Comm. Math. Sci.
Vol. 4, No. 1, pp. 179–212

NONLINEAR INVERSE SCALE SPACE METHODS*

MARTIN BURGER[†], GUY GILBOA[‡], STANLEY OSHER[§], AND JINJUN XU[¶]

Abstract. In this paper we generalize the iterated refinement method, introduced by the authors in a recent work, to a time-continuous inverse scale-space formulation. The iterated refinement procedure yields a sequence of convex variational problems, evolving toward the noisy image.

The inverse scale space method arises as a limit for a penalization parameter tending to zero, while the number of iteration steps tends to infinity. For the limiting flow, similar properties as for the iterated refinement procedure hold. 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. 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.

Key words. Iterated refinement techniques, inverse scale space methods, image restoration, total variation, Denoising, evolution equations.

AMS subject classifications. 47A52, 49M30, 65J22, 68U18

1. Introduction

The processing of noisy images is a central task in mathematical imaging. Over the last decades, a variety of methods have been proposed ranging from filtering methods over variational approaches to techniques based on the solution of partial differential equations (cf. [10]). 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 [26] of the form

$$\frac{\partial u}{\partial t} = \mathrm{div}(\gamma(|\nabla u|^2)\nabla u), \tag{1.1}$$

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

---

[*] Received: November 24, 2005; accepted (in revised version): January 17, 2006. Communicated by Shi Jin.

[†] Institut für Industriemathematik, Johannes Kepler Universität, Altenbergerstr. 69, A 4040 Linz, Austria (martin.burger@jku.at).

[‡] Department of Mathematics, UCLA, 520 Portola Plaza, Los Angeles, CA 90095, USA (gilboa@math.ucla.edu).

[§] Department of Mathematics, UCLA, 520 Portola Plaza, Los Angeles, CA 90095, USA (sjo@math.ucla.edu).

[¶] Department of Mathematics, UCLA, 520 Portola Plaza, Los Angeles, CA 90095, USA (jjxu@math.ucla.edu).

nonlinear scale-spaces see [2, 15, 19, 26, 33, 34] and the references therein. Diffusion filters can be related to regularization theory (cf. [30]) with certain regularization functionals, but theoretical foundations of choosing optimal stopping times are still missing (see [16, 22] for two recent studies concerning the stopping time problem).

Recently, *inverse scale space methods* have been introduced in [29], 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 we normalize 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. [17, 29]) with the same regularization functionals as for diffusion filters. The construction of inverse scale space methods in [29] 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 paper we present a different version of constructing inverse scale space methods as the limit of an iterated refinement procedure previously introduced by the authors (cf. [24]) and demonstrate its applicability to image restoration. With the new approach we are able to easily implement nonlinear inverse scale space methods even for the total variation functional, and, in contrast to diffusion filters, we obtain a rigorously justified and simple stopping criterion for the methods.

2. Iterated refinement

In [24], an iterated refinement procedure for total variation restoration was introduced, motivated by the variational problem

$$u=\underset{u\in BV(\Omega)}{\operatorname{argmin}}\left\{|u|_{BV}+\frac{\lambda}{2}\|f-u\|_{L^2}^2\right\} \tag{2.1}$$

for some scale parameter $\lambda>0$, where $BV(\Omega)$ denotes the space of functions with bounded variation on Ω, equipped with the BV seminorm which is formally given by

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

also referred to as the *total variation* (TV) of u. Problem (2.1) is called the ROF model, introduced to the field of image restoration in [28].

In [24] the authors showed that an iterative procedure (which turned out to be equivalent to Bregman's relaxation method, cf. [7], and proximal point algorithms, cf. [12]) could be used to improve the quality of regularized solutions to inverse problems, based on regularization functionals as in (2.1). Given a convex functional $J:\mathcal{U}\to\mathbb{R}$ (on a suitable Banach space $\mathcal{U}$), e.g., $J(u)=|u|_{BV}$, the *iterated refinement method* defines a sequence $\{u_k\}$ by:

- Set $u_0=0$, $p_0=0$;
- Given u_{k-1} and $p_{k-1}\in\partial J(u_{k-1})$, $k\geq 1$, perform the following two steps:
  (i) compute $u_k=\underset{u}{\operatorname{arg\,min}}\, Q_k(u)$ with

$$Q_k: u\longmapsto J(u)-J(u_{k-1})-\langle p_{k-1},u-u_{k-1}\rangle+\frac{\lambda}{2}\left\|f-u\right\|_{L^2}^2, \tag{2.2}$$

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

(ii) update the dual variable $p_k = p_{k-1} + \lambda(f - u_k) \in \partial J(u_k)$.

