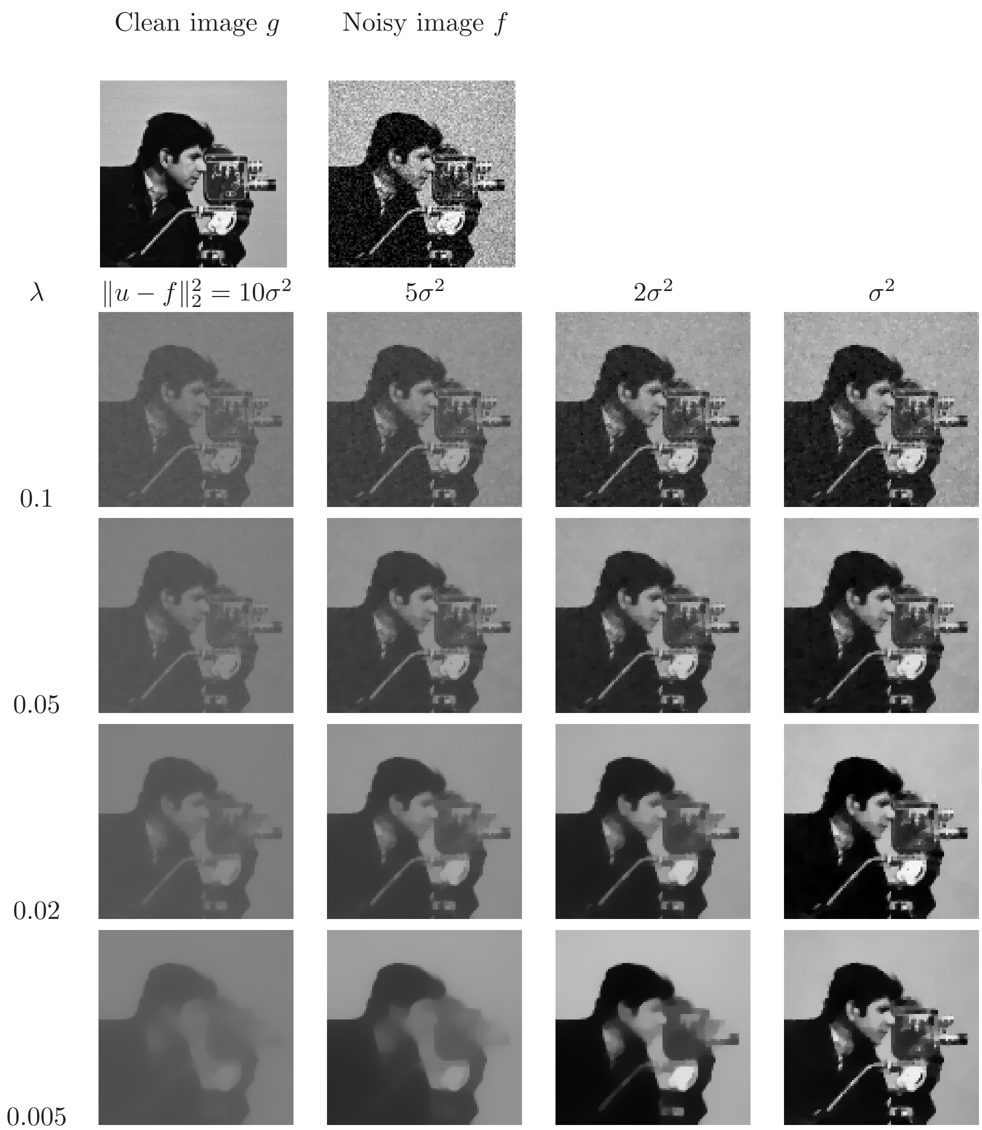


Figure 3.5: Denoising with different values of λ. Top row: original image (left) and noisy image (right). Second to fifth rows: u for the following values of λ: $0.1, 0.05, 0.02, 0.005$, respectively. Each column, from left to right, depicts the following L^2 norm of the residual part $\|u-f\|_{L^2}^2$: $10\sigma^2, 5\sigma^2, 2\sigma^2, \sigma^2$, respectively. [$u(0) = v(0) = 0$, $\alpha = \lambda/4$, $\sigma = 20$].

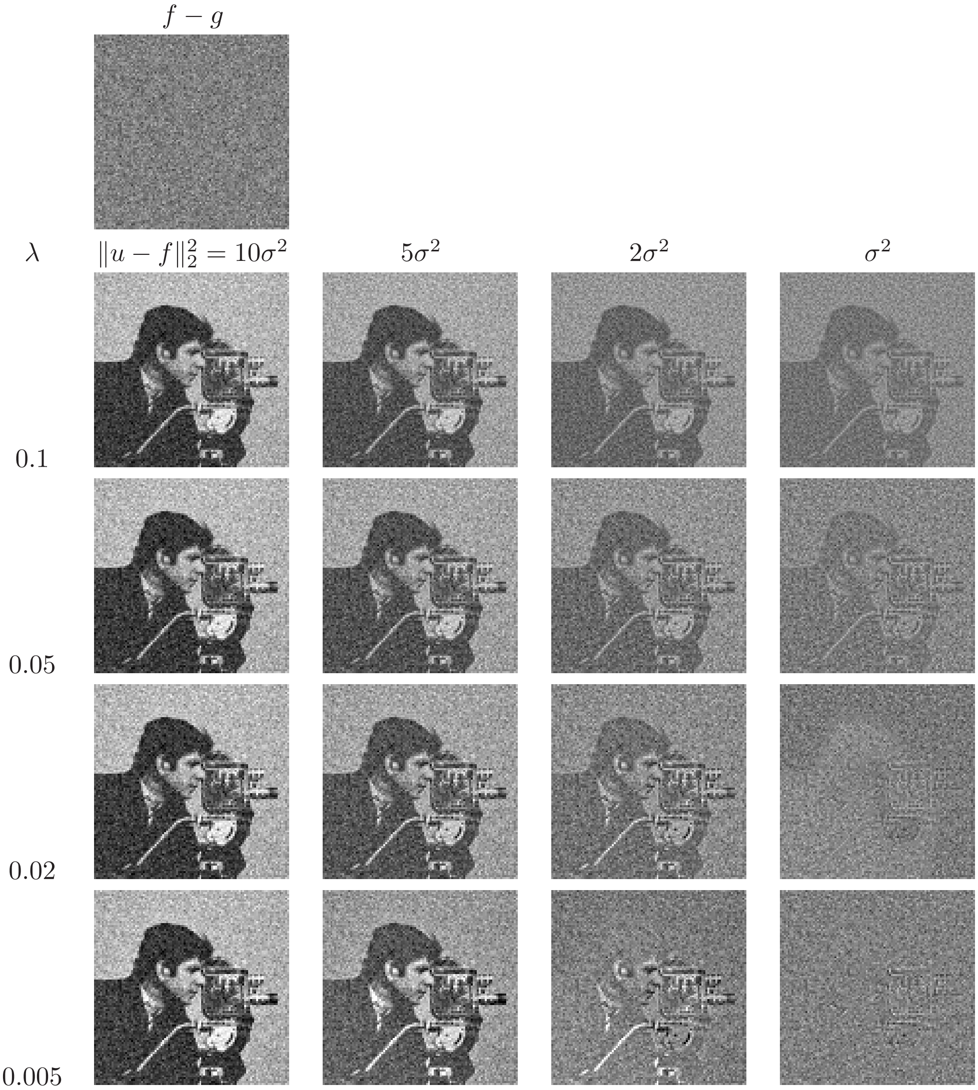


Figure 3.6: Denoising with different values of λ. Top row: $f-g$, an instance of white Gaussian noise ($\sigma = 20$) which was added to the clean image. Second to fifth rows: $f-u$ for the following values of λ: $0.1, 0.05, 0.02, 0.005$, respectively. Each column, from left to right, depicts the following L^2 norm of the residual part $\|u-f\|_{L^2}^2$: $10\sigma^2, 5\sigma^2, 2\sigma^2, \sigma^2$, respectively.

3.5.2.4 Complexity

The relaxed inverse scale space flow, in most cases, has about the same complexity as the standard gradient descent to steady state approach of ROF. The rate of the flow depends on λ and α and the evolution time monotonically increases with the values of these parameters decreasing. For most applications, however, the implementation is efficient enough and much faster than an equivalent series of Bregman iterations. For the linear case of the direct inverse scale space flow, as shown in [SG01], we obtain a step size of order one yielding stability, i.e., no severe restriction on large time steps.

3.5.2.5 Disk example

Analyzing the evolution of a disk image can be very illuminating, since the characteristic function f of a disk is a basic shape with respect to the BV semi-norm (e.g. satisfies the source condition $q \in \partial J(f) \cap L^2$) and in some cases allows a direct computation of a solution (cf. e.g. [Mey01] for the ROF model). As we shall see also for inverse scale space methods in the following, this example can provide insight into the multiscale properties of the flow, at least for piecewise constant images with smooth discontinuity sets.

Here we will analyze both the direct and the relaxed flows (which are based on the ROF energy).

For the sake of simplicity we shall restrict our attention to the case of $\Omega \subset \mathbb{R}^2$ being the ball of radius R_0 . We start with the indicator function of height h and rescale it to a function of mean zero in order to apply the above analysis, i.e., we assume

$$f(x) = \begin{cases} h\left(1 - \dfrac{R^2}{R_0^2}\right) & \text{if } |x| < R, \\ -h\dfrac{R^2}{R_0^2} & \text{otherwise} \end{cases} \tag{3.34}$$

for $h \in \mathbb{R}$ and $R \in (0, R_0)$. For convenience we denote $c_0 = 1 - \frac{R^2}{R_0^2}$ and without loss of generality we assume that $h > 0$.

Let us start with a simple property of subgradients of the total variation functional at f:

Proposition 3.5.2. *Let $J : BV(\Omega) \to \mathbb{R}$ be the total variation seminorm and let f be defined via* (3.34)*. Then the function p defined via*

$$p(x) = \begin{cases} \dfrac{2}{R} & \textit{if } |x| < R, \\ -\dfrac{2R}{c_0 R_0^2} & \textit{otherwise} \end{cases} \tag{3.35}$$

satisfies $p \in \partial J(f)$.

Proof. It is straightforward to compute

$$\int_\Omega p = 0, \quad \int_\Omega pf = J(f) = 2\pi Rh.$$

Now let q be the unique solution with mean zero of the problem

$$\begin{aligned} -\Delta q &= p \quad \text{in } \Omega, \\ \partial_n q &= 0, \quad \text{on } \partial\Omega. \end{aligned}$$

By a simple computation in polar coordinates it is straightforward to see that

$$\nabla q = q_r \left(\frac{x}{\sqrt{x^2+y^2}}, \frac{y}{\sqrt{x^2+y^2}} \right),$$

with the scalar function

$$q_r(x) = \begin{cases} \dfrac{\sqrt{x^2+y^2}}{R} & \text{if } |x| < R, \\ -\dfrac{R}{c_0 r}\left(1 - \dfrac{r^2}{R_0^2}\right) & \text{otherwise.} \end{cases}$$

One observes that $\|\nabla q\|_\infty \leq \|q_r\|_\infty \leq 1$ and hence, by the definition of the total variation functional

$$\begin{aligned} \int_\Omega p(u-f) &= \int_\Omega \nabla \cdot (-\nabla q) u - J(f) \\ &\leq \sup_{\vec{\zeta}, \|\vec{\zeta}\|_\infty \leq 1} \int_\Omega \nabla \cdot \vec{\zeta}\, u - J(f) = J(u) - J(f), \end{aligned}$$

which implies that p is indeed a subgradient. □

This result shows that the subgradient has the same structure as f, namely p a piecewise constant function with discontinuity at the circle with radius R and p has mean zero. This motivates to look for solutions of the form

$$(u(x,t), p(x,t), v(x,t)) = \begin{cases} c_0(u_1(t), p_1(t), v_1(t)) & \text{if } |x| < R, \\ (c_0 - 1)(u_1(t), p_1(t), v_1(t)) & \text{otherwise.} \end{cases}$$

We start with the original inverse scale space method, where the above Ansatz yields the ODE

$$\frac{dp_1}{dt}(t) = h - u_1(t), \qquad p_1(0) = u_1(0) = 0.$$

By the same technique as in the proof of Proposition 3.5.2 it is easy to see that for $|p_1| \leq \frac{2}{Rc_0}$, $p \in \partial J(0)$. Thus, in the initial phase of the evolution, where p_1 is still small we will always have $u \equiv 0$ and $p \in \partial J(0) = \partial J(u)$. We denote this time interval by $(0, t_1)$ and look for a solution with $u \equiv 0$ for $t < t_1$. This means that

$$\frac{dp_1}{dt}(t) = h \quad \Rightarrow \quad p_1(t) = th.$$

Since by the above argument we need $|p_1(t)| \leq \frac{2}{Rc_0}$ for $t \leq t_1$, this yields

$$t_1 = \frac{2}{Rc_0 h}, \tag{3.36}$$

and by a simple integration

$$p_1(t_1) = t_1 h = \frac{2}{Rc_0}.$$

From Proposition 3.5.2 we obtain that $p(t_1) \in \partial J(f)$ and hence, we can continue the solution via $p_1(t) = p_1(t_1)$ and $u_1(t) = h$ (and thus $u(\cdot, t) = f$) for $t > t_1$, which is clearly a solution since

$$\frac{dp_1}{dt}(t) = f - u_1(t) = 0, \quad p(t) \in \partial J(f) = \partial J(u(\cdot, t)).$$

Hence, we have found a solution for the original flow in this way, and in particular the reconstructed image satisfies

$$u(\cdot,t) = \begin{cases} 0 & t < t_1, \\ f & t > t_1. \end{cases} \tag{3.37}$$

This means that after an initial time interval of length t_1, where the image remains zero and just the dual variable p grows, the reconstruction suddenly jumps to the correct image f and remains constant afterwards. The length of the time interval t_1 needed to obtain the correct image also gives an indication of the multiscale properties of the model. Note that the scale of the image can be characterized by the product of radius and height, i.e., by Rh. Our analysis shows that t_1 is inversely proportional to Rh (note that c_0 is of order 1 if R_0 is sufficiently large) and hence, larger scales appear faster than smaller ones. This property can be seen as the fundamental reason why the inverse scale space method is a good denoising technique, since it first reconstructs the large scale features and only later the very small scale ones (which are usually caused by noise). An appropriate stopping rule as the one proposed above will ensure that the flow is stopped before too small scales enter.

It is rather straightforward to extend the above reasoning to the relaxed inverse scale space method. If we look for an initial time period ($t < t_1$) where $u \equiv 0$, then we are led to the ODE (with the above notation):

$$\frac{dv_1}{dt} = \alpha h, \quad p_1 = \lambda(h + v_1), \quad p_1(0) = v_1(0) = 0.$$

Hence, $v_1(t) = \alpha h t$ and $p_1(t) = \lambda h(1+\alpha t)$. We know that $p \in \partial J(0)$ if $|p_1| \leq \frac{2}{Rc_0}$, which means that we obtain

$$t_1 = \begin{cases} \dfrac{2 - Rc_0 h\lambda}{\alpha\lambda Rhc_0} & \lambda \leq \dfrac{2}{Rhc_0}, \\ 0 & \text{otherwise.} \end{cases} \tag{3.38}$$

Note that as $\lambda \to 0$, we obtain the same value for $\alpha\lambda t_1$ as for the critical time t_1 in the original flow, Eq. (3.36) (the additional factor $\alpha\lambda$ corresponds again to the time rescaling discussed before). In the second phase ($t > t_1$) we know that $p_1 = \frac{2}{Rc_0}$ and therefore we can write the evolution as the coupled system

$$\frac{dr}{dt}(t) = q(t) - \lambda r(t), \quad \frac{dq}{dt}(t) = -\alpha\lambda r(t) \tag{3.39}$$

with $r(t) := u_1(t) - h$ and $q(t) := \lambda v_1(t) - \frac{2}{Rc_0}$. The eigenvalues of this linear dynamical system are given by

$$e_\pm = -\frac{\lambda}{2} \pm \frac{1}{2}\sqrt{\lambda(\lambda - 4\alpha)}.$$

Similar to the linear analysis, in order to have solutions with real roots we require $\lambda \geq 4\alpha$.

In the case of unbounded domain $R_0 \to \infty$ and having $\alpha = \frac{\lambda}{4}$, the solution of the disk problem for the relaxed flow is

$$u(x,t) = \begin{cases} 0, & 0 \leq t \leq t_1 \\ \left(-(\frac{\lambda h}{2}(t - t_1) + h)e^{-(t-t_1)\lambda/2} + h\right) f(x), & t_1 < t < \infty \end{cases} \tag{3.40}$$

with t_1 as defined in (3.38). We will use this simple equation as a reference in the following numerical experiment.