- Increase k by 1 and continue.

Here $\partial J(u_k)$ is the subgradient of the functional J, given by

$$\partial J(u) = \{p \in \mathcal{U}^* \mid \langle p, v-u\rangle \leq J(v) - J(u),\ \forall\ v \in \mathcal{U}\},$$

which is single-valued, i.e., $\partial J(u) = \{J'(u)\}$ if J is Fréchet-differentiable. This procedure improves the quality of reconstruction for many problems with discontinuous solutions, e.g., deblurring and denoising of images (cf. [18, 24]) when the least-squares term $\|f-u\|_{L^2}^2$ is replaced by an appropriate fitting term for individual examples.

Note that the regularization term used in the first step is a so-called generalized *Bregman distance* between u and u_{k-1}, defined as follows,

$$D(u,v) = J(u) - J(v) - \langle u-v, p\rangle, \quad p \in \partial J(v), \tag{2.3}$$

Note that the subgradient may contain more than one element if the functional J is not differentiable, so that the distance would depend on the specific choice of the subgradient. However, we shall suppress the dependence on the subgradient in the notation below. Note that for smooth, strictly convex functionals the subgradient contains at most one element and $D(u,v)$ is a scalar "distance", which is strictly positive for $u \neq v$. We can then rewrite the functional Q_k minimized in each iteration step as:

$$Q_k(u) = D(u, u_{k-1}) + \frac{\lambda}{2}\|f-u\|_{L^2}^2.$$

The Bregman distance and the associated iteration were not used in this fashion previously, but they have been rather employed to minimize functions $H(u,f)$ where H is a (usually complicated) convex function of u having a unique minimum (cf. e.g., [12]).

It was shown in [24] that the iterated refinement method yields a well-defined sequence of minimizers u_k and subgradients $p_k \in \partial J(u_k)$. Moreover, it was proved that the sequence $\{u_k\}$ yields a monotonously decreasing residual $\|u_k - f\|_{L^2}$, and

$$f \in BV(\Omega) \quad \Rightarrow \quad \|u_k - f\|_{L^2}^2 \leq \frac{J(f)}{k},$$

i.e., u_k converges monotonically to f in $L^2(\Omega)$ with a rate of $\frac{1}{\sqrt{k}}$. Of course, this convergence result does not give useful information on the behavior of the method as a denoising method, in particular for the typical case of a noisy image f.

The key denoising result obtained in [24] is as follows: for $g \in BV(\Omega)$ we have

$$D(g, u_k) < D(g, u_{k-1}) \quad \text{if} \quad \|f - u_k\|_{L^2} \geq \tau \|g - f\|_{L^2} \tag{2.4}$$

for any $\tau > 1$. Thus, the Bregman distance between a restored image u_k and a possible exact image g is decreasing until the L^2-distance of f and u_k is larger than the L^2-distance of f and g. This result can be used to construct a stopping rule for our iterative procedure. If we have an estimate of the variance of the noise, i.e., $f = g + n, \|n\|_{L^2} = \sigma$, where $g \in BV(\Omega)$ is the noise-free image and n is the noise, then we can stop the iteration at the first k for which $\|f - u_{k+1}\|_{L_2} < \tau\sigma$. The choice of τ allows some freedom to apply the stopping rule also in the case when we only know an upper bound for σ.

It is interesting to note that for denoising with any convex regularization $J(u)$, the sequence $\{u_k\}$ has the following interpretation (cf. [24]):

- Define $u_0 = 0$, $v_0 = f$.
- Then inductively for $k \geq 1$, let

$$u_k = \arg\min_u \left\{ J(u) + \frac{\lambda}{2} \|f + v_{k-1} - u\|_{L^2}^2 \right\}$$

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

In other words, we *add* the “small scales” v_{k-1} back to f and perform (e.g. ROF) minimization with f replaced by $f + v_{k-1}$ and decompose this function into “large scales” (u_k) plus “small scales” (v_k). This interpretation already yields a multiscale interpretation of the method, since the “small scales” are somehow doubled in each step and so their larger parts can be incorporated into the large scale part after the next iteration. A related procedure involving the ROF model using Tikhonov-Morozov rather than the Bregman iteration which multiplies λ by two in each step yields a multiscale method suggested in [17] and analyzed in [31]. This procedure does not obviously extend to the nonlinear inverse scale-space method, discussed below, and does not appear to satisfy an estimate analogous to that of equation (2.4).