3.5.2.6 Disk Numerical Experiment

Below we display the results of a numerical experiment of an evolution of a disk. The disk radius is $R = 10$ and its height is $h = 1$. The evolution is for $t \in [0, 125]$. For simplicity, in order not to implement a circular domain, we take a large rectangular Ω and use the model for $R_0 \to \infty$, $c_0 \to 1$. For $\lambda = 0.12, \alpha = \frac{\lambda}{4}$ the disk should start to appear at $t_1 = 22.2$. In Figure 3.7 the original disk f and $u(t)$ at some time along the evolution ($t = 50$) are shown. In the image of u we also super-imposed a cross-section at the center of the disk and the center point of the disk, for which the values are plotted (Figure 3.8). In Figure 3.8, left, the values of a cross-section of the disk are plotted for 40 equally spaced time points. On the right of Figure 3.8 the theoretical model of equation (3.40) (dashed, red) is compared to the simulation at the central point of the disk $u(x_p, y_p, t)$ (solid, blue). In our case the mean value is not zero (here the domain is 100×100 pixels and therefore the mean value is approximately $\frac{\pi}{100}$). Therefore the initial condition $u(x, y, 0)$ (plotted in dashed line near 0) is different than Eq. (3.40). Apart from that, the evolution is quite faithful to the model.

3.5.2.7 The Initial Phase for General Images

In the following we are going to extend the results for the disk to general images, at least the behaviour in the initial stage. We assume that $f \in L^2$ is a function with mean zero. Following Meyer [Mey01] the $*$-norm (the dual of the BV seminorm) is defined as

$$\|w\|_* = \inf_{\vec{\xi}:\nabla\cdot\vec{\xi}=w} \sup_{x,y} |\vec{\xi}|$$

Then we can use a characterization of subgradients obtained in [Mey01] namely that $p \in \partial J(0)$ if and only if $\|p\|_* \leq 1$. Thus, we immediately obtain the following

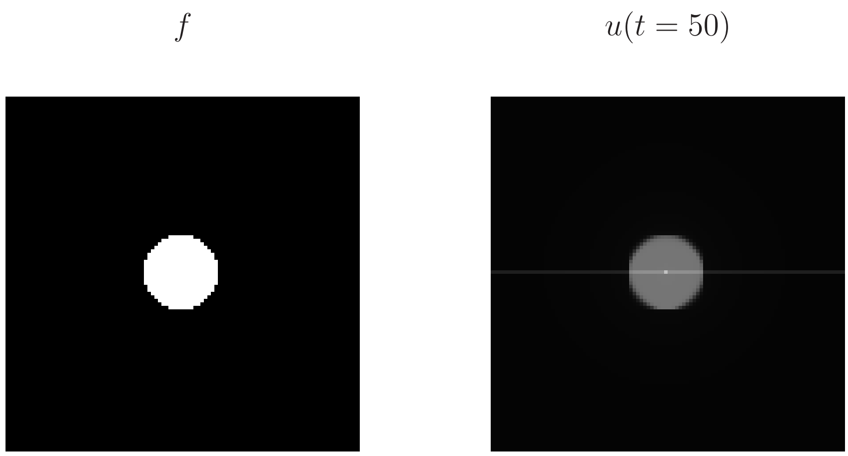


Figure 3.7: Disk evolution: f (left) and u (right). The brighter line and point in u show which values of u are plotted in Figure 3.8.

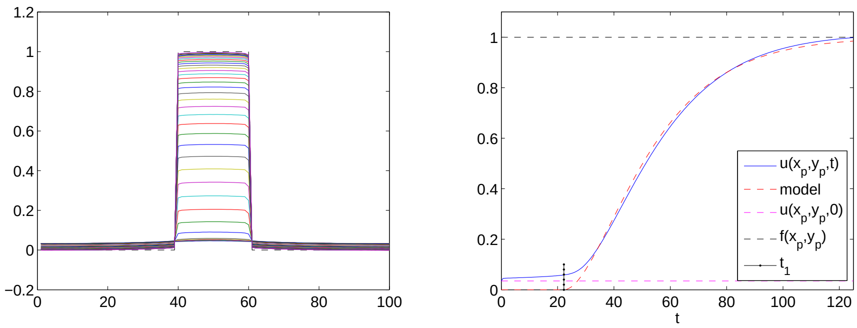


Figure 3.8: Disk evolution. Left: plots of a cross-section of u at equally spaced time points. Dashed - cross-section of f. Right: comparison between the simulation (solid, blue) and the theoretical model (dashed, red).

generalization of the behaviour in the initial phase

Theorem 3.5.3. *Let $f \neq 0$ be as above and let $t_1 = \frac{1}{\|f\|_*}$. Then the solution (u,p) of* (3.5) *satisfies*

$$u(\cdot,t) \equiv 0, \quad p(\cdot,t) = tf, \qquad \text{for } t < t_1.$$

Moreover, t_1 is maximal with this property, i.e., $u(\cdot,t)$ is not identically 0 for $t > t_1$.

Proof. For $t < t_1$ we obtain that $p(\cdot,t) \in \partial J(0)$ from the above reasoning and one easily checks that (u,p) defined as above is a solution of (3.5). Now assume that t_1 is not maximal, i.e., $u \equiv 0$ in the time interval $(0,t_2)$ for $t_2 > t_1$. Then from (3.5) we obtain $p = tf$, but $p(\cdot,t) \notin J(0)$ for $t > t_1$ a contradiction. Hence, u is not identically 0 at least in the time interval $(t_1, t_1 + \epsilon)$ for some $\epsilon > 0$. Since we know that the residual $\|f - u(t)\|_{L^2}$ is non-increasing in time (see Section 3.2.2), we deduce

$$\|f - u(t)\|_{L^2} \leq \|f - u(s)\|_{L^2} < \|f - 0\|_{L^2}$$

for $t > s$ and $s \in (t_1, t_1 + \epsilon)$. Hence, $u \neq 0$ for all $t > t_1$. $\square$

Note that all computations of the subgradients in the disk example were implicitly computing the $*$-norm of the functions f and 0, so this generalization is not completely surprising.

We can also give a multiscale interpretation of Theorem 3.5.3. We can have f scaled such that $\|f\|_{L^2} = 1$. These properties can always be achieved by rescaling for f different from a constant (and if f is constant the inverse scale space method and its relaxed version are both stationary at the correct image anyway). We can write then t_1 as the ratio $t_1 = \frac{\|f\|_{L^2}}{\|f\|_*}$ which actually can be considered as a definition of scale.

Since the L^2-norm is stronger than the $*$-norm, it will be large for high frequency (small scale) features and small for low frequency (large scale) ones. This again explains to some point why large scale features are incorporated earlier than small scale ones. We also mention that by a standard inequality for dual norms (cf. [Mey01, p.32] or [AE84]) we have for "clean images" $f \in BV(\Omega)$

$$t_1 = \frac{\|f\|_{L^2}}{\|f\|_*} \leq \frac{J(u)}{\|f\|_{L^2}},$$

which yields a similar interpretation of scales in terms of the ratio of total variation and the L^2-norm, e.g., in the disk example above one obtains equality in these ratios as $R_0 \to \infty$.

In the following proposition we state the analogue property which holds for the relaxed flow:

Proposition 3.5.4. *Let $f \neq 0$, $\int_\Omega f = 0$, $\lambda\|f\|_* \leq 1$ and let $t_1 = \frac{1-\lambda\|f\|_*}{\alpha\lambda\|f\|_*}$. Then each solution (u, v) of (3.25) satisfies*

$$u(\cdot, t) \equiv 0, \quad v(\cdot, t) = \alpha t f, \qquad \textit{for } t < t_1.$$

Proof. Let us define the following energy

$$E(t) := J(u) + \frac{\lambda}{2}\|f + v - u\|_{L^2}^2.$$

Then the flow of u in (3.25) can be viewed at each time point as a steepest descent of this energy.

Using the decomposition result of [Mey01, p. 32] and the condition $\lambda\|f\|_* \leq 1$ we can verify that the initial condition $u(0) = 0$, $v(0) = 0$ at time $t = 0$ is a stationary point, where the energy $E(0) = J(0) + \frac{\lambda}{2}\|f\|_{L^2}^2$ is minimal. Therefore $\partial_t u|_{t=0} = 0$ and consequently $u|_{t=0^+} = 0$. By a similar argument u will stay zero as long as $\lambda\|f+v\|_* \leq 1$. Solving for v we have a simple ODE $\frac{d}{dt}v = \alpha f$, yielding $v(t) = \alpha f t$ for $t \in (0, t_1)$. This evolution is valid until at some time t_1 we have

$$\lambda\|f + v(t_1)\|_* = 1.$$

As $f \neq 0$, $\|f + v\|_* = (1 + \alpha t)\|f\|_*$ is increasing with time and this equality will be reached in a finite time. $\square$

Note that as in the disk example, we can obtain the time of appearance t_1 of the direct solution (stated in Theorem 3.5.3) by multiplying the expression for t_1 of the relaxed flow by $\alpha\lambda$ and letting $\lambda \to 0$.

3.6 Numerical Examples

In this section we present some numerical examples. Some of the results were previously presented in [BOX05, BGO06]. We show an example of a one-dimensional (1D) problem solved by the direct inverse scale space flow (DISS, Section 3.3) and by the relaxed inverse scale space flow (RISS, Section 3.4, Eq. (3.25)) in order to test and compare their behaviors. The 1D example reveals a striking resemblance of the direct and relaxed processes, which well justifies our interpretation of the relaxed flow as a good approximation of the direct flow. Motivated by the agreement between the one-dimensional results, we proceed by processing images using the relaxed flow.

We focus on denoising by ROF-based flows. A single example of image decomposition by the TV-L^1-based inverse scale space flow (Eq. (3.26)) is also given. Further generalizations and experiments with other J and H functionals are currently studied and will appear in the future work.

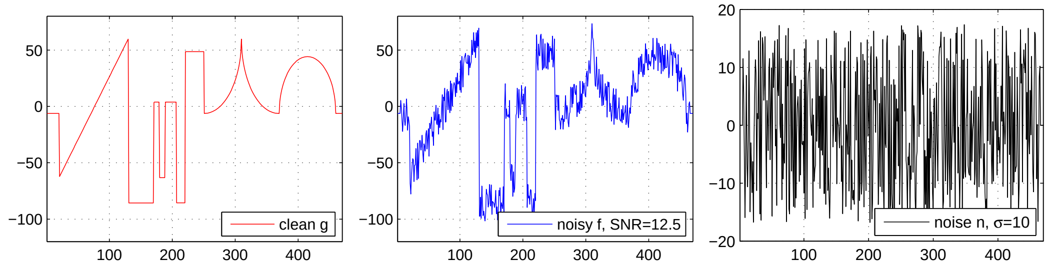


Figure 3.9: 1D signal denoising. From left to right: clean signal g, noisy signal f ($SNR = 12.5$); Gaussian noise n, $\sigma = \|n\|_{L^2} = 10$.

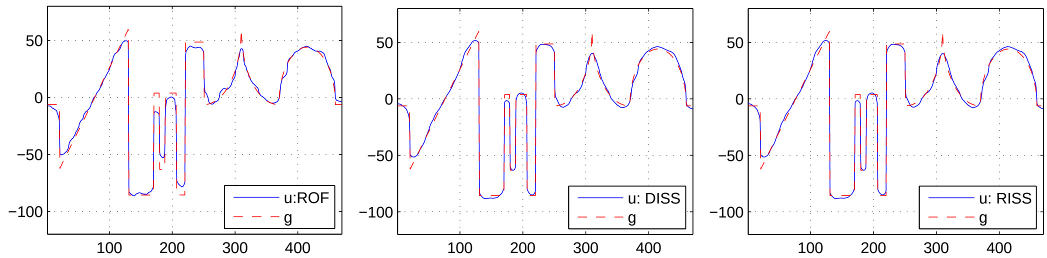


Figure 3.10: 1D signal denoising. From left to right: restored signal u from ROF, the direct inverse flow (DISS) and the relaxed inverse flow (RISS). SNR: ROF(u)=17.73, DISS(u)=21.94, RISS(u)=21.95.

Example 1 (1D signal): We first consider a 1D denoising problem. Figure 3.9 shows the clean signal g, the noisy signal f and the Gaussian noise n ($\sigma = \|n\|_{L^2} = 10 \approx 25\%\|g\|_{L^2}$). Figure 3.10 shows the solutions u obtained by ROF, the direct inverse flow (DISS) and the relaxed inverse flow (RISS). Figure 3.12 shows a comparison between these solutions at a region [180,220]. The typical signal loss can be observed in the result of ROF, and, as expected, the loss is much smaller for the inverse scale space flows. The signal-to-noise-ratio (SNR) of the inverse flow results (21.94 and 21.95) are also much higher than that of ROF (17.7). This supports our theoretical arguments (see Proposition 3.2.1 and the following discussion) that the TV-based inverse scale space flow yields better restorations than the original ROF model.

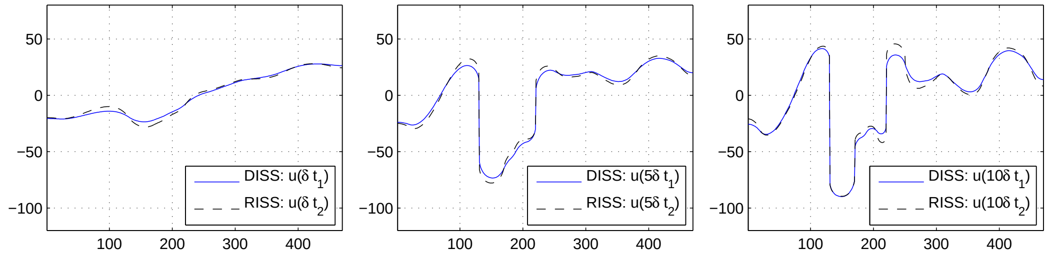


Figure 3.11: 1D signal denoising. Comparisons between the evolutions of DISS and RISS flows at different time points δt, $5\delta t$ and $10\delta t$ ($\delta t_1 = 2 \times 10^{-4} = 2 \times 10^5 \Delta t_1$ for DISS and $\delta t_2 = 10^5 = 2 \times 10^5 \Delta t_2$ for RISS).

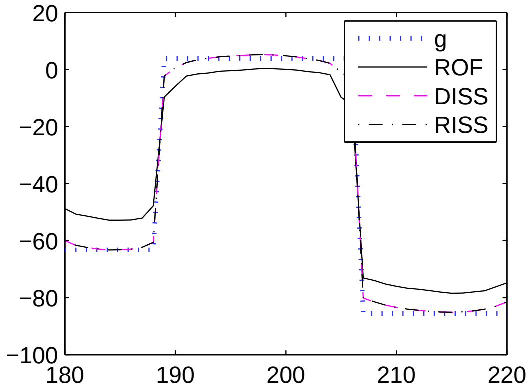


Figure 3.12: 1D signal denoising. Comparison between the solutions u from ROF, DISS flow and RISS flow (part, at $[180, 220]$).

In Figure 3.11 we show a comparison between the evolutions of DISS and RISS flows at three different time points. We can see that they agree to each other very well. In Figure 3.12 one part ([180,220]) of the three solutions are plotted on the same grid, where the direct and relaxed solutions virtually coincide although the flows' equations and their implementations are very different. Both SNR results are also almost identical. This validates our view of the relaxed flow as a faithful representation of the direct one. We choose $\epsilon = h = 1$ for all three experiments, which is a relatively large value, due to the sensitivity of DISS to numerical errors for small ϵ. Moreover, we used $\lambda = 0.01$ for ROF, $\Delta t_1 = 10^{-9}$ for DISS, $\lambda = 10^{-4}, \Delta t_2 = 0.5$ for RISS. The difference of Δt in the two inverse flow experiments are only due to the different scaling.

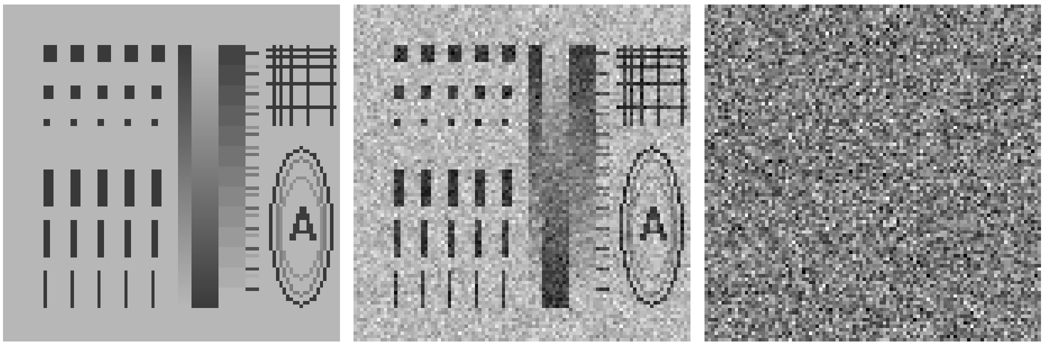


Figure 3.13: Synthetic shape image. From left to right: original image g; noisy image f ($SNR = 7.4$); Gaussian noise n ($\sigma = 40$).

Example 2 (Synthetic shape image): We now turn to the denoising of 2D images. In this example we consider an image with different scales and shapes and corrupted by Gaussian noise, which is shown in Figure 3.13. $SNR(f) = 7.4, \sigma = \|f - g\|_{L^2} = 40$. Figure 3.14 shows the results obtained by ROF, iterated TV refinement (Bregman ROF, Algorithm 2.1 in Chapter 2) and relaxed inverse TV flow, column-by-column respectively. The restoration result u, the corresponding residual part $w = f - u$ and an enlargement of part of w are displayed for each method (again, to enhance the visibility, we shift w by adding 128 and then plot its image). One observes that for the ROF model visible components of the clean signal are contained in w (e.g. the small blocks and grids) whereas almost no trace of the signal is visible in the other two models. In this example RISS gave the best SNR result: $SNR(u) = 9.9, 11.8, 12.5$ for ROF, iterated refinements and RISS flow respectively.