3. Inverse scale space methods

In the following we generalize the concept of inverse scale space theory introduced in [17, 29] in the context of Tikhonov regularization for the case

$$J(u) = \frac{1}{2} \int_\Omega |\nabla u|^2. \tag{3.1}$$

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

$$J(u) = \int_\Omega \sqrt{|\nabla u|^2 + \epsilon^2}, \tag{3.2}$$

$\epsilon > 0$ is a small constant.

Recall that for a special $\lambda > 0$ the iterative refinement procedure constructs sequences u_k^λ of primal and p_k^λ of dual variables such that $u_0^\lambda = p_0^\lambda = 0$,

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

From the Euler-Lagrange equation

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

we are led to the dual iteration

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

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 $t_k = k\Delta t$, $p^{\Delta t}(t_k) = p_k^{\Delta t}$, and $u^{\Delta t}(t_k) = u_k^{\Delta t}$, we have $p_{k-1}^{\Delta t} = p^{\Delta t}(t_{k-1}) = p^{\Delta t}(t_k - \Delta t)$ and thus

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

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

$$\frac{\partial p}{\partial t}(t) = f - u(t), \qquad p(t) \in \partial J(u(t)), \tag{3.3}$$

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 straight-forward way for $\int_\Omega f \neq 0$, with the only difference being 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 (3.3) exists and is well behaved (which can be shown under reasonable assumptions on the functional J, in particular for total variation, cf. [8]), it is an inverse scale space method in the sense of [17]. 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 (3.3) 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 iterated refinement 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 7, but the differences are small. (Of course both methods are significantly better than solving the standard variational problem (2.1), even with the best choice of the parameter λ). We mention that with a 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 (2.1), 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.

3.1. Behaviour for quadratic regularization. We start by briefly reviewing the results obtained in [17] for the quadratic regularization (3.1). In this case we obtain from the variation of the functional J the boundary value problem

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

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. 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 $\mathbb{R}^N$ 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. [13]), defined by

$$J^*(p) := \sup_u \left\{ \langle u, p \rangle - J(u) \right\}. \tag{3.4}$$

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 (3.3) 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 $H^* = \partial_{pp}^2 J^*$ for the Hessian of the dual functional. If J^* is strictly convex, then there exists a constant $a > 0$ such that

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

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

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

and using Gronwall's inequality we have

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

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

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

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. [9]). From (3.3) we proceed formally to

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

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 $D(f,u(t))$ and $D(u(t),f)$ converge to zero. Moreover, we can integrate the last inequality from time zero to t, which gives

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

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

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

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

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

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

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

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 $\|f-g\|_{L^2} \leq \sigma$. The above estimate guarantees that the distance $D(g,u)$ 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 t_* such that $\|f-u(t_*)\|_{L^2} = \sigma$. This stopping criterion is well-known in the theory of iterative regularization of inverse problems as the so-called *discrepancy principle* (cf. [14, 27] 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 regularizations 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,

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

(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 $D(g,u)$ can vanish for g not identical to u, it is fairly easy to show that $D(g,u) = 0$ implies that $g = R(u)$, R being a nondecreasing 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 [1]. The proof can be outlined as follows: $D(g,u) = 0$ 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 $R' = r$.

3.3. 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 (3.3) 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. *Under the above conditions and notations, there are the following connections between u, p and $\tilde{u}$, $\tilde{p}$:*

- *If J is quadratic (such as* (3.1)*), 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 (3.3) 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 (3.3) with subgradient $\tilde{p}(t)$. Again, by uniqueness, we obtain that for a solution u of (3.3) 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 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.