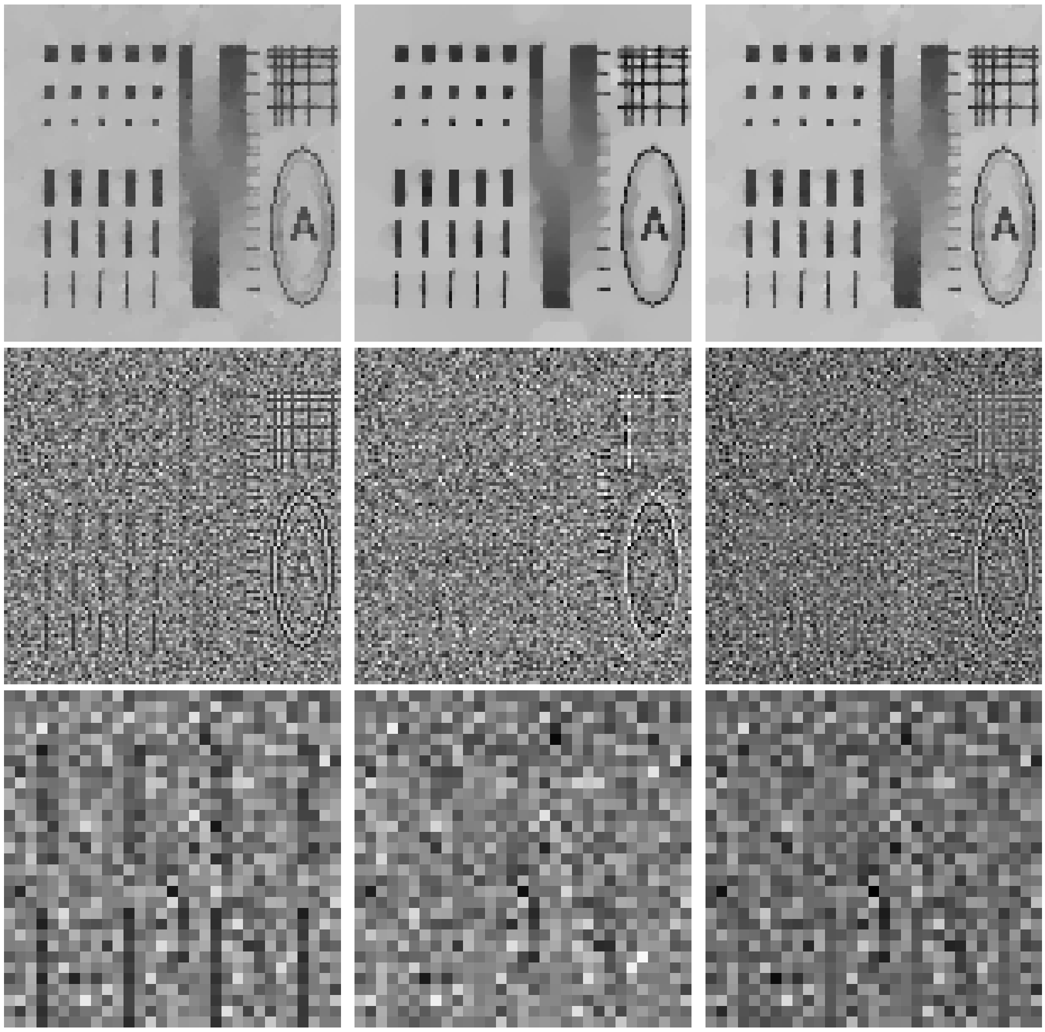


Figure 3.14: Synthetic shape image. Denoising results. From top to bottom: denoised u, residual $w = f - u$ and part of w from ROF, Bregman ROF and RISS flow (column-by-column). SNR : ROF(u)=9.9, Bregman(u)=11.8, RISS(u)=12.5.

Natural images are being processed in the following examples.

Example 3 (Cameraman):. In Figure 3.15 we compare the denoising of the Cameraman image by ROF and RISS. RISS retains great contrast, which is most visible in the residual part $f - u$ (bottom row), where the coat, tripod and camera details are much less degraded. Both methods have the same L^2 norm residual $\|f - u\|_{L^2} = \sigma$. SNR results, ROF: 15.76, RISS: 16.42.

Example 4 (Sailboat): In Figures 3.16 and 3.17 another comparison is made between ROF and RISS. Here it is very clear that thin lines, which get eroded by ROF, are better preserved by our method (e.g. the poles and the number on the sail). Again, in both methods we have $\|f - u\|_{L^2} = \sigma$. SNR results, ROF: 11.43, RISS: 11.98.

In Figure 3.18 some more information of the evolution of Examples 3 and 4 is given. Three performance criterion which measure the closeness of u to the clean image g are shown as a function of the evolution time t. In the first row SNR is plotted (which is based on the L^2 distance between g and u), the second row depicts the L^1 distance, $\|g - u\|_{L^1}$, and the third row depicts the Bregman distance $D(g, u)$. From the value of SNR and the Bregman distance we can see that $u(t)$ is approaching g. In the fourth row the L^2 convergence of u to the noisy image f is plotted as a function of time. One can observe that $\|f - u\|_{L^2}$ is monotonically decreasing in time and it is straightforward to select a stopping time based on the discrepancy principle.

Example 5 (TV+L^1 Decomposition, Barbara): In Figure 3.19 we present one example of image decomposition using the TV-L^1 inverse scale space flow. Eq. (3.26) is evolved in order to separate a clean image f into its geometrical part u and its textural part $w = f - u$. The stopping time in this case was chosen manually. We note that qualitatively similar results were obtained within

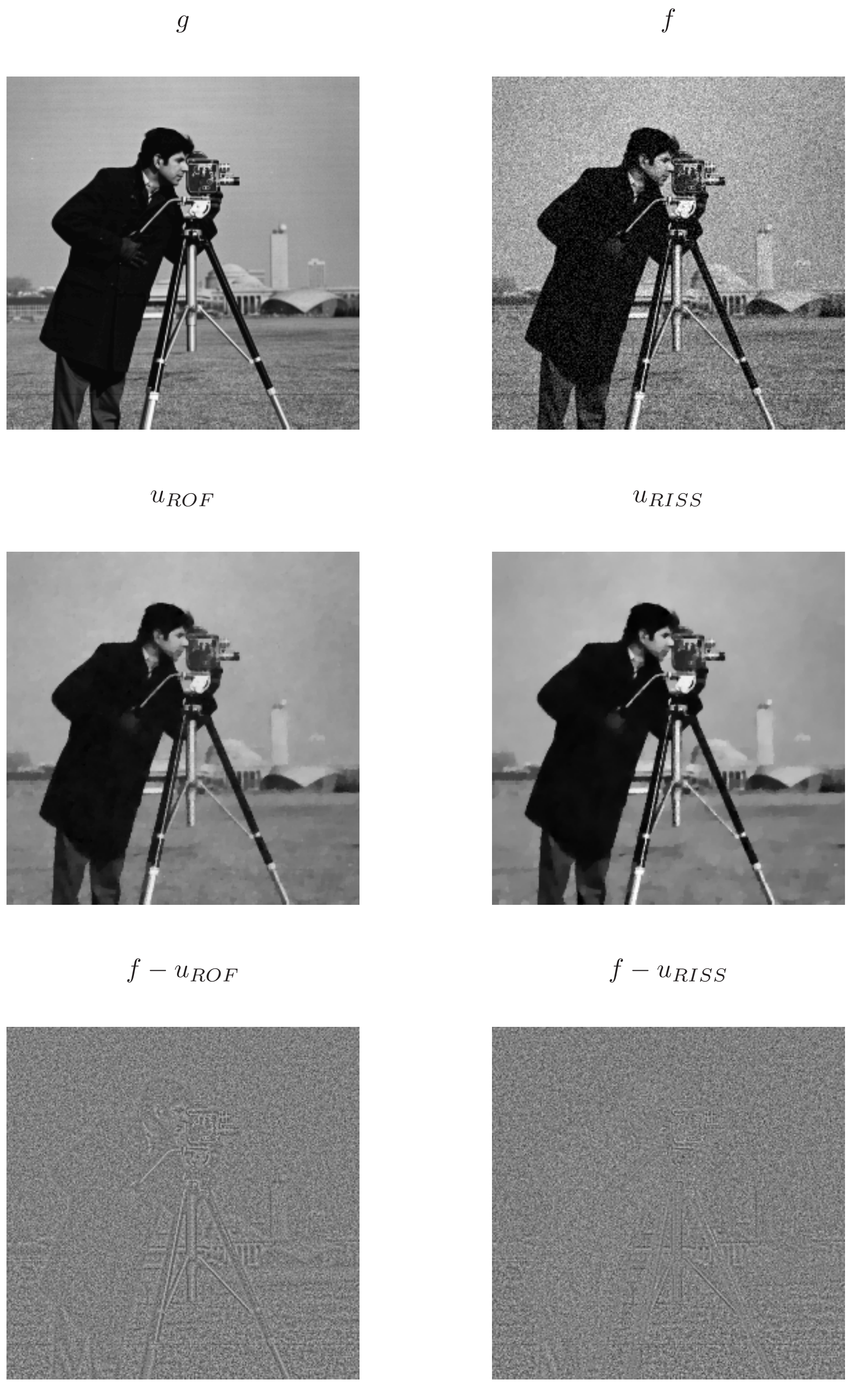


Figure 3.15: Cameraman image ($\sigma = 20$, SNR=9.89). Top row: clean image g (left), noisy image f (right). Second row: denoised image u by ROF, SNR=15.76 (left) and by RISS flow (Eq. (3.25)), SNR=16.42 (right). Bottom row: corresponding residual parts $w = f - u$. [$\lambda = 0.01$].

g f

Figure 3.16: Sailboat image (σ = 20, SNR=4.40). Clean image g (left), noisy image f (right).

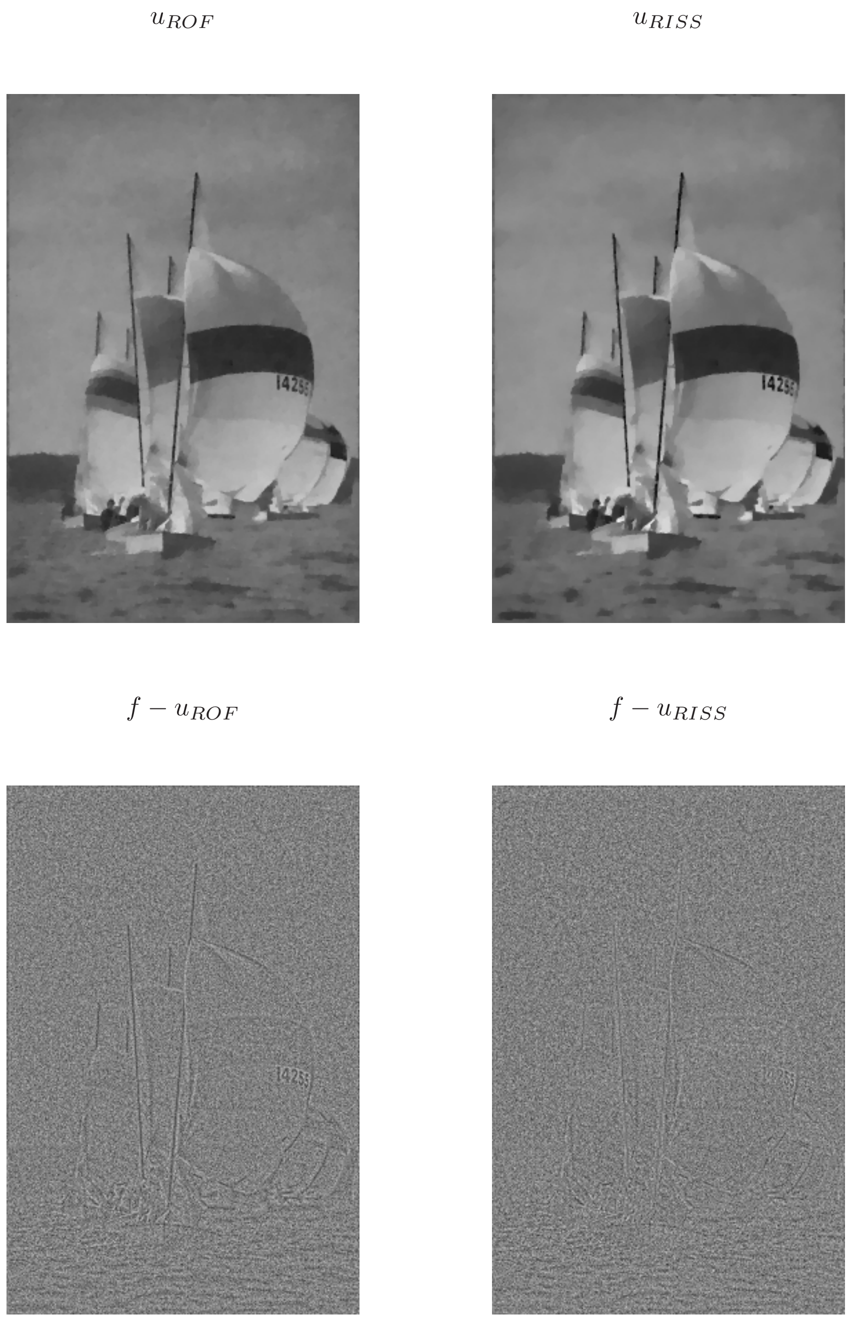


Figure 3.17: Sailboat image (cont'). Top row: denoised image u by ROF, SNR=11.43 (left) and by RISS flow (Eq. (3.25)), SNR=11.98 (right). Bottom row: corresponding residual parts $f - u$. [$\lambda = 0.01$].

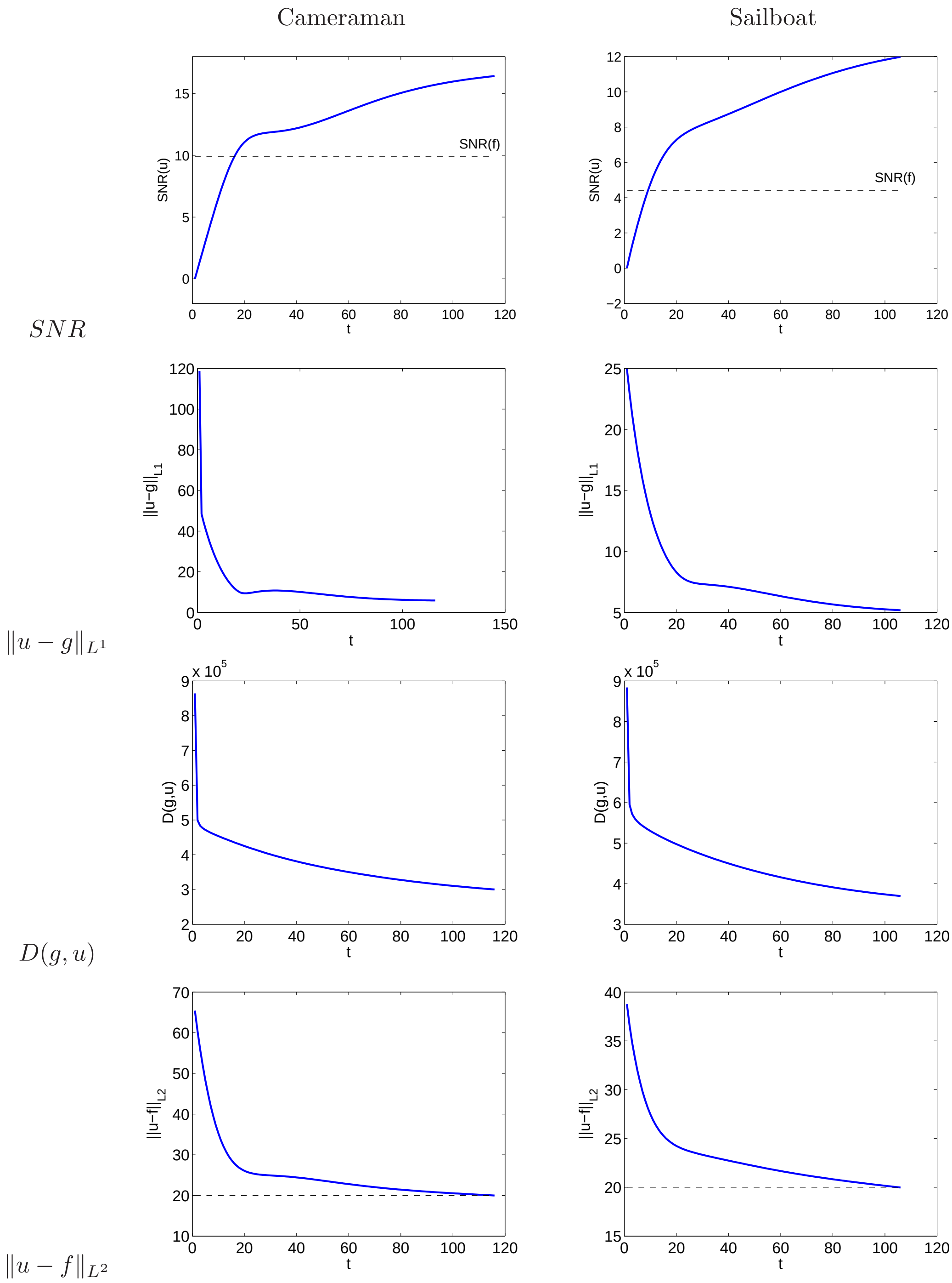


Figure 3.18: Performance and convergence plots as a function of time. From top: $SNR(u)$, $\|u-g\|_{L^1}$, $D(g,u)$, $\|u-f\|_{L^2}$. Left – Cameraman image, right – Sailboat image.

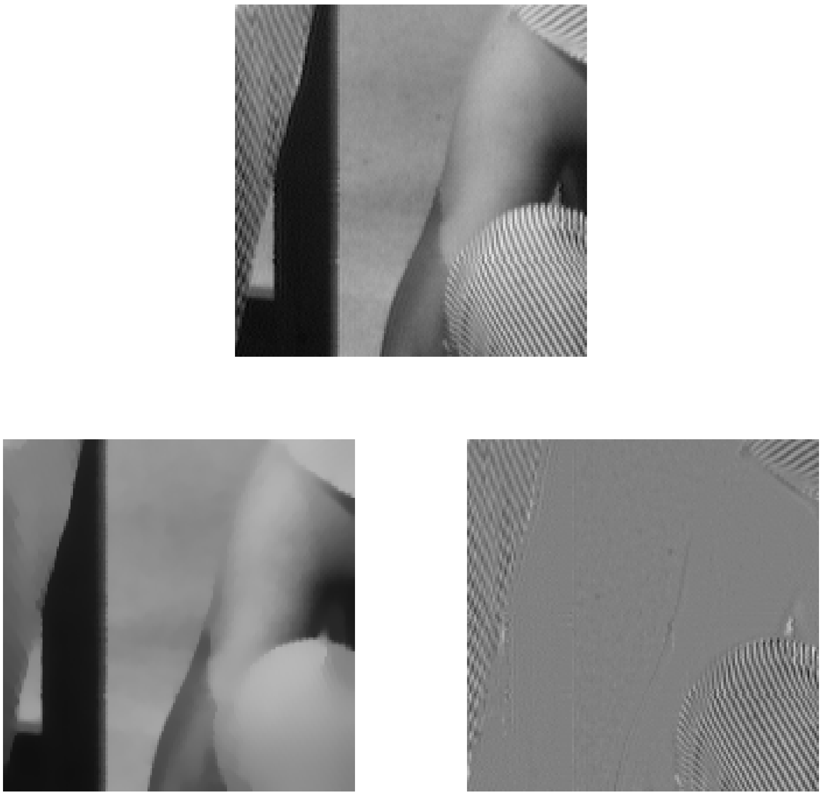


Figure 3.19: Decomposition of part of Barbara using TV-L^1 inverse scale space flow (Eq. (3.26)). Top: original. Bottom: geometric part u (left) and textural part $f - u$ (right). $[\lambda = 0.02]$.

quite a large evolution duration, therefore it seems that the process is not sensitive to a very specific choice of the stopping time in order to obtain meaningful decomposition results. Further study of this evolution and comparison to other decomposition methods will be studied in the future.

3.7 Discussion and Conclusion

Two new types of nonlinear processes (inverse scale space flow and relaxed inverse scale space flow) are presented in this chapter for image simplification and regularization. Both extend the Bregman iterations procedure introduced in previous chapter to a continuous formulation, creating stable flows going from a zero signal to the input image.

Two basic characteristics distinguish these flows from the various variations of forward linear and nonlinear scale spaces (e.g. [ALM92, GSZ04, KMS00, PM90, Wei99, Wit83]): First, the flows advance in the inverse direction from the most simplified image (zero or mean value) to the most detailed image (input image). This allows fast denoising of large objects, which appear very early in the evolution. Second, the flows are based on both energy terms - the regularization term $J(u)$ and the fidelity term $H(u,f)$. This is different from forward scale space methods which, at least in some cases, can be viewed as steepest descent flows of the regularization term $J(u)$. Thus it is possible to construct new PDE-based evolutions for problems which until now were solved primarily in the variational setting. For example, one may evolve a deconvolution scale space (with $H = \frac{1}{2}\|f - Ku\|^2_{L^2}$) or to have a flow based on the L^1 fidelity term for removal of impulsive noise or for structure-texture decomposition [Nik04, CE05, YGO05]. Other types of fidelity terms may be considered in the future, for instance ones based on the $*$-norm [AAB05, Mey01, VO03], H^{-1} norm [OSV03] or on Gabor functions [AGC06]. A scale space, as opposed to the variational setting, naturally introduces a continuous set of solutions. Whereas for denoising usually a single solution is selected, for decomposition or segmentation purposes several solutions may be preferred, understood as a multiscale representation of the input image.

The proposed direct inverse scale space flow is based on evolving in time the subgradient of the regularized image u. Various theoretical properties are shown concerning the convergence of the flow to the input f. Moreover, the monotonic approach of u to the clean image g (in the Bregman distance sense) is proved as long as the L^2 norm of the residual is larger than that of the noise. This well justifies theoretically the use of a discrepancy principle as a stopping criterion. To the best of our knowledge, no similar property is available in any forward scale space evolution. We have presented a way to compute the direct flow in one dimension.

The relaxed inverse scale space flow can be viewed as either two coupled equations which are first order in time or as a single second order in time PDE. Its implementation is very standard and can be achieved by applying the ordinary numerical techniques used in variational minimizations. Although further theoretical study is needed we have shown the similarity of the relaxed flow to the direct flow for small λ (after time rescaling). Numerical solutions in one dimension (in which both flows could be computed and compared) indicate a very high degree of similarity of the flows. Convergence of the relaxed flow to the input image f was proved by Lie and Nordbotten [LN05]. The flow produces excellent denoising results and retains very good contrast of larger objects (in some cases contrast may be even slightly enhanced). Open questions concerning the relaxed flow include whether in some sense u approaches the clean image g (as shown for the direct flow) and whether $\|f - u(t)\|_{L^2}$ is monotonically decreasing.

The numerical indications are promising. The inverse scale space flow can be generalized to many types of variational models and inverse problems, e.g., see a recent paper [BFO06] by Burger et. al.. In next chapter we will generalize both ideas of iterative regularization and inverse scale space to wavelet-based image restoration.

Acknowledgments

Useful and stimulating discussions are acknowledged to Mark Green and Barry Merriman (both UCLA).

CHAPTER 4

Iterative Regularization and Inverse Scale Space Applied to Wavelet Based Image Denoising

In this chapter we generalize the iterative regularization procedure and inverse scale space to wavelet based image denoising.

4.1 Introduction

Given an orthonormal wavelet basis $\{\psi_{\mathbf{j}}(x)\}$, $\mathbf{j} = (j_1, j_2, j_3)$, which is generated by $\{\psi^{(j_3)}(x)\}_{j_3=1}^{2^d-1}$ as $\psi_{\mathbf{j}}(x) = 2^{j_1}\psi^{(j_3)}(2^{j_1}x - j_2)$, $j_1 \in \mathbb{Z}, j_2 \in \mathbb{Z}^d, x \in \mathbb{R}^d$, the wavelet transform of an image f can be represented as (cf., e.g., [Dau92, Mey92, CDV93]):

$$f = \sum_{\mathbf{j}} \tilde{f}_{\mathbf{j}}\psi_{\mathbf{j}} = \sum_{\mathbf{j}} \langle f, \psi_{\mathbf{j}} \rangle \psi_{\mathbf{j}},$$

where $\langle \cdot, \cdot \rangle$ is the L^2 inner product. We denote $\tilde{f} = \{\tilde{f}_{\mathbf{j}}\}$.

In general, wavelet shrinkage attempts to denoise images via the following three steps (cf., e.g., [DJ94, Don95]):

(1) Analysis. Transform the noisy image f to the wavelet coefficients $\tilde{f} = \{\tilde{f}_{\mathbf{j}}\}$;

(2) Shrinkage. Apply a shrinkage operator $\mathcal{D}$ with a threshold parameter τ related to the noise level to the wavelet coefficient $\tilde{f}$: $\tilde{u} := \mathcal{D}_\tau(\tilde{f})$;

(3) Synthesis. Reconstruct the denoised solution u from the shrunken wavelet coefficients:

$$u = \sum_{\mathbf{j}} \tilde{u}_{\mathbf{j}} \psi_{\mathbf{j}} = \sum_{\mathbf{j}} \mathcal{D}_\tau(\tilde{f}_{\mathbf{j}}) \psi_{\mathbf{j}} \,.$$

Remark. *In the literature (cf., e.g., [SWB04]) the wavelet basis is divided into two parts: lowpass scaling function(s) $\varphi(x)$ and bandpass wavelet function(s) $\psi(x)$. Correspondingly the wavelet coefficients are divided into two parts: scaling coefficients (or "approximation coefficients") and detail coefficients. Then in the shrinkage step above one can choose to apply the shrinkage operator on all wavelet coefficients or only on the detail ones. In our discussion in this chapter we will consider shrinkage on all the wavelet coefficients. Since the summation in our new models is separable, we will see that our discussion can be easily generalized to the case of shrinkage on detail coefficients only.*

There are various types of shrinkage operators being discussed in the literatures. We list those that we will use here:

- Soft shrinkage (cf. [DJ94]), for $\tau > 0$

$$\mathcal{S}_\tau(w) = \begin{cases} w - \tau \mathrm{sign}(w), & \text{if } |w| > \tau, \\ 0, & \text{if } |w| \le \tau. \end{cases} \tag{4.1}$$

- Hard shrinkage (cf. [DJ94]), for $\tau > 0$

$$\mathcal{H}_\tau(w) = \begin{cases} w, & \text{if } |w| > \tau, \\ 0, & \text{if } |w| \le \tau. \end{cases} \tag{4.2}$$

- Firm shrinkage (cf. [GB97]), for $\tau_2 > \tau_1 > 0$

$$\mathcal{F}_{\tau_1,\tau_2}(w) = \begin{cases} w, & \text{if } |w| > \tau_2, \\ \dfrac{\tau_2}{\tau_2 - \tau_1}(w - \tau_1 \mathrm{sign}(w)), & \text{if } \tau_1 < |w| \le \tau_2, \\ 0, & \text{if } |w| \le \tau_1. \end{cases} \tag{4.3}$$

One important feature of the wavelet-based denoising methods is their simplicity and computational efficiency compared to, e.g. TV-based image restoration. But their performance is heavily dependent on the type of noise and there may exist some artifacts in the restored solutions, especially near edges. Some work has been done to solve these problems, e.g., cf. [DC95] for a translation-invariant denoising technique. The study of the relation between TV regularization methods and wavelet-based methods and furthermore combining their advantages is an interesting research topic and has been investigated by several authors, e.g., in [CDL98, SWB04, SED05, DT05]. In this chapter we will generalize the iterative regularization method and the inverse scale space method from previously discussed TV-based image restoration to wavelet based image restoration.

First we recall that the total variation based ROF model is

$$u = \operatorname*{arg\,min}_{u \in BV(\Omega)} \left\{ |u|_{BV} + \frac{\lambda}{2}\|f - u\|_{L^2}^2 \right\}, \tag{4.4}$$

where $\lambda > 0$ is a scale parameter, $BV(\Omega)$ is a bounded variation space equipped with a seminorm $|u|_{BV} = \int_\Omega |\nabla u|$. Then we consider the Besov space $B_1^1(L^1(\Omega))$, which contains, roughly speaking, functions with first order derivatives in $L^1(\Omega)$. (For formal definition of Besov spaces $B_q^\alpha(L^p(\Omega))$, cf., e.g., [DL92b]). It has been found that the discrete l^1 norm of the wavelet coefficients is equivalent to the norm in the Besov space $B_1^1(L^1(\Omega))$, which is a subset of $BV(\Omega)$ for $\Omega \subset \mathbb{R}^2$

(cf., e.g., [Mey90, DL92a, CDL98, CDP99, CDD03, DT05]). Now we replace the BV-seminorm in (4.4) by the l^1 norm of wavelet coefficients

$$J(\tilde{u}) = \sum_{j} |\tilde{u}_{\mathbf{j}}| \approx \|u\|_{B^1_1(L^1)}. \tag{4.5}$$

Here '≈' is used to represent the equivalence between the two norms. Using Parseval's identity we have

$$\|f - u\|^2_{L^2} = \|\tilde{f} - \tilde{u}\|^2_{L^2} = \sum_{\mathbf{j}} |\tilde{f}_{\mathbf{j}} - \tilde{u}_{\mathbf{j}}|^2. \tag{4.6}$$

We then approximate the TV-based ROF model (4.4) using the following wavelet-based method:

$$\begin{aligned} \tilde{u} &= \operatorname*{argmin}_{\tilde{u}} \left\{ J(\tilde{u}) + \frac{\lambda}{2}\|\tilde{f} - \tilde{u}\|^2_{L^2} \right\} \\ &= \operatorname*{argmin}_{\tilde{u}} \left\{ \sum_{\mathbf{j}} |\tilde{u}_{\mathbf{j}}| + \frac{\lambda}{2} \sum_{\mathbf{j}} |\tilde{f}_{\mathbf{j}} - \tilde{u}_{\mathbf{j}}|^2 \right\} \end{aligned} \tag{4.7}$$

where $\tilde{f}_{\mathbf{j}}$ are the wavelet coefficients of the noisy image f and the restored image u is the wavelet reconstruction of $\tilde{u}$.

Note that because the summation in (4.7) is separable, it suffices to solve a sequence of scalar minimization problems $\min_{\tilde{u}_{\mathbf{j}}} \phi_{\tilde{f}_{\mathbf{j}}}(\tilde{u}_{\mathbf{j}})$, where

$$\phi_{\tilde{f}_{\mathbf{j}}}(\tilde{u}_{\mathbf{j}}) = |\tilde{u}_{\mathbf{j}}| + \frac{\lambda}{2}(\tilde{f}_{\mathbf{j}} - \tilde{u}_{\mathbf{j}})^2. \tag{4.8}$$

The minimizer of (4.7) is

$$\tilde{u}_{\mathbf{j}} = \begin{cases} \tilde{f}_{\mathbf{j}} - \dfrac{1}{\lambda}\mathrm{sign}(\tilde{f}_{\mathbf{j}}), & \text{if } |\tilde{f}_{\mathbf{j}}| > \dfrac{1}{\lambda}, \\ 0, & \text{if } |\tilde{f}_{\mathbf{j}}| \leq \dfrac{1}{\lambda}, \end{cases} \tag{4.9}$$

which is precisely the soft shrinkage algorithm (4.1) with threshold $\tau = \frac{1}{\lambda}$:

$$\tilde{u} = \mathcal{S}_{\frac{1}{\lambda}}(\tilde{f}).$$

This conclusion (of the relation between ROF and soft shrinkage) was observed by Chambolle et. al. in [CDL98].

Remark. *Denote* $F(\tilde{u}_{\boldsymbol{j}}) = |\tilde{u}_{\boldsymbol{j}}|$ *a scalar function, then* $J(\tilde{u}) = \sum_{\boldsymbol{j}} F(\tilde{u}_{\boldsymbol{j}})$ *and we can write the subgradient* $\partial J(\tilde{u}) = \{\partial F(\tilde{u}_{\boldsymbol{j}})\}$*, where*

$$\partial_{\tilde{u}_{\boldsymbol{j}}} F(\tilde{u}_{\boldsymbol{j}}) = \begin{cases} sign(\tilde{u}_{\boldsymbol{j}}), & if \;\; \tilde{u}_{\boldsymbol{j}} \neq 0, \\ [-1, 1], & if \;\; \tilde{u}_{\boldsymbol{j}} = 0. \end{cases} \tag{4.10}$$

The Euler-Lagrange equation of (4.7) *is*

$$\partial F(\tilde{u}_{\boldsymbol{j}}) + \lambda(\tilde{u}_{\boldsymbol{j}} - \tilde{f}_{\boldsymbol{j}}) \ni 0, \quad for\ all\ \boldsymbol{j},$$

Denote $\tilde{p}_{\boldsymbol{j}} = \lambda(\tilde{f}_{\boldsymbol{j}} - \tilde{u}_{\boldsymbol{j}}) \in \partial F(\tilde{u}_{\boldsymbol{j}})$, $\tilde{v}_{\boldsymbol{j}} = \frac{\tilde{p}_{\boldsymbol{j}}}{\lambda}$*, then* $\tilde{v}_{\boldsymbol{j}} = \tilde{f}_{\boldsymbol{j}} - \tilde{u}_{\boldsymbol{j}}$ *and we have a decomposition* $\tilde{f}_{\boldsymbol{j}} = \tilde{u}_{\boldsymbol{j}} + \tilde{v}_{\boldsymbol{j}}$*. From* (4.9) *we have*

$$\tilde{p}_{\boldsymbol{j}} = \begin{cases} sign(\tilde{f}_{\boldsymbol{j}})\,, & if\ |\tilde{f}_{\boldsymbol{j}}| > \dfrac{1}{\lambda}, \\ \lambda \tilde{f}_{\boldsymbol{j}}\,, & if\ |\tilde{f}_{\boldsymbol{j}}| \leq \dfrac{1}{\lambda}, \end{cases} \tag{4.11}$$

and

$$\tilde{v}_j = \begin{cases} \frac{1}{\lambda} sign(\tilde{f}_j)\,, & if\ |\tilde{f}_j| > \frac{1}{\lambda}, \\ \tilde{f}_j\,, & if\ |\tilde{f}_j| \leq \frac{1}{\lambda}, \end{cases} \tag{4.12}$$

for all j.