3.4. Comparison to ROF scale space. In contrast to the evolution (3.3) 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 (2.1), is a less appealing alternative. Note that the improvement seen with respect to (2.1) seen in numerical experiments is one of the major motivations for investigating (3.3). Let us consider the inverse scale space defined by $w(t)$ which satisfies:

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

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 $H(f,u)$ 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

$$\partial_t p(t) + t\partial_t w(t) = f - w(t). \qquad w(0)=0, p(0)=0. \tag{3.5}$$

We claim that this evolution equation is not as useful as our inverse scale space equation (3.3) (which it resembles). To show this, we first examine the convergence to f:

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

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

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

and an analogous reasoning as in Proposition 3.1 is not apparent due to the second term. In fact, for the quadratic case (3.1) we have

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

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 (2.1) can lose a lot of the image variation (cf. [21]) if $\lambda < \infty$.

4. Inverse scale space for signals

In the following we discuss the numerical solution of (3.3) in spatial dimension one. 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 $J(u) = \int \sqrt{|\nabla u|^2 + \epsilon^2}$, which yields

$$J'(u(t)) = -\operatorname{div}\left(\frac{\nabla u(t)}{\sqrt{|\nabla u(t)|^2 + \epsilon^2}}\right) = p(t). \tag{4.1}$$

Note that since $\partial J(u+c) = \partial J(u)$, the solution of (4.1) is not unique if we take the standard assumption that u satisfies the homogeneous Neumann boundary condition. In this case, the solvability condition is $\int p(x,t)\,dx = 0$ for all t and the conservation of mean value discussed above provides an additional property implying uniqueness, namely $\int u \; dx \equiv \int f \; dx = 0$.

For a fixed time t, we have to solve

$$-\left(\frac{u_x}{\sqrt{u_x^2 + \epsilon^2}}\right)_x = p \text{ in } I = (a,b), \qquad \int_a^b u\,dx = 0. \tag{4.2}$$

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

$$q(x,t) = -\int_a^x p(s,t)\,ds = \int_x^b p(s,t)\,ds \tag{4.3}$$

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

$$u(x,t) = \epsilon \int_a^x \frac{q(y,t)}{\sqrt{1 - q^2(y,t)}}\,dy + C \tag{4.4}$$

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.

5. 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 to write the general expressions of the discrete Bregman procedure and the direct inverse scale-space having any convex smooth fidelity terms $H(f,u)$.

5.1. General fidelity term $H(f,u)$.

The process is based on a convex variational problem of the general form

$$\min_u \{J(u) + \lambda H(f,u)\}, \tag{5.1}$$

where $H(f,u)$ is usually a fidelity to a known image (or signal) f, (in the L^2 case studied before $H(f,u) = \frac{1}{2}\|f-u\|_{L^2}^2$). In our discussion we assume for simplicity that f is normalized beforehand to have a zero mean: $\int_\Omega f = 0$. We also assume $p(0) = 0$.

The general form of the series of variational minimizations, introduced in [24], is

$$u_k = \underset{u}{\operatorname{argmin}}\left\{ D(u,u_{k-1}) + \lambda H(f,u) \right\} \tag{5.2}$$

where $u_0 = 0$ and $p_0 = 0$. Expanding $D(\cdot,\cdot)$ according to (2.3) and omitting constant parts which are not relevant to the minimization yields

$$u_k = \underset{u}{\operatorname{argmin}}\left\{ J(u) - \langle u, p(u_{k-1})\rangle + \lambda H(f,u) \right\}. \tag{5.3}$$

The Euler-Lagrange equation of (5.3) is

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

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

$$\partial_t p = -\partial_u H(f,u(t)), \qquad p \in \partial J(u) \tag{5.4}$$

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

5.2. Relaxed inverse scale-space method. The concise formulation of (5.4) 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 (5.4) by using standard variational formulations, which are simple to compute.

Let us revisit the series of Bregman iterations stated in equation (5.3). Using the update formula for the decomposed"noise" $v_k = \frac{p_k}{\lambda}$ (cf. [24]) and the first order optimality conditions, we deduce

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

and hence,

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

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

$$u_k = \underset{u}{\operatorname{argmin}}\left\{ J(u) + \lambda H(f,u) + \lambda \sum_{j=0}^{k-1} \langle u, \partial_u H(f,u_j)\rangle \right\}. \tag{5.5}$$

or, coupled for u_k and v_k

$$\begin{aligned} u_k &= \underset{u}{\operatorname{argmin}}\left\{ J(u) + \lambda (H(f,u) - \langle u, v_{k-1}\rangle) \right\}, \\ v_k &= v_{k-1} - \partial_u H(f,u_k), \end{aligned} \tag{5.6}$$