Note that although $\partial F(0) = [-1, 1]$ *is a multivalued set,* $\tilde{p}_j$ *defined above is unique.*

4.2 Iterative Regularization Applied to Wavelet Denoising

We start by reviewing the generalized Bregman distance discussed in Section 2.3.1. The generalized Bregman distance associated with $J(\tilde{u})$ in (4.5) can be defined as

$$D^{\tilde{p}}(\tilde{u}, \tilde{w}) = J(\tilde{u}) - J(\tilde{w}) - \langle \tilde{u} - \tilde{w}, \tilde{p} \rangle. \tag{4.13}$$

where $\tilde{p} \in \partial J(\tilde{w})$. Again we note that for $\tilde{w}_{\mathbf{j}} = 0$, $\partial F(\tilde{w}_{\mathbf{j}})$ is a multivalued set. However, as we shall see below, the proposed iterative regularization algorithm will automatically select a unique subgradient $\tilde{p}_{\mathbf{j}}$.

Following the same idea as in Section 2.3, we replace $J(\tilde{u})$ in (4.7) by the Bregman distance (4.13) and then obtain a sequence of minimization problems on $\tilde{u}_{\mathbf{j}}$ and the update of its dual variable $\tilde{p}_{\mathbf{j}}$ as follows

$$\tilde{u}^{(k)} = \underset{\tilde{u}}{\operatorname{argmin}} \left\{ D^{\tilde{p}^{(k-1)}}(\tilde{u}, \tilde{u}^{(k-1)}) + \frac{\lambda}{2} \|\tilde{f} - \tilde{u}\|_{L^2}^2 \right\}, \tag{4.14}$$

$$\tilde{p}^{(k)} = \tilde{p}^{(k-1)} + \lambda(\tilde{f} - \tilde{u}^{(k)}), \tag{4.15}$$

with $k \geq 1, \tilde{u}^{(0)} = 0, \tilde{p}^{(0)} = 0$. We shall show that such $\tilde{p}^{(k)} \in \partial J(\tilde{u}^{(k)})$.

If we denote $\tilde{v}^{(k)} = \frac{\tilde{p}^{(k)}}{\lambda}$, then $\tilde{v}^{(0)} = 0$, by plugging (4.13) into (4.14) and dropping the constant terms from the minimization, after some simplification we can rewrite (4.14) as

$$\tilde{u}^{(k)} = \underset{\tilde{u}}{\operatorname{argmin}} \left\{ J(\tilde{u}) + \frac{\lambda}{2} \left\| \left(\tilde{f} + \frac{\tilde{p}^{(k-1)}}{\lambda} \right) - \tilde{u} \right\|_{L^2}^2 \right\} \tag{4.16}$$

or

$$\tilde{u}^{(k)} = \underset{\tilde{u}}{\operatorname{argmin}} \left\{ J(\tilde{u}) + \frac{\lambda}{2} \left\| (\tilde{f} + \tilde{v}^{(k-1)}) - \tilde{u} \right\|_{L^2}^2 \right\}. \tag{4.17}$$

We have the following decomposition for all $k \geq 1$:

$$\tilde{f} + \tilde{v}^{(k-1)} = \tilde{u}^{(k)} + \tilde{v}^{(k)}. \tag{4.18}$$

We note that at the k^{th} iteration we simply replace the wavelet coefficient $\tilde{f}$ in the original minimization (4.7) by $\tilde{f} + \frac{\tilde{p}^{(k-1)}}{\lambda}$ or equivalently, $\tilde{f} + \tilde{v}^{(k-1)}$, and proceed to solve the same minimization procedure as for (4.7). The minimizer of (4.17) is then

$$\tilde{u}_{\mathbf{j}}^{(k)} = \begin{cases} (\tilde{f}_{\mathbf{j}} + \tilde{v}_{\mathbf{j}}^{(k-1)}) - \frac{1}{\lambda}\operatorname{sign}(\tilde{f}_{\mathbf{j}} + \tilde{v}_{\mathbf{j}}^{(k-1)})\,, & \text{if } |\tilde{f}_{\mathbf{j}} + \tilde{v}_{\mathbf{j}}^{(k-1)}| > \frac{1}{\lambda}, \\ 0\,, & \text{if } |\tilde{f}_{\mathbf{j}} + \tilde{v}_{\mathbf{j}}^{(k-1)}| \leq \frac{1}{\lambda}, \end{cases} \tag{4.19}$$

or simply,

$$\tilde{u}_{\mathbf{j}}^{(k)} = \mathcal{S}_{\frac{1}{\lambda}}(\tilde{f}_{\mathbf{j}} + \tilde{v}_{\mathbf{j}}^{(k-1)}), \tag{4.20}$$

where $k \geq 1$, $\tilde{v}_{\mathbf{j}}^{(0)} = 0$ and

$$\tilde{v}_{\mathbf{j}}^{(k)} = \tilde{f}_{\mathbf{j}} + \tilde{v}_{\mathbf{j}}^{(k-1)} - \tilde{u}_{\mathbf{j}}^{(k)}. \tag{4.21}$$

We have the following observations for the above $\tilde{u}_{\mathbf{j}}^{(k)}$ and $\tilde{v}_{\mathbf{j}}^{(k)}$:

Theorem 4.2.1. *For the solutions $\tilde{u}_j^{(k)}$ and $\tilde{v}_j^{(k)}$ defined in the updates* (4.19) *and* (4.21)*, $k \geq 1$, we have*

(1)

$$\tilde{v}_j^{(k)} = \begin{cases} \frac{1}{\lambda} sign(\tilde{f}_j)\,, & if\ |\tilde{f}_j| > \frac{1}{k\lambda}, \\ k\tilde{f}_j\,, & if\ |\tilde{f}_j| \leq \frac{1}{k\lambda}, \end{cases} \tag{4.22}$$

and $sign(\tilde{v}_j^{(k)}) \equiv sign(\tilde{f}_j)$.

(2)

$$\tilde{u}_j^{(k)} = \begin{cases} \tilde{f}_j\,, & if\ \ |\tilde{f}_j| > \frac{1}{(k-1)\lambda}, \\ k\tilde{f}_j - \frac{1}{\lambda} sign(\tilde{f}_j), & if\ \ \frac{1}{k\lambda} < |\tilde{f}_j| \leq \frac{1}{(k-1)\lambda}, \\ 0, & if\ \ |\tilde{f}_j| \leq \frac{1}{k\lambda}, \end{cases} \tag{4.23}$$

and $sign(\tilde{u}_j^{(k)}) = sign(\tilde{f}_j)$, if $\tilde{u}_j^{(k)} \neq 0$;

(3) for $\tilde{p}_j = \lambda\tilde{v}_j$, $\tilde{p}_j \in \partial_{\tilde{u}_j} F(\tilde{u}_j)$.

Proof. (1) Plugging (4.19) into (4.21) we have

$$\tilde{v}_{\mathbf{j}}^{(k)} = \begin{cases} \frac{1}{\lambda}\mathrm{sign}(\tilde{f}_{\mathbf{j}} + \tilde{v}_{\mathbf{j}}^{(k-1)})\,, & \text{if } |\tilde{f}_{\mathbf{j}} + \tilde{v}_{\mathbf{j}}^{(k-1)}| > \frac{1}{\lambda}, \\ \tilde{f}_{\mathbf{j}} + \tilde{v}_{\mathbf{j}}^{(k-1)}\,, & \text{if } |\tilde{f}_{\mathbf{j}} + \tilde{v}_{\mathbf{j}}^{(k-1)}| \leq \frac{1}{\lambda}, \end{cases} \tag{4.24}$$

for $k \geq 1$. Since $\tilde{v}_{\mathbf{j}}^{(0)} = 0$, we have $\mathrm{sign}(\tilde{v}_{\mathbf{j}}^{(1)}) = \mathrm{sign}(\tilde{f}_{\mathbf{j}})$. By induction, $\mathrm{sign}(\tilde{v}_{\mathbf{j}}^{(k)}) = \mathrm{sign}(\tilde{f}_{\mathbf{j}})$ for all $k \geq 1$. Next we also prove (4.22) by induction. For $k = 1$, we have (4.22) from (4.12). For $k \geq 2$,

(i) If $|\tilde{f}_{\mathbf{j}}| > \frac{1}{(k-1)\lambda}$, then $\tilde{v}_{\mathbf{j}}^{(k-1)} = \frac{1}{\lambda}\mathrm{sign}(\tilde{f}_{\mathbf{j}})$, and $|\tilde{f}_{\mathbf{j}} + \tilde{v}_{\mathbf{j}}^{(k-1)}| \geq |\tilde{f}_{\mathbf{j}}| > \frac{1}{(k-1)\lambda} > \frac{1}{k\lambda}$. From (4.24), $\tilde{v}_{\mathbf{j}}^{(k)} = \frac{1}{\lambda}\mathrm{sign}(\tilde{f}_{\mathbf{j}})$;

(ii) If $|\tilde{f}_{\mathbf{j}}| \leq \frac{1}{(k-1)\lambda}$, then $\tilde{v}_{\mathbf{j}}^{(k-1)} = (k-1)\tilde{f}_{\mathbf{j}}$ and $|\tilde{f}_{\mathbf{j}} + \tilde{v}_{\mathbf{j}}^{(k-1)}| = |k\tilde{f}_{\mathbf{j}}|$. From

(4.24), if $|k\tilde{f}_{\mathbf{j}}| > \frac{1}{\lambda}$, i.e., $|\tilde{f}_{\mathbf{j}}| > \frac{1}{k\lambda}$, then $\tilde{v}_{\mathbf{j}}^{(k)} = \frac{1}{\lambda}\text{sign}(\tilde{f}_{\mathbf{j}} + \tilde{v}_{\mathbf{j}}^{(k-1)}) = \frac{1}{\lambda}\text{sign}(\tilde{f}_{\mathbf{j}})$, otherwise, $|k\tilde{f}_{\mathbf{j}}| \leq \frac{1}{\lambda}$, $\tilde{v}_{\mathbf{j}}^{(k)} = \tilde{f}_{\mathbf{j}} + \tilde{v}_{\mathbf{j}}^{(k-1)} = k\tilde{f}_{\mathbf{j}}$.

This validates (4.22).

(2) Now we prove (4.23). From (4.21) we have $\tilde{u}_{\mathbf{j}}^{(k)} = \tilde{f}_{\mathbf{j}} + \tilde{v}_{\mathbf{j}}^{(k-1)} - \tilde{v}_{\mathbf{j}}^{(k)}$. Using (4.22), for $k > 1$ we have

(i) if $|\tilde{f}_{\mathbf{j}}| \leq \frac{1}{k\lambda} < \frac{1}{(k-1)\lambda}$, then

$$\tilde{v}_{\mathbf{j}}^{(k-1)} = (k-1)\tilde{f}_{\mathbf{j}},\ \tilde{v}_{\mathbf{j}}^{(k)} = k\tilde{f}_{\mathbf{j}}, \quad \Longrightarrow \quad \tilde{u}_{\mathbf{j}}^{(k)} = 0;$$

(ii) if $|\tilde{f}_{\mathbf{j}}| > \frac{1}{(k-1)\lambda} > \frac{1}{k\lambda}$, then

$$\tilde{v}_{\mathbf{j}}^{(k-1)} = \tilde{v}_{\mathbf{j}}^{(k)} = \frac{1}{\lambda}\text{sign}(\tilde{f}_{\mathbf{j}}), \quad \Longrightarrow \quad \tilde{u}_{\mathbf{j}}^{(k)} = \tilde{f}_{\mathbf{j}};$$

(iii) if $\frac{1}{k\lambda} < |\tilde{f}_{\mathbf{j}}| \leq \frac{1}{(k-1)\lambda}$, then

$$\tilde{v}_{\mathbf{j}}^{(k-1)} = (k-1)\tilde{f}_{\mathbf{j}},\ \tilde{v}_{\mathbf{j}}^{(k)} = \frac{1}{\lambda}\text{sign}(\tilde{f}_{\mathbf{j}}), \quad \Longrightarrow \quad \tilde{u}_{\mathbf{j}}^{(k)} = k\tilde{f}_{\mathbf{j}} - \frac{1}{\lambda}\text{sign}(\tilde{f}_{\mathbf{j}}),$$

and $\text{sign}(\tilde{u}_{\mathbf{j}}^{(k)}) = \text{sign}(\tilde{f}_{\mathbf{j}})\text{sign}(k|\tilde{f}_{\mathbf{j}}| - \frac{1}{\lambda}) = \text{sign}(\tilde{f}_{\mathbf{j}})$.

Note that for $k = 1$, we have $\frac{1}{(k-1)\lambda} = \infty$ and (4.23) reduces to (4.9). This validates (4.23).

(3) From (4.22) we have

$$\tilde{p}_{\mathbf{j}}^{(k)} = \begin{cases} \text{sign}(\tilde{f}_{\mathbf{j}}), & \text{if } |\tilde{f}_{\mathbf{j}}| > \dfrac{1}{k\lambda}, \\ k\lambda\tilde{f}_{\mathbf{j}}, & \text{if } |\tilde{f}_{\mathbf{j}}| \leq \dfrac{1}{k\lambda}, \end{cases} \tag{4.25}$$

For the first part, we have $\tilde{u}_{\mathbf{j}}^{(k)} \neq 0$ and $\partial J(\tilde{u}_{\mathbf{j}}^{(k)}) = \text{sign}(\tilde{u}_{\mathbf{j}}^{(k)}) = \tilde{p}_{\mathbf{j}}^{(k)}$. For the second part, we have $\tilde{u}_{\mathbf{j}}^{(k)} = 0$, and $|\tilde{p}_{\mathbf{j}}| \leq 1$, $\tilde{p}_{\mathbf{j}} \in \partial J(\tilde{u}_{\mathbf{j}}^{(k)})$. This also shows that the dual variable $\tilde{p}^{(k)}$ updated via (4.15) is automatically a subgradient of $J(\tilde{u}^{(k)})$.

□

We note that (4.23) gives firm shrinkage (4.3) with thresholds $\tau^{(k)} = \frac{1}{k\lambda}$ and $\tau^{(k-1)} = \frac{1}{(k-1)\lambda}$. One can also see that these thresholds are monotonically decreasing with respect to the iterates k. Therefore, the iterative soft shrinkage provides a multiscale wavelet denoising sequence, in the sense that bigger coefficients in $\tilde{u}_{\mathbf{j}}$ are saved earlier than smaller ones.

Now we need a stopping criterion for the iterations. We first observe that the distance between f and $u^{(k)}$, which equals to the distance between $\tilde{f}$ and $\tilde{u}^{(k)}$, is monotonically decreasing with respect to k. Then we can use the same stopping criterion as was used for iterated total variation based models: we stop the iteration (4.14) and (4.15) at the last $k = \bar{k}$ where

$$\|\tilde{f} - \tilde{u}^{(k)}\|_{L^2} \geq \sigma,$$

where $\sigma = \|f - g\|_{L^2} = \|\tilde{f} - \tilde{g}\|_{L^2}$ is the variance of the noise $f - g$, g is used to denote the original clean image. Note that this stopping criterion corresponds to the commonly used L^2 constraint in denoising problems. In general g is unknown, however, as we discussed in previous chapter, in typical imaging situations, an estimate for the noise variance is known, which yields a bound of the form $\|f - g\|_{L^2} \leq \sigma$.

4.2.1 Bregman distance

We are interested in the Bregman distance between the original clean image g and the restored image u. In the wavelet space, we now turn to compute the Bregman distance (4.13) between the wavelet coefficients $\tilde{g}$ and $\tilde{u}^{(k)}$:

$$
\begin{aligned}
D^{\tilde{p}^{(k)}}(\tilde{g}, \tilde{u}^{(k)}) &= J(\tilde{g}) - J(\tilde{u}^{(k)}) - \langle \tilde{g} - \tilde{u}^{(k)}, \tilde{p}^{(k)} \rangle \\
&= J(\tilde{g}) - \langle \tilde{g}, \tilde{p}^{(k)} \rangle \\
&= \sum_{\mathbf{j}} |\tilde{g}_{\mathbf{j}}| - \sum_{\mathbf{j}} \tilde{p}_{\mathbf{j}}^{(k)} \tilde{g}_{\mathbf{j}} \\
&= \sum_{\mathbf{j}:|\tilde{f}_{\mathbf{j}}|>1/k\lambda} (|\tilde{g}_{\mathbf{j}}| - \mathrm{sign}(\tilde{f}_{\mathbf{j}})\tilde{g}_{\mathbf{j}}) + \sum_{\mathbf{j}:|\tilde{f}_{\mathbf{j}}|\leq 1/k\lambda} (|\tilde{g}_{\mathbf{j}}| - k\lambda \tilde{f}_{\mathbf{j}} \tilde{g}_{\mathbf{j}}) \qquad (4.26) \\
&\geq 0.
\end{aligned}
$$

And we also have

$$
\begin{aligned}
& D^{\tilde{p}^{(k)}}(\tilde{g}, \tilde{u}^{(k)}) - D^{\tilde{p}^{(k-1)}}(\tilde{g}, \tilde{u}^{(k-1)}) \\
&= -\langle \tilde{g}, \tilde{p}^{(k)} - \tilde{p}^{(k-1)} \rangle \\
&\leq -\langle \tilde{g} - \tilde{u}^{(k)}, \tilde{p}^{(k)} - \tilde{p}^{(k-1)} \rangle \\
&= -\langle \tilde{g} - \tilde{u}^{(k)}, \lambda(f - \tilde{u}^{(k)}) \rangle \\
&\leq \lambda(-\frac{1}{2}\|\tilde{f} - \tilde{u}^{(k)}\|_{L^2}^2 + \frac{1}{2}\|\tilde{f} - \tilde{g}\|_{L^2}^2) \\
&< 0
\end{aligned}
$$

as long as $\|\tilde{f} - \tilde{u}^{(k)}\|_{L^2} > \|\tilde{f} - \tilde{g}\|_{L^2} = \sigma$.