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

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:

$$\frac{\partial u}{\partial t} = -p + \lambda(-\partial_u H(f,u) + v_k), \quad p \in \partial J(u), \quad u|_{t=0} = u_{k-1}, \tag{5.7}$$

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 $\partial_u H(f,u_k)$ and update v_{k+1}. Then k is incremented by one and the process resumes, such that in each iteration (5.7) 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 (5.6) can be viewed as an iterative descent in v_k for minimizing $H(f,u_k)$. 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:

$$\partial_\tau v = -\partial_u H(f,u_k), \quad v|_{\tau=0} = v_k, \tag{5.8}$$

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 (5.7) and (5.8), 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 (5.7) by $v(t)$ and u_k in (5.8) by $u(t)$ yields the *relaxed inverse-scale-space* flow:

$$\begin{aligned} \partial_t u &= -p + \lambda(-\partial_u H(f,u) + v), \\ \partial_t v &= -\alpha \partial_u H(f,u), \end{aligned} \tag{5.9}$$

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

5.2.1. Second order in time formulation. If H and J 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 (5.9) and substituting for v_t by using the second equation, yielding the following evolution:

$$\partial^2_{tt} u = -\partial_t p(u) - \lambda(\partial_t(\partial_u H(f,u)) - \alpha \partial_u H(f,u)), \tag{5.10}$$

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

$$\partial^2_{tt} u = -(J''(u) + \lambda \partial^2_{uu} H(u,f))\partial_t u - \lambda\alpha \partial_u H(f,u). \tag{5.11}$$

5.2.2. Relation to the direct flow. In order to understand the relation to the original inverse scale space formulation, Eq. (5.4), we consider the special case $H(f,u) = \frac{1}{2}\|u - f\|^2$, 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

$$\begin{aligned} \alpha\lambda \partial_{\hat{t}} u &= -p + \lambda(f - u) + w, \\ \partial_{\hat{t}} w &= f - u. \end{aligned} \tag{5.12}$$

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 7 (Figs. 7.2 and 7.3).

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

- *Linear model:* $J = \frac{1}{2}\|\nabla u\|_2^2$, $H = \frac{1}{2}\|f-u\|_2^2$.

$$\begin{aligned} u_t &= \Delta u + \lambda(f-u+v), \\ v_t &= \alpha(f-u). \end{aligned} \tag{5.13}$$

- *ROF model:* $J = \int_\Omega |\nabla u|$, $H = \frac{1}{2}\|f-u\|_2^2$.

$$\begin{aligned} u_t &= \operatorname{div}\left(\frac{\nabla u}{|\nabla u|}\right) + \lambda(f-u+v), \\ v_t &= \alpha(f-u). \end{aligned} \tag{5.14}$$

- *$TV-L^1$ model (cf. [11]):* $J = \int_\Omega |\nabla u|$, $H = \|f-u\|_1$.

$$\begin{aligned} u_t &= \operatorname{div}\left(\frac{\nabla u}{|\nabla u|}\right) + \lambda(\operatorname{sign}(f-u)+v), \\ v_t &= \alpha \operatorname{sign}(f-u). \end{aligned} \tag{5.15}$$

  [Note that H is not strictly convex or smooth here and sign is just the notation for an element in the subgradient of H.]

- *Deconvolution by ROF:* $J = \int_\Omega |\nabla u|$, $H = \frac{1}{2}\|f - K*u\|_2^2$, where K is a real blurring kernel, $\hat{K}(x,y) = K(-x,-y))$ and $*$ denotes convolution.

$$\begin{aligned} u_t &= \operatorname{div}\left(\frac{\nabla u}{|\nabla u|}\right) + \lambda\left(\hat{K}*(f-K*u)+v\right), \\ v_t &= \alpha \hat{K}*(f-K*u). \end{aligned} \tag{5.16}$$

6. Properties of the relaxed method

6.1. 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 [20], 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 (5.11). In the linear case the subgradient is unique and given by $p = -\Delta u$, and $\partial_u H(f,u) = u - f$. The flow can be written as:

$$\partial_{tt} u + (\Delta + \lambda)\partial_t u + \lambda\alpha u = \lambda\alpha f, \tag{6.1}$$