Therefore, the Bregman distance (4.13) monotonically decrease for k less than $\bar{k}$, which is the last iterate such that $\|\tilde{f} - \tilde{u}^{(k)}\|_{L^2} \geq 0$.

4.2.2 Limiting case

If we reinterpret $\lambda = \Delta t$ as a timestep and $k\lambda = t^k$, then (4.23) becomes

$$\tilde{u}_{\mathbf{j}}(t^k) = \begin{cases} \tilde{f}_{\mathbf{j}}\,, & \text{if } |\tilde{f}_{\mathbf{j}}| \geq \dfrac{1}{t^k - \Delta t}, \\ k(\tilde{f}_{\mathbf{j}} - \dfrac{1}{t^k}\mathrm{sign}(\tilde{f}_{\mathbf{j}}))\,, & \text{if } \dfrac{1}{t^k} \leq |\tilde{f}_{\mathbf{j}}| < \dfrac{1}{t^k - \Delta t}, \\ 0\,, & \text{if } |\tilde{f}_{\mathbf{j}}| < \dfrac{1}{t^k}. \end{cases}$$

Let $\Delta t \searrow 0$, then for $k \gg 1$, $t - \Delta t \to t$. Dropping k we have the following solution

$$\tilde{u}_{\mathbf{j}}(t) = \begin{cases} \tilde{f}_{\mathbf{j}}\,, & |\tilde{f}_{\mathbf{j}}| \geq \dfrac{1}{t}, \\ 0\,, & |\tilde{f}_{\mathbf{j}}| < \dfrac{1}{t}, \end{cases} \tag{4.27}$$

which turns out to be hard shrinkage (4.2) with threshold $\tau = \frac{1}{t}$:

$$\tilde{u} = \mathcal{H}_{\frac{1}{t}}(\tilde{f}).$$

The Bregman distance $D(\tilde{g}, \tilde{u}(t))$ is same as the one stated in (4.26), with $k\lambda$ replaced by t, i.e.,

$$D(\tilde{g}, \tilde{u}(t)) = \sum_{\mathbf{j}:|\tilde{f}_{\mathbf{j}}|>1/t} (|\tilde{g}_{\mathbf{j}}| - \mathrm{sign}(\tilde{f}_{\mathbf{j}})\tilde{g}_{\mathbf{j}}) - \sum_{\mathbf{j}:|\tilde{f}_{\mathbf{j}}|\leq 1/t} (|\tilde{g}_{\mathbf{j}}| - t\tilde{f}_{\mathbf{j}}\tilde{g}_{\mathbf{j}}) \tag{4.28}$$

4.3 Regularized Wavelet Denoising and Inverse Scale Space

In this section we will generalize the above iterative regularization procedure to a time-continuous inverse scale space. First we need to borrow the usual regularization technique from the TV-based imaging community: we approximate

$F(\tilde{u}_{\mathbf{j}}) = |\tilde{u}_{\mathbf{j}}|$ as

$$F_\epsilon(\tilde{u}_{\mathbf{j}}) = \sqrt{\tilde{u}_{\mathbf{j}}^2 + \epsilon}, \tag{4.29}$$

where $\epsilon > 0$ is a small constant (and independent of $\mathbf{j}$). Note that

$$\tilde{p}_{\mathbf{j}} = \partial F_\epsilon(\tilde{u}_{\mathbf{j}}) = \frac{\tilde{u}_{\mathbf{j}}}{\sqrt{\tilde{u}_{\mathbf{j}}^2 + \epsilon}} \tag{4.30}$$

is well-defined and unique everywhere now, and we can invert $\tilde{u}_{\mathbf{j}}$ from $\tilde{p}_{\mathbf{j}}$.

If we replace $F(\tilde{u}_{\mathbf{j}})$ with $F_\epsilon(\tilde{u}_{\mathbf{j}})$ in (4.7), we have

$$\tilde{u}_{\mathbf{j}} = \underset{\tilde{u}_{\mathbf{j}}}{\operatorname{argmin}} \left\{ \sqrt{\tilde{u}_{\mathbf{j}}^2 + \epsilon} + \frac{\lambda}{2}(\tilde{f}_{\mathbf{j}} - \tilde{u}_{\mathbf{j}})^2 \right\}, \quad \forall j. \tag{4.31}$$

The corresponding Euler-Lagrange equation now is

$$\frac{\tilde{u}_{\mathbf{j}}}{\sqrt{\tilde{u}_{\mathbf{j}}^2 + \epsilon}} + \lambda(\tilde{u}_{\mathbf{j}} - \tilde{f}_{\mathbf{j}}) = 0. \tag{4.32}$$

This is a nonlinear equation for $\tilde{u}_{\mathbf{j}}$ which can be solved numerically, e.g., by a simple fixed point method. A slight extra computational cost comes with it as compared to (4.23).

4.3.1 Inverse scale space

We start from the Bregman decomposition (4.18). Using $\tilde{v}^{(k)} = \frac{\tilde{p}^{(k)}}{\lambda}$, we have for each $\mathbf{j}$,

$$\tilde{p}_{\mathbf{j}}^{(k)} - \tilde{p}_{\mathbf{j}}^{(k-1)} = \lambda(\tilde{f}_{\mathbf{j}} - \tilde{u}_{\mathbf{j}}^{(k)}), \quad k \geq 1 \tag{4.33}$$

$$\tilde{u}_{\mathbf{j}}^{(0)} = \tilde{p}_{\mathbf{j}}^{(0)} = 0. \tag{4.34}$$

Let $\lambda = \Delta t$, $k\Delta t \to t$, the equation becomes

$$\frac{d\tilde{p}_{\mathbf{j}}}{dt} = \tilde{f}_{\mathbf{j}} - \tilde{u}_{\mathbf{j}}, \quad \tilde{u}_{\mathbf{j}}(0) = 0. \tag{4.35}$$

Since $\dfrac{d\tilde{p}_{\mathbf{j}}}{d\tilde{u}_{\mathbf{j}}} = \dfrac{\epsilon}{(\tilde{u}_{\mathbf{j}}^2 + \epsilon)^{3/2}}$, we have an inverse scale space flow for each $\tilde{u}_{\mathbf{j}}$ as follows

$$\frac{d\tilde{u}_{\mathbf{j}}}{dt} = \frac{(\tilde{u}_{\mathbf{j}}^2 + \epsilon)^{3/2}}{\epsilon}(\tilde{f}_{\mathbf{j}} - \tilde{u}_{\mathbf{j}}), \quad \tilde{u}_{\mathbf{j}}(0) = 0. \tag{4.36}$$

4.3.2 Convergence analysis

We now study the behavior of the above regularized inverse scale space model (4.36). First,

$$\begin{aligned}
\frac{d}{dt}\|\tilde{f} - \tilde{u}(t)\|_{L^2}^2 &= \frac{d}{dt}\sum(\tilde{f}_{\mathbf{j}} - \tilde{u}_{\mathbf{j}})^2 \\
&= -2\sum_j(\tilde{f}_{\mathbf{j}} - \tilde{u}_{\mathbf{j}})^2\left(\frac{\epsilon}{(\tilde{u}_{\mathbf{j}}^2 + \epsilon)^{3/2}}\right)^{-1} \\
&\leq -2\epsilon^{1/2}\sum_j(\tilde{f}_{\mathbf{j}} - \tilde{u}_{\mathbf{j}})^2 \\
&\leq -2\epsilon^{1/2}\|\tilde{f} - \tilde{u}(t)\|_{L^2}^2.
\end{aligned}$$

From Granwall's inequality we have

$$\|\tilde{f} - \tilde{u}(t)\|_{L^2}^2 \leq e^{-\epsilon^{1/2}(t-s)}\|\tilde{f} - \tilde{u}(s)\|_{L^2}^2, \quad t \geq s.$$

If $f \in L^2$, let $s = 0$, we have $t \to \infty$,

$$\|\tilde{f} - \tilde{u}(t)\|_{L^2}^2 \leq e^{-\epsilon^{1/2}(t)}\|\tilde{f}\|_{L^2}^2 \searrow 0, \text{ as } t \nearrow \infty.$$

Therefore, $\tilde{u}(t) \to \tilde{f}$ in L^2 as $t \to \infty$, and as a consequence the reconstructed result $u(t) \to f$ as $t \to \infty$.

Second, the Bregman distance between $\tilde{g}$ and $\tilde{u}$ is $\sum_{\mathbf{j}} d_{\mathbf{j}}(\tilde{g}_{\mathbf{j}}, \tilde{u}_{\mathbf{j}})$, where

$$\begin{aligned} 0 \le d_{\mathbf{j}}(\tilde{g}_{\mathbf{j}}, \tilde{u}_{\mathbf{j}}) &= F_\epsilon(\tilde{g}_{\mathbf{j}}) - F_\epsilon(\tilde{u}_{\mathbf{j}}) - (\tilde{g}_{\mathbf{j}} - \tilde{u}_{\mathbf{j}})\frac{\partial F_\epsilon}{\partial \tilde{u}_{\mathbf{j}}}(\tilde{u}_{\mathbf{j}}) \\ &= \sqrt{\tilde{g}_{\mathbf{j}}^2 + \epsilon} - \frac{\epsilon + \tilde{g}_{\mathbf{j}}\tilde{u}_{\mathbf{j}}}{\sqrt{\tilde{u}_{\mathbf{j}}^2 + \epsilon}} \\ &\to |\tilde{g}_{\mathbf{j}}| - \tilde{g}_{\mathbf{j}} \operatorname{sign}(\tilde{u}_{\mathbf{j}}) \quad \text{as } \epsilon \searrow 0. \end{aligned}$$

For any $g \in B_1^1(L^1)$,

$$\begin{aligned} \frac{d}{dt}D(\tilde{g}, \tilde{u}) &= \sum_{\mathbf{j}} (\tilde{g}_{\mathbf{j}} - \tilde{u}_{\mathbf{j}})\frac{d\tilde{p}_{\mathbf{j}}}{dt} = -\sum_{\mathbf{j}} (g - \tilde{u}_{\mathbf{j}})(f - \tilde{u}_{\mathbf{j}}) \\ &\le -\sum_{\mathbf{j}} \frac{(\tilde{f}_{\mathbf{j}} - \tilde{u}_{\mathbf{j}})^2}{2} + \sum_{\mathbf{j}} \frac{(\tilde{f}_{\mathbf{j}} - \tilde{g}_{\mathbf{j}})^2}{2} \\ &< 0, \end{aligned}$$

as long as $\|\tilde{f} - \tilde{u}(t)\|_{L^2} > \|\tilde{f} - \tilde{g}\|_{L^2}$.

We may rewrite

$$d_{\mathbf{j}}(\tilde{g}_{\mathbf{j}}, \tilde{u}_{\mathbf{j}}) = \frac{\epsilon(\tilde{u}_{\mathbf{j}} - \tilde{g}_{\mathbf{j}})^2}{\sqrt{\tilde{u}_{\mathbf{j}}^2 + \epsilon}(\sqrt{\tilde{u}_{\mathbf{j}}^2 + \epsilon}\sqrt{\tilde{g}_{\mathbf{j}}^2 + \epsilon} + \epsilon + \tilde{g}_{\mathbf{j}}\tilde{u}_{\mathbf{j}})} \tag{4.37}$$

The factor ϵ can be removed and we may rewrite, for any $\epsilon > 0$

$$\frac{d}{dt}\frac{D(\tilde{g}, \tilde{u})}{\epsilon} < 0, \quad \text{as long as } \|\tilde{f} - \tilde{u}\|_{L^2} > \|\tilde{f} - \tilde{g}\|_{L^2} = \sigma.$$

This is an interesting estimate which holds uniformly in ϵ. We also have a stopping criterion which is similar to the one for iterative refinement: we can stop the evolution (4.36) at the last $t = \bar{t}$ where $\|\tilde{f} - \tilde{u}(t)\|_{L^2} \geq \sigma$.

4.4 Numerical Examples

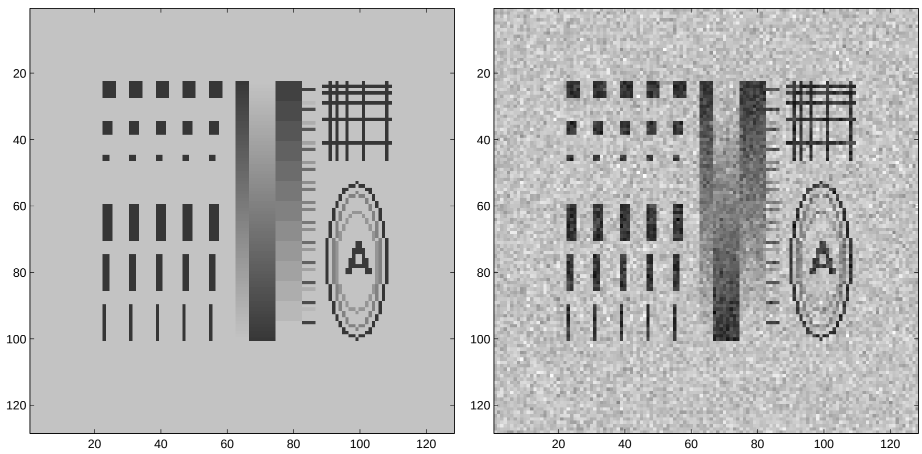


Figure 4.1: shape image, 128×128. left: original image; right: noisy image, $\sigma = 30$, $SNR = 7.29$.

In this section we present two numerical examples of the wavelet denoising with soft shrinkage, hard shrinkage, the new iterative regularization method (W-IRM, which is equivalent with a sequence of firm shrinkage) and the inverse scale space (W-ISS) flow we mentioned above in this paper. We add Gaussian i.i.d. noise to the original clean image g and get a noisy image f to use for experiments. For different thresholds and parameters, there are two ways to define 'optimal' results numerically: (i) the SNR of the restored image u is biggest among all; (ii) $\|f - u\|_{L^2} \approx \sigma$. In general we may have an estimate of σ but no information about g, therefore we can only use the second criterion in those cases. Moreover, as indicated in previous sections, (ii) is also our stopping criterion for the iterative method and inverse scale space flow. To compare results in this numerical section we choose thresholds τ for soft shrinkage and hard shrinkage such that their results satisfy $\|f-u\|_{L^2}=\sigma$ and we use stopping criterion (ii) for all our new methods.

Figure 4.1 shows the original image g, which is composed of different shapes and scales, and the noisy image f. $\sigma=\|f-g\|_{L^2}=30$, the signal-to-noise-ratio (SNR) is 7.29.

We choose the Haar basis and level 3 for wavelet decomposition in this example. In Figure 4.2, the first row shows the results u from soft shrinkage (threshold $\tau = 49$) and hard shrinkage ($\tau = 101$), with their corresponding $SNR = 12.03$ and 13.04 respectively; the second row shows the results from iterative regularization (W-IRM, $\lambda = 0.001, \bar{k} = 11$) and inverse scale space (W-ISS, $dt = 0.001, \bar{t} = 0.014, \epsilon = 0.01$), with their corresponding $SNR = 13.56$ and 13.45 respectively. We can see that these two new results are close to the result of hard shrinkage. Their $SNRs$ are slightly higher than that of hard shrinkage and much higher than that of soft shrinkage.