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

$$r_\pm = \frac{-(\lambda+|\xi|^2) \pm \sqrt{(\lambda+|\xi|^2)^2 - 4\alpha\lambda}}{2}. \tag{6.2}$$

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

$$r_\pm \approx \frac{-(\lambda+|\xi|^2)(1 \pm (1-\frac{2\alpha\lambda}{(\lambda+|\xi|^2)^2}))}{2}, \tag{6.3}$$

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

$$U(\xi) = (c_+ e^{r_+ t} + c_- e^{r_- t} + 1)F(\xi) \tag{6.4}$$

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

6.2. ROF model.

The ROF (TV-L^2) 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.

6.2.1. Convergence to steady-state.

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

$$e(t) = \frac{1}{2\lambda}\|u-f\|_2^2 + \frac{1}{2\alpha}\|v - \frac{q}{\lambda}\|_2^2, \tag{6.5}$$

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 [20]:

Proposition 6.1. *Let $u(0)$, $v(0)$ be an initial value such that $e(0) < \infty$, and let $u(t)$, $v(t)$ be the solution of (5.14) 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*

$$\|f - u(t_k)\| \to 0, \quad D(f, u(t_k)) \to 0, \quad D(u(t_k), f) \to 0. \tag{6.6}$$

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

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

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

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

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

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

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

Note that we slightly changed the original proof from [20] 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 (5.14).

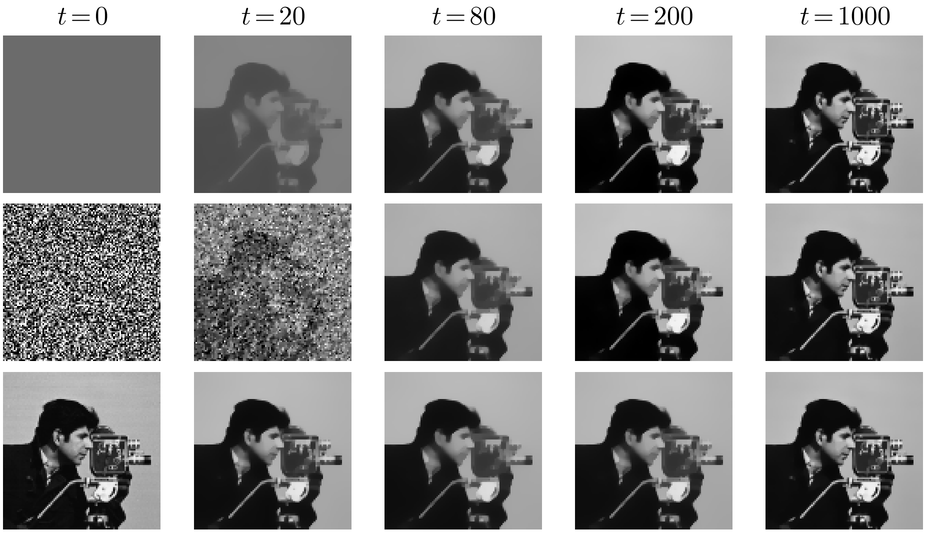


FIG. 6.1. *Evolution of* u *towards a clean image* f *with different initial conditions. Top* $u(0)=0$*, second row* $u(0)=n$ *(white Gaussian noise), bottom:* $u(0)=f$*.* [$v(0)=0, \lambda=0.02, \alpha=\lambda/4$].

6.2.2. 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 6.1 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 6.2

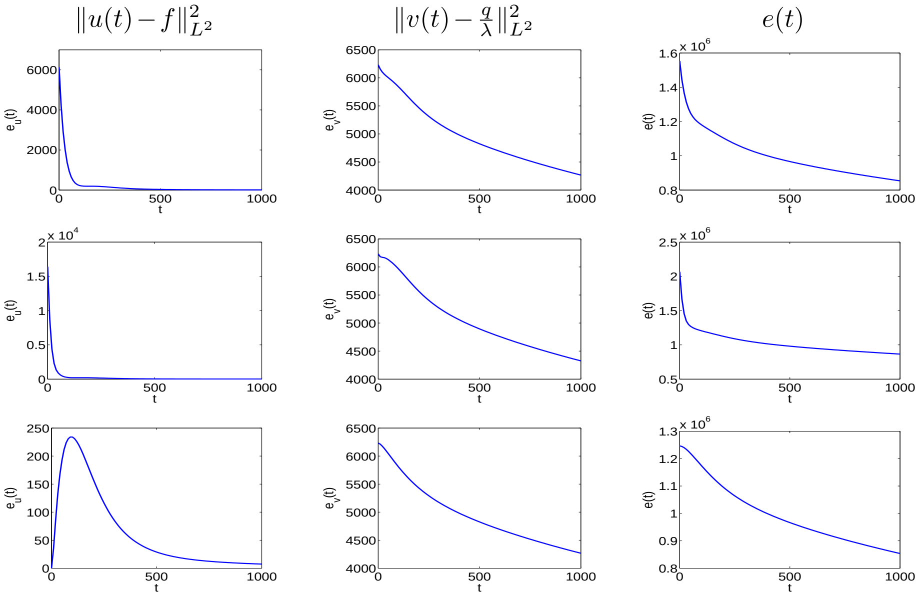


FIG. 6.2. *Distance from steady state as a function of time: u (left), v (middle) and the joint energy e (right) for different initial conditions. Top* $u(0)=0$, *second row* $u(0)=n$ *(white Gaussian noise), bottom:* $u(0)=f$.

the L^2 distance to the the steady states of u and v and the joint energy e (Eq. (6.5)) 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 $\|u-f\|^2$ 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.

6.2.3. 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 Section 6.2.5), where we have the same bound on α for monotone convergence. We show in Figure 6.3 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 6.4 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 6.3), 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 $\|u-f\|_2^2$ is decreasing monotonically in most cases for zero initial conditions and $\alpha\leq\frac{\lambda}{4}$. We have been able

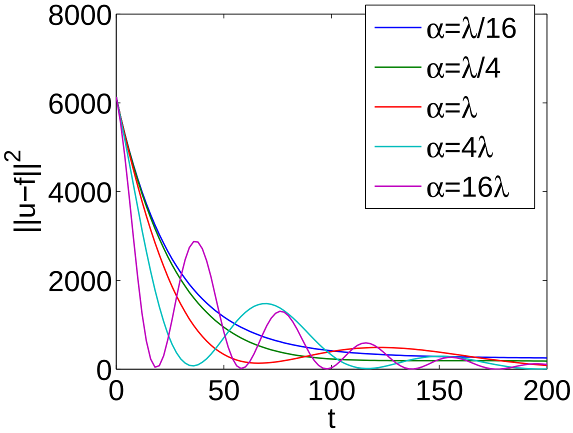


Fig. 6.3. *$\|u-f\|_2^2$ as a function of time for different values of α: $\alpha\in\{\frac{1}{16},\frac{1}{4},1,4,16\}\lambda$. [Cameraman image, $u(0)=v(0)=0$, $\lambda=0.02$].*

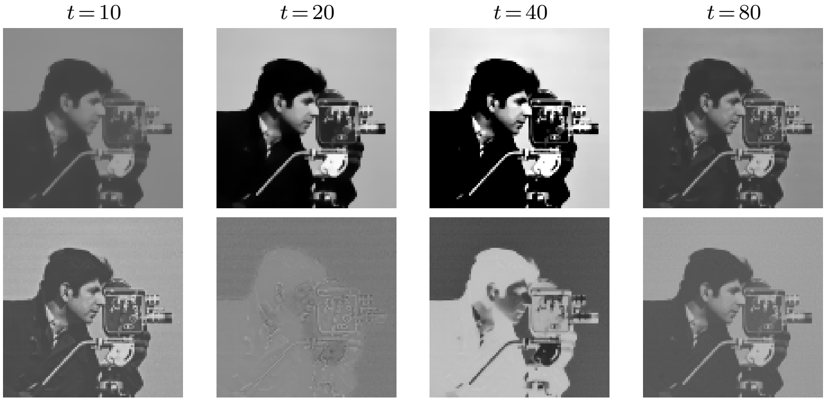


Fig. 6.4. *Example of an evolution with large α. $u(t)$ (top) and $f-u(t)$ (bottom) at times: $10,20,40,80$, for $\alpha=16\lambda$. [$u(0)=v(0)=0$, $\lambda=0.02$].*

to produce rare synthetic cases where $\|u-f\|_2^2$ is not monotone. This happens for very large features when λ is very large. However, from the 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 Figs. 6.5 and 6.6 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.