In Figure 4.2 there are some artifacts in the results. This is a common defect of wavelet imaging. We point out here that some techniques such as translation invariant cycle-spinning (cf. [DC95]) can be easily incorporated into our new methods. Moreover, in the W-ISS method proposed above, we introduced a regularized parameter ϵ, which can also be used to decrease the artifacts if its value is taken to be big. In Figure 4.3 we show a result u of W-ISS with a bigger $\epsilon = 10$, which has much fewer artifacts than the previous results. The corresponding $SNR = 12.84$ is higher than that of soft shrinkage. Furthermore, we also plotted the residual part $v = f - u$ of this result (to enhance the visibility, we plotted $v + 128$ here). We can see that it contains very little visible signal, which is similar to the residual of hard shrinkage. In soft shrinkage, we removed

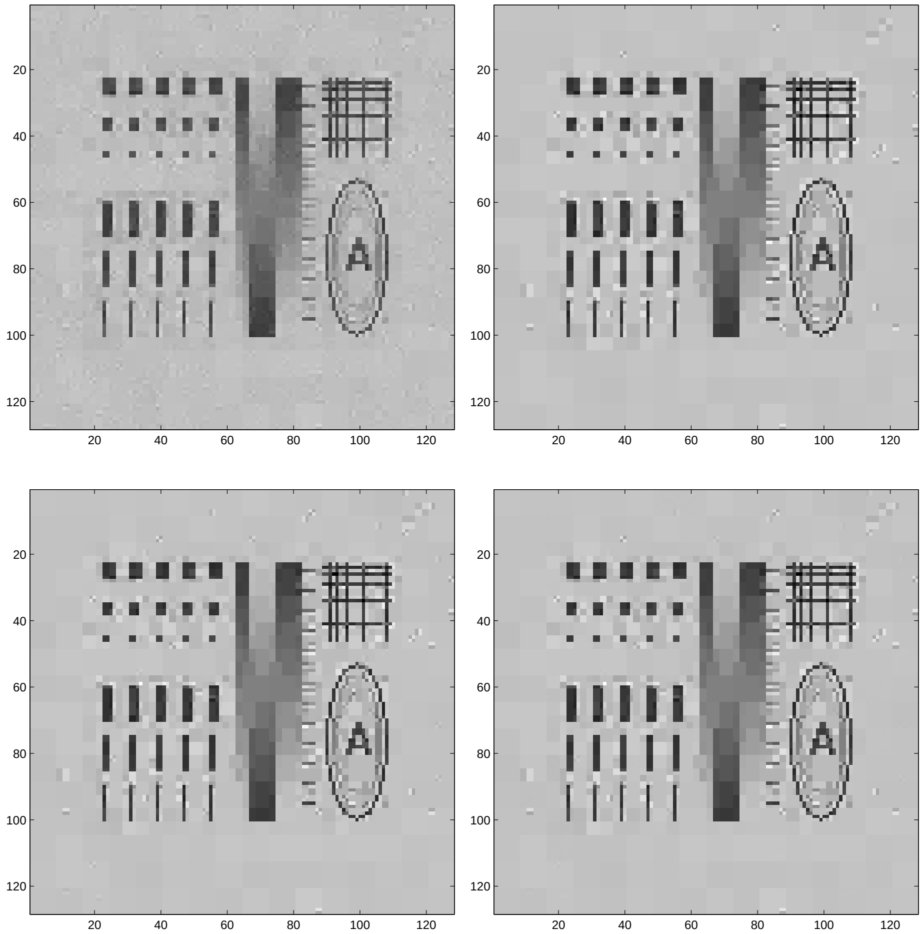


Figure 4.2: First row: denoised results from soft shrinkage (left, $SNR = 12.03$) and hard shrinkage (right, $SNR = 13.04$); Second row: denoised results from iterative regularization (4.23) (left, $SNR = 13.56$) and inverse scale space (4.36) (right, $SNR = 13.45$). All $\|f - u\|_{L^2} \approx \sigma = 30$.

some signal along with the residual.

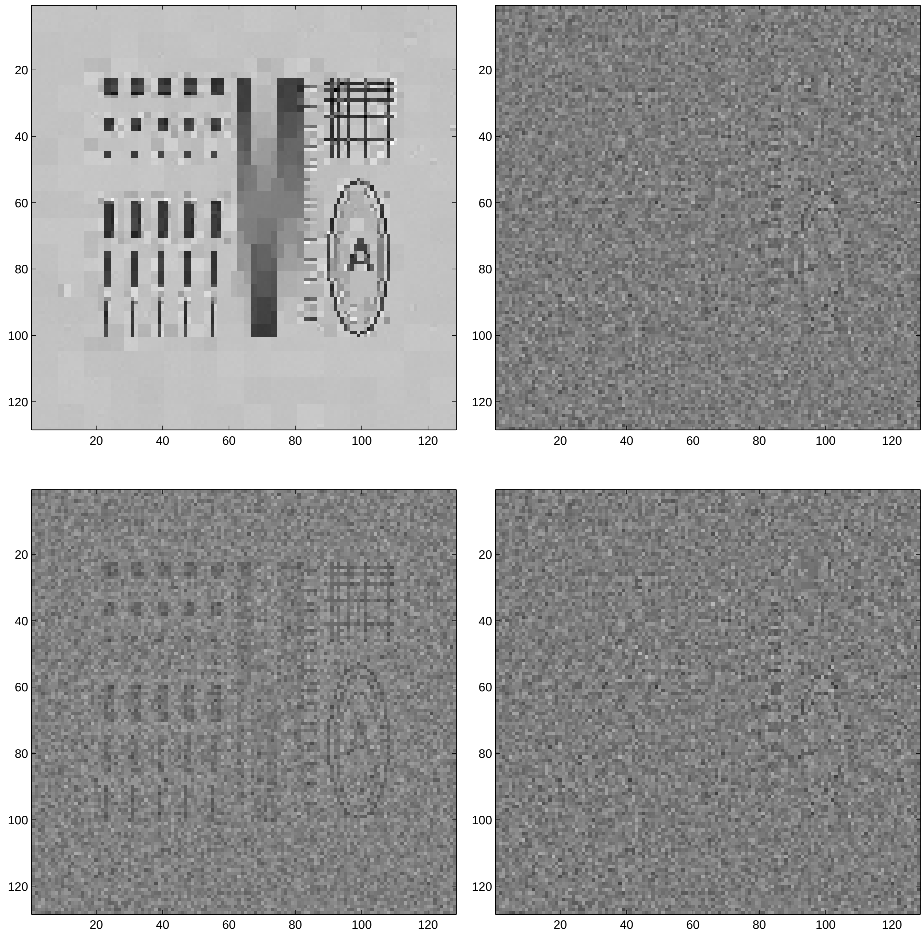


Figure 4.3: First row: denoised result u from inverse scale space with $\epsilon = 10$ (left, $SNR = 12.84$) and corresponding residual $v = f - u$ (+128, right); Second row: residuals $v + 128$ of soft shrinkage (left) and hard shrinkage (right) in Figure 4.2.

In Figure 4.4 we show the result from total variation based relaxed inverse scale space discussed in previous chapter. The denoised result is better than that from wavelet methods, with no artifacts and higher SNR, but the computational cost is much more expensive due to the evolution of nonlinear partial differential equations.

In the second example we denoised a magnetic resonance (MR) image. Figure

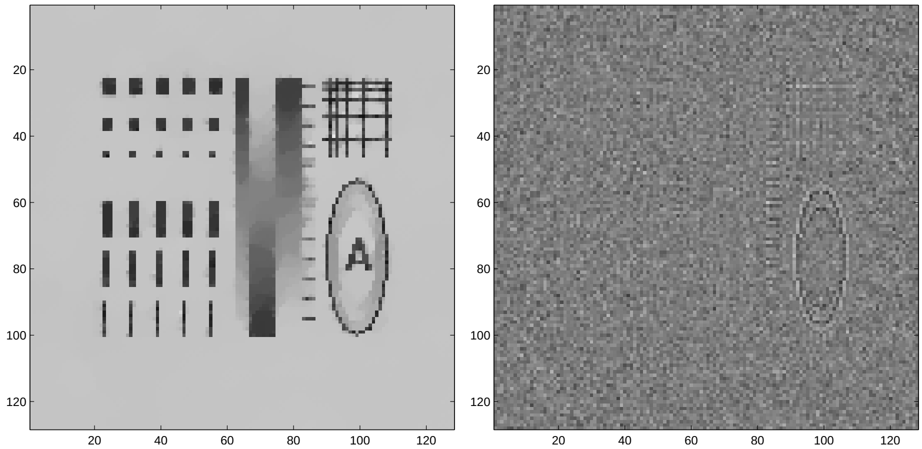


Figure 4.4: Result from TV relaxed ISS. left: denoised u ($SNR = 14.96$, $\|f-u\|_{L^2} = \sigma = 30$); right: residual $v+128$ ($v = f-u$).

4.5 shows the original image g and noisy image f. $\sigma = 30$ and SNR of f is 4.43. We use db3 basis and level 3 for wavelet decomposition in this example. Figure 4.6 shows the results: the first row shows u and v from soft shrinkage ($\tau = 57, SNR = 11.72$); the second row shows u from hard shrinkage ($\tau = 83, SNR = 11.01$) and the W-IRM method ($\lambda = 0.0008, \bar{k} = 16, SNR = 11.01$); the third row shows u from the W-ISS method ($dt = 0.001, \bar{t} = 0.012, \epsilon = 10, SNR = 11.94$). In this example we can see that: compared with soft shrinkage, although the $SNRs$ of hard shrinkage and W-IRM are lower, using a relative large $\epsilon = 10$ in W-ISS we obtained a result with fewer artifacts, higher SNR, and much less visible signal in the residual.

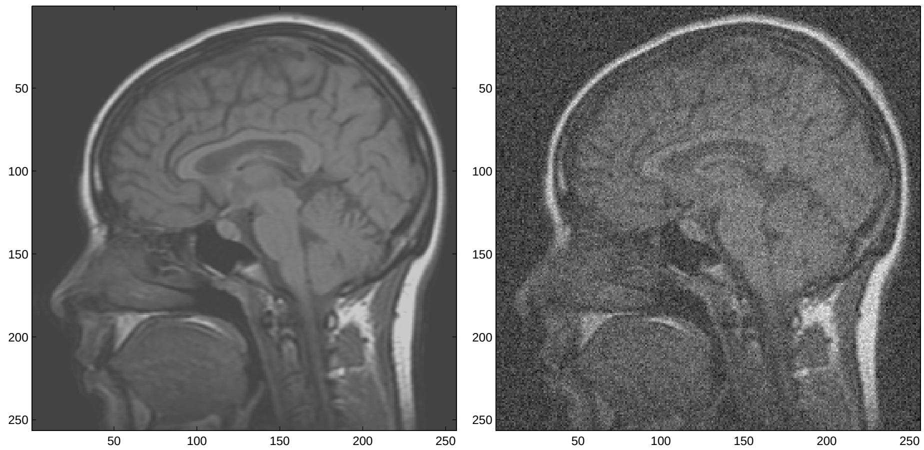


Figure 4.5: MR image, 256×256. left: original image; right: noisy image, $\sigma = 30$, $SNR = 4.43$,

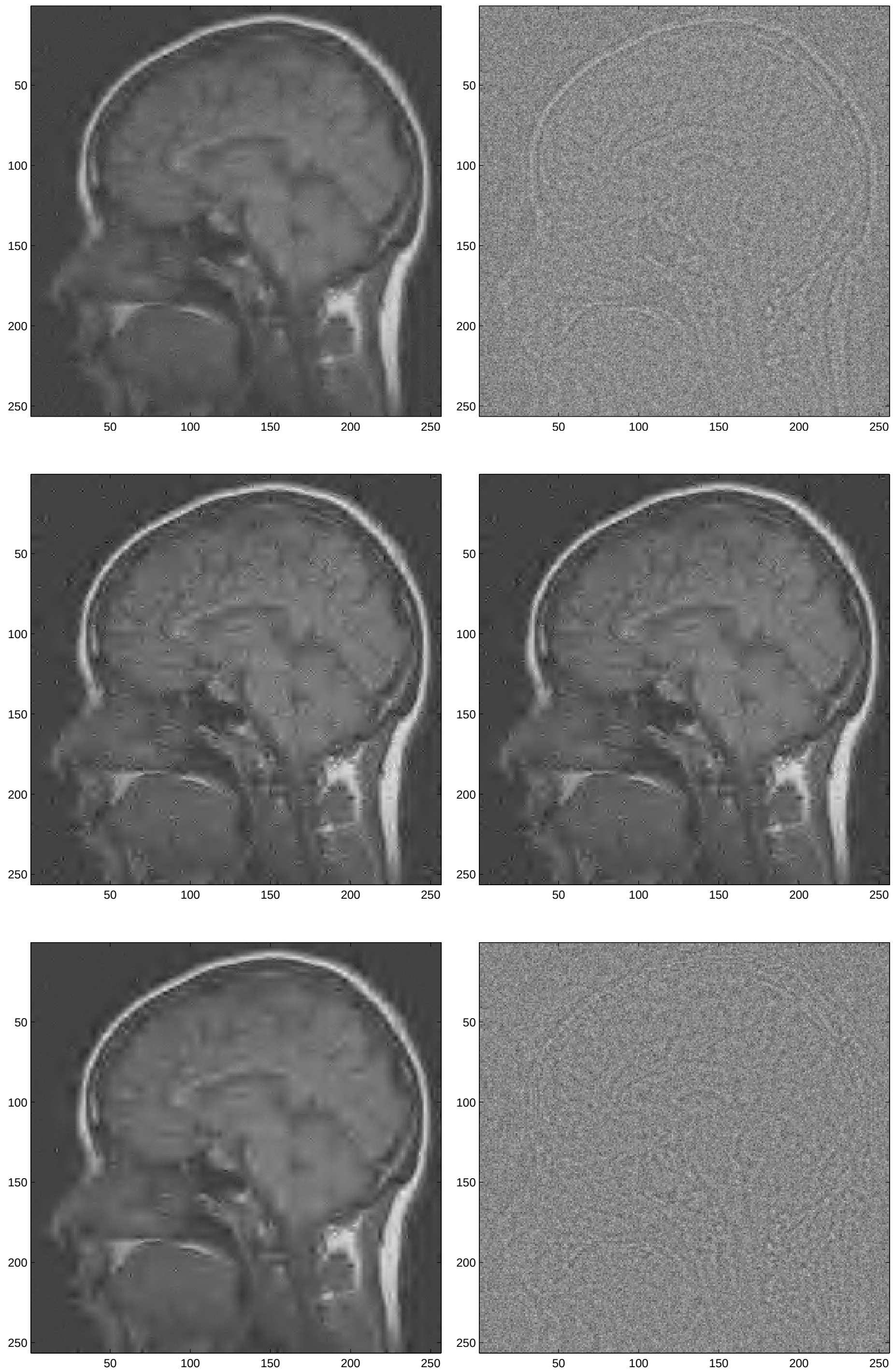


Figure 4.6: First row: denoised result from soft shrinkage (left, $SNR = 11.72$) and corresponding residual $v = f - u$ (+128,right); Second row: denoised results from hard shrinkage (left, $SNR = 11.01$) and iterative regularization (4.23) (right, $SNR = 11.01$); Third row: denoised result from inverse scale space (4.36) (left, $\epsilon = 10, SNR = 11.94$) and corresponding residual $v + 128$ (right). All $\|f - u\|_{L^2} \approx \sigma = 30$.

4.5 Discussion and Conclusion

We have presented two alternatives to soft shrinkage. These involve the iterative regularization (W-IRM) using generalized Bregman distances and the inverse scale space (W-ISS) ideas borrowed from total variation based image restoration. The solution of W-IRM turns out to be a firm shrinkage with the thresholds determined by the scale parameter λ and iterates k. It appears that the new methods, especially W-ISS, perform better than soft shrinkage from the SNR point of view and result in less loss of signal into the residual. All these methods are fast and easy to implement. Since the summation part in the minimization functions is separable, our methods can be easily combined with some other techniques of wavelet denoising, such as shrinkage on detail coefficients only and keep the scaling coefficients unchanged, and/or cycle spinning. The open questions include the studies of the scale parameter λ in W-IRM and the regularized parameter ϵ in W-ISS.

Acknowledgments

We thank Gabriele Steidl (University of Mannheim) for introducing firm shrinkage to us.

References

[AAB05] Jean-Francois Aujol, Gilles Aubert, Laure Blanc-Féraud, 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." *International Journal of Computer Vision*, **63**:85–104, 2005.

[AE84] Jean-Pierre Aubin and Ivar Ekeland. *Applied Nonlinear Analysis*. John Wiley & Sons, New York, 1984.

[AGC06] Jean-Francois Aujol, Guy Gilboa, Tony Chan, and Stanley Osher. "Structure-Texture Image Decomposition - Modeling, Algorithms, and Parameter Selection." *International Journal of Computer Vision*, **67**(1):111–136, 2006.

[AGL93] Luis Alvarez, Frederic Guichard, Pierre-Louis Lions, and Jean-Michel Morel. "Axioms and Fundamental Equations of Image Processing." *Arch. Rat. Mech. and Anal.*, **123**(3):199–257, 1993.

[ALM92] Luis Alvarez, Pierre-Louis Lions, and Jean-Michel Morel. "Image Selective Smoothing and Edge Detection by Nonlinear Diffusion (II)." *SIAM J. Numer. Anal.*, **29**(3):845–866, 1992.

[AV94] Robert Acar and Curt Vogel. "Analysis of Bounded Variation Penalty Methods for Ill-Posed Problems." *Inverse Problems*, **10**:1217–1229, 1994.

[BCM05] Antoni Buades, Bartomeu Coll, and Jean-Michel Morel. "A Review of Image Denoising Algorithms, With a New One." *Multiscale Model. Simul.*, **4**(2):490–530, 2005.

[BFO06] Martin Burger, Klaus Frick, Stanley Osher, and Otmar Scherzer. "Inverse Total Variation Flow." *UCLA-CAM-Report 06-24, Los Angeles, CA, USA*, 2006.

[BGO06] Martin Burger, Guy Gilboa, Stanley Osher, and Jinjun 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.

[BOX05] Martin Burger, Stanley Osher, Jinjun Xu, and Guy Gilboa. "Nonlinear Inverse Scale Space Methods for Image Restoration." *Lecture Notes in Computer Science*, **3752**:25–36, 2005.

[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.

[BSC00] Marcelo Bertalmio, Guillermo Sapiro, Vicent Caselles, and Coloma Ballester. "Image Inpainting." In *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques, ACM*, pp. 417–424, New York, 2000.

[BVS03] Marcelo Bertalmio, Luminita Vese, Guillermo Sapiro, and Stanley Osher. "Simultaneous Structure and Texture Image Inpainting." *IEEE Trans. Image Proc.*, **12**(8):882–889, 2003.

[Car02] Jamylle Laurice Carter. *Dual Methods for Total Variation-Based Image Restoration*. PhD thesis, UCLA, 2002.

[CDD03] Albert Cohen, Wolfgang Dahmen, Ingrid Daubechies, and Ronald DeVore. "Harmonic Analysis of the Space BV." *Rev. Mat. Iberoamericana*, **19**:235–263, 2003.

[CDL98] Antonin Chambolle, Ronald DeVore, Nam-Yong Lee, and Bradley Lucier. "Nonlinear Wavelet Image Processing: Variational Problems, Compression, and Noise Removal Through Wavelet Shrinkage." *IEEE Trans. Image Proc.*, **7**(3):319–335, 1998.

[CDP99] Albert Cohen, Ronald DeVore, Pencho Petrushev, and Hong Xu. "Nonlinear Approximation and the Space $BV(\mathbb{R}^2)$." *American J. of Math.*, **121**(3):587–628, 1999.

[CDV93] Albert Cohen, Ingrid Daubechies, and Pierre Vial. "Wavelets on the Interval and Fast Wavelet Transforms." *Appl. Comput. Harmon. Anal.*, **1**:54–81, 1993.

[CE05] Tony Chan and Selim Esedoḡlu. "Aspects of Total Variation Regularized L^1 Function Approximation." *SIAM J. Appl. Math.*, **65**(5):1817–1837, 2005.

[CEP05a] Tony Chan, Selim Esedoḡlu, and Frederick Park. "A Fourth Order Dual Method for Staircase Reduction in Texture Extraction and Image Restoration Problems." *UCLA-CAM-Report 05-28, Los Angeles, CA, USA*, 2005.

[CEP05b] Tony Chan, Selim Esedoḡlu, and Frederick Park. "Image Decomposition Combining Staircase Reduction and Texture Extraction." *UCLA-CAM-Report 05-18, Los Angeles, CA, USA*, 2005.

[CEP05c] Tony Chan, Selim Esedoḡlu, Frederick Park, and Andy Yip. *Mathematical Models in Computer Vision: The Handbook*, chapter Total Variation Image Restoration: Overview and Recent Developments, pp. 17–32. Springer, 2005.

[Cet89] Enis Cetin. "Reconstruction of Signals from Fourier Transform Samples." *Signal Process.*, **16**:129–148, 1989.

[Cet91] Enis Cetin. "An Iterative Algorithm for Signal Reconstruction from Bispectrum." *IEEE Trans. Signal Proc.*, **39**:2621–2628, 1991.

[CGM00] Tony Chan, Gene Golub, and Pep Mulet. "A Nonlinear Primal-Dual Method for Total Variation Based Image Restoration." *SIAM J. Sci. Comput.*, **22**:503–516, 2000.

[Cha04] Antonin Chambolle. "An Algorithm for Total Variation Regularization and Denoising." *Journal of Mathematical Imaging and Vision*, **20**:89–97, 2004.

[CK97] Guy Chavent and Karl Kunisch. "Regularization of Linear Least Squares Problems by Total Bounded Variation." *ESAIM Control Optim. Calc. Var.*, **2**:359–376, 1997.

[CKP99] Eduardo Casas, Karl Kunisch, and Cecilia Pola. "Regularization by Functions of Bounded Variation and Applications to Image Enhancement." *Appl. Math. Optim.*, **40**:229–257, 1999.

[CL97] Antonin Chambolle and Pierre-Louis Lions. "Image Recovery Via Total Variational Minimization and Related Problems." *Numer. Math.*, **76**:167–188, 1997.

[CMM00] Tony Chan, Antonio Marquina, and Pep Mulet. "High-Order Total Variation-Based Image Restoration." *SIAM J. Sci. Comput.*, **22**:503–516, 2000.

[CS05] Tony Chan and Jianhong Shen. *Image Processing and Analysis.* SIAM, Philadelphia, 2005.

[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 Chan and Chiu-Kwong Wong. "Total Variation Blind Deconvolution." *IEEE Trans. Image Proc.*, **7**:370–375, 1998.

[Dau92] Ingrid Daubechies. *Ten Lectures on Wavelets.* SIAM, Philadelphia, 1992.

[DC95] David Donoho and Ronald Coifman. "Translation-Invariant De-Noising." In A. Antoniadis and G. Oppenheim, editors, *Wavelets and Statistics*, pp. 125–150, New York, 1995. Springer-Verlag.

[DJ94] David Donoho and Iain Johnstone. "Ideal Spatial Adaptation Via Wavelet Shrinkage." *Biometrika*, **81**:425–455, 1994.

[DJ95] David Donoho and Iain Johnstone. "Adapting to Unknown Smoothness Via Wavelet Shrinkage." *J. Amer. Stat. Assoc.*, **90**:1200–1224, 1995.

[DL92a] Ronald DeVore and Bradley Lucier. "Fast Wavelet Techniques for Near-Optimal Image Processing." In *IEEE Military Commun. Conf. Rec.*, pp. 1129–1135, San Diego, CA, 1992. IEEE Press.

[DL92b] Ronald DeVore and Bradley J. Lucier. "Wavelets." *Acta Numerica*, **1**:1–56, 1992.

[Don95] David Donoho. "Denoising by Soft Thresholding." *IEEE. Trans. Inform. Theory*, **41**(3):613–627, 1995.

[DS96] David Dobson and Otmar Scherzer. "Analysis of Regularized Total Variation Penalty Methods for Denoising." *Inverse Problems*, **12**:601–617, 1996.

[DT05] Ingrid Daubechies and Gerd Teschke. "Variational Image Restoration by Means of Wavelets: Simultaneous Decomposition, Deblurring and Denoising." *Applied and Computational Harmonic Analysis*, **19**(1):1–16, 2005.

[Egg93] Paul P. B. Eggermont. "Maximum Entropy Regularization for Fredholm Integral Equations of the First Kind." *SIAM J. Math. Anal.*, **24**:1557–1576, 1993.

[EHN96] Heinz Engl, Martin Hanke, and Andreas Neubauer. *Regularization of Inverse Problems.* Kluwer Academic Publishers, Dordrecht, The Netherlands, 1996.

[EL93] Heinz Engl and Gerhard Landl. "Convergence Rates for Maximum Entropy Regularization." *SIAM J. Numer. Anal.*, **30**:1509–1536, 1993.

[EO04] Selim Esedoḡlu and Stanley Osher. "Decomposition of Images by the Anisotropic Rudin-Osher-Fatemi Model." *Comm. Pure Appl. Math.*, **57**(12):1609–1626, 2004.

[ET99] Ivar Ekeland and Roger Téman. *Convex Analysis and Variational Problems.* Classics Appl. Math. 28, SIAM, Philadelphia, 1999.

[GB97] Hong-Ye Gao and Andrew G. Bruce. "WaveShrink with Firm Shrinkage." *Statist. Sinica*, **7**(4):855–874, 1997.

[GS00] Charles Groetsch and Otmar Scherzer. "Nonstationary Iterated Tikhonov-Morozov Method and Third Order Differential Equations for the Evaluation of Unbounded Operators." *Math. Methods Appl. Sci.*, **23**:1287–1300, 2000.

[GSZ04] Guy Gilboa, Nir Sochen, and Yehoshua Y. Zeevi. "Image Enhancement and Denoising by Complex Diffusion Processes." *IEEE Trans. Pattern Analysis. and Machine Intelligence*, **25**(8):1020–1036, 2004.

[GSZ05] Guy Gilboa, Nir Sochen, and Yehoshua Y. Zeevi. "Estimation of Optimal PDE-Based Denoising in the SNR Sense." *UCLA-CAM-Report 05-48, Los Angeles, CA, USA*, 2005. to appear in IEEE Trans. Image Proc.

[GY04] Donald Goldfarb and Wotao Yin. "Second-Order Cone Programming Methods for Total Variation based Image Restoration." *CORC Report TR-2004-05, Columbia University, New York, NY*, 2004.

[Han97] Martin Hanke. "A Regularizing Levenberg-Marquardt Scheme, with Applications to Inverse Groundwater Filtration Problems." *Inverse Problems*, **13**:79–95, 1997.

[HBO05] Lin He, Martin Burger, and Stanley Osher. "Iterative Total Variation Regularization with Non-Quadratic Fidelity." *UCLA-CAM Report 05-38, Los Angeles, CA, USA*, 2005. to appear in Journal of Mathematical Imaging and Vision.

[HG98] Martin Hanke and Charles Groetsch. "Nonstationary Iterated Tikhonov Regularization." *J. Optim. Theory Appl.*, **98**:37–53, 1998.

[HM04] Ali Haddad and Yves Meyer. "Variational Methods in Image Processing." *UCLA-CAM-Report 04-52, Los Angeles, CA, USA*, 2004.

[HMO05] Lin He, Antonio Marquina, and Stanley Osher. "Blind Deconvolution Using TV Regularization and Bregman Iteration." *International Journal of Imaging Systems and Technology*, **15**(1):74–83, 2005.

[HS06] Walter Hinterberger and Otmar Scherzer. "Variational Methods on the Space of Functions of Bounded Hessian for Convexification and Denoising." *Computing*, **76**:109–133, 2006.

[Kiw97] Krzysztof Kiwiel. "Proximal Minimization Methods with Generalized Bregman Functions." *SIAM J. Control Optim.*, **35**:1142–1168, 1997.

[KMS00] Ron Kimmel, Ravikanth Malladi, and Nir Sochen. "Images as Embedding Maps and Minimal Surfaces: Movies, Color, Texture, and Volumetric Medical Images." *International Journal of Computer Vision*, **39**(2):111–129, 2000.

[KO05] Stefan Kindermann and Stanley Osher. "Saddle Point Formulation for a Cartoon-Texture Decomposition." *UCLA-CAM-Report 05-42, Los Angeles, CA, USA*, 2005.

[LN05] Johan Lie and Jan Nordbotten. "Inverse Scale Spaces for Nonlinear Regularization." *UCLA-CAM-Report 05-67, Los Angeles, CA, USA*, 2005. submitted to Journal of Mathematical Imaging and Vision.

[LOT04] Marius Lysaker, Stanley Osher, and Xue-Cheng Tai. "Noise Removal Using Smoothed Normals and Surface Fitting." *IEEE Trans. Image Proc.*, **13**:1345–57, 2004.

[LV05a] Triet Le and Luminita Vese. "Image Decomposition Using Total Variation and Div(BMO)." *Multiscale Model. Simul.*, **4**(2):390–423, 2005.

[LV05b] Linh Lieu and Luminita Vese. "Image Restoration and Decompostion Via Bounded Total Variation and Negative Hilbert-Sobolev Spaces." *UCLA-CAM-Report 05-33, Los Angeles, CA, USA*, 2005.

[Mey90] Yves Meyer. *Ondelettes et Opérateurs I: Ondelettes*. Hermann, Paris, 1990.

[Mey92] Yves Meyer. *Wavelets and Operators.* Cambridge Univ. Press, New York, 1992.

[Mey01] Yves Meyer. *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations.* AMS, Providence, RI, 2001.

[MN03] Pavel Mrázek and Mirko Navara. "Selection of Optimal Stopping Time for Nonlinear Diffusion Filtering." *International Journal of Computer Vision*, **52**(2/3):189–203, 2003.

[MO00] Antonio Marquina and Stanley Osher. "Explicit Algorithms for a New Time Dependent Model Based on Level Set Motion for Nonlinear Deblurring and Noise Removal." *SIAM J. Sci. Comput.*, **22**(2):387–405, 2000.

[Nik97] Mila Nikolova. "Estimées Localement Fortement Homogénes." *C. R. Acad. Sci. Paris Sér. I Math.*, **325**:665–670, 1997.

[Nik00] Mila Nikolova. "Local Strong Homogeneity of a Regularized Estimator." *SIAM J. Appl. Math.*, **61**:633–658, 2000.

[Nik04] Mila Nikolova. "Weakly Constrained Minimization. Application to the Estimation of Images and Signals Involving Constant Regions." *Journal of Mathematical Imaging and Vision*, **21**:155–175, 2004.

[OBG05] Stanley Osher, Martin Burger, Donald Goldfarb, Jinjun Xu, and Wotao Yin. "An Iterative Regularization Method for Total Variation Based Image Restoration." *Multiscale Model. and Simul.*, **4**:460–489, 2005.

[OR90] Stanley Osher and Leonid Rudin. "Feature-Oriented Image Enhancement Using Shock Filters." *SIAM J. Numer. Anal.*, **27**(4):919–940, 1990.

[OS88] Stanley Osher and James Sethian. "Fronts Propagating with Curvature Dependent Speed; Algorithms Based on Hamilton-Jacobi Formulations." *J. Comput. Phys.*, **79**:12–49, 1988.

[OSV03] Stanley Osher, Andrés Solé, 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. "The Discrepancy Principle for Iterative and Parametric Methods to Solve Linear Ill-Posed Problems." *Numer. Math.*, **75**:99–120, 1996.

[PM90] Pietro Perona and Jitendra Malik. "Scale-Space and Edge Detection Using Anisotropic Diffusion." *IEEE Trans. PAMI*, **12(7)**:629–639, 1990.

[Rin00] Wolfgang Ring. "Structural Properties of Solutions of Total Variation Regularization Problems." *M2AN Math. Model. Numer. Anal.*, **34**:799–810, 2000.

[RLO03] Leonid Rudin, Pierre-Louis Lions, and Stanley Osher. "Multiplicative Denoising and Deblurring: Theory and Algorithms." In Stanley Osher and Nikos Paragios, editors, *Geometric Level Set Methods in Imaging Vision and Graphics*, pp. 103–121. Springer, New York, 2003.

[RO94] Leonid Rudin and Stanley Osher. "Total Variation Based Image Restoration with Free Local Constraints." In *Proc. IEEE Int. Conf. Image Proc.*, volume 1, pp. 31–35, Austin, TX, USA, 1994. IEEE Press.

[ROF92] Leonid Rudin, Stanley Osher, and Emad Fatemi. "Nonlinear Total Variation Based Noise Removal Algorithms." *Phys. D*, **60**:259–268, 1992.

[SC03] David Strong and Tony Chan. "Edge-Preserving and Scale-Dependent Properties of Total Variation Regularization." *Inverse Problems*, **19**:S165–S187, 2003.

[SED05] Jean-Luc Starck, Michael Elad, and David Donoho. "Image Decomposition Via the Combination of Sparse Representations and a Variational Approach." *IEEE Trans. Image Proc.*, **14**(10):1570–1582, 2005.

[SG01] Otmar Scherzer and Charles Groetsch. "Inverse Scale Space Theory for Inverse Problems." *Lecture Notes in Computer Sciences*, **2106**:317–325, 2001.

[SW00] Otmar Scherzer and Joachim Weickert. "Relations Between Regularization and Diffusion Filtering." *Journal of Mathematical Imaging and Vision*, **12**:43–63, 2000.

[SWB04] Gabriele Steidl, Joachim Weickert, Thomas Brox, Pavel MráZek, and Martin Welk. "On the Equvalence of Soft Wavelet Shrinkage, Total Variation Diffusion, Total Variation Regularization, and SIDEs." *SIAM J. Numer. Anal.*, **42**(2):686–713, 2004.

[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.

[TO05] Yen-Hsi Tsai and Stanley Osher. "Total Variation and Level Set Based Methods in Image Science." *Acta Numerica*, **14**:509–573, 2005.

[TWB03] Tolga Tasdizen, Ross Whitaker, Paul Burchard, and Stanley Osher. "Geometric Processing Via Normal Maps." *ACM Trans. Graphics*, **22**:1012–1033, 2003.

[VO02] Luminita Vese and Stanley Osher. "Numerical Methods for P-Harmonic Flows and Applications to Image Processing." *SIAM J. Numer. Anal.*, **40**:2085–2104, 2002.

[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.

[Vog02] Curt Vogel. *Computational Methods for Inverse Problems*. Frontiers Appl. Math. 23, SIAM, Philadelphia, 2002.

[Wei99] Joachim Weickert. "Coherence-Enhancing Diffusion Filtering." *International Journal Computer Vision*, **31**:111–127, 1999.

[Wit83] Andrew P. Witkin. "Scale Space Filtering." In *Proc. Int. Joint Conf. On Artificial Intelligence*, pp. 1019–1023, 1983.

[XO06] Jinjun Xu and Stanley Osher. "Iterative Regularization and Nonlinear Inverse Scale Space Applied to Wavelet Based Denoising." *UCLA-CAM-Report 06-11, Los Angeles, CA, USA*, 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 L^1 Functional." *Lecture Notes in Computer Science*, **3752**:73–84, 2005.