6.2.4. Complexity. The relaxed inverse scale space flow, in most cases, has about the same complexity as the standard gradient descent to steady state approach

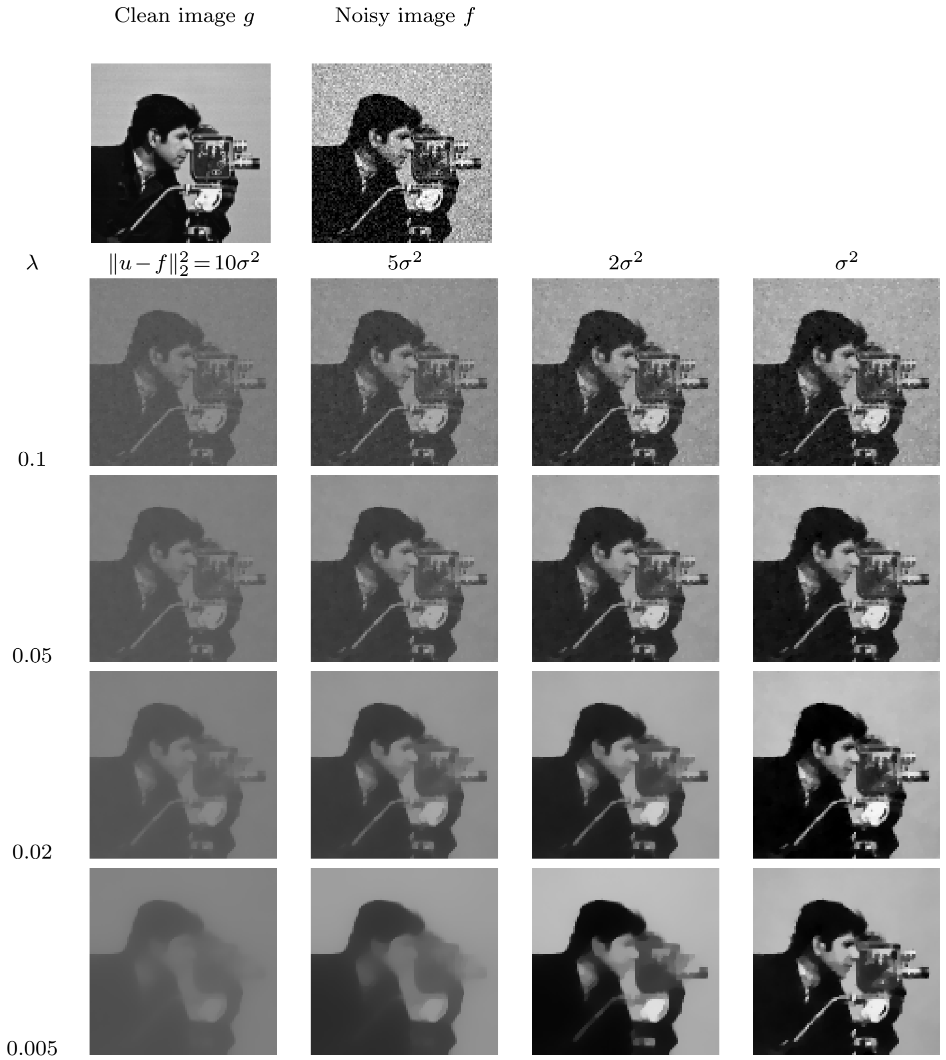


FIG. 6.5. *Denoising with different values of* λ. *Top row: original image (left) and noisy image (right). Second to fifth rows:* u *for the following values of* λ*:* $0.1, 0.05, 0.02, 0.005$*, respectively. Each column, from left to right, depicts the following* L^2 *norm of the residual part* $\|u-f\|_2^2$*:* $10\sigma^2, 5\sigma^2, 2\sigma^2, \sigma^2$*, respectively. [*$u(0)=v(0)=0$*,* $\alpha=\lambda/4$*,* $\sigma=20$*].*

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 [29], we obtain a step size of order one yielding stability, i.e., no severe restriction on large time steps.

6.2.5. 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 TV semi-norm (e.g. satisfies the source condition $q \in \partial J(f) \cap L^2$) and in some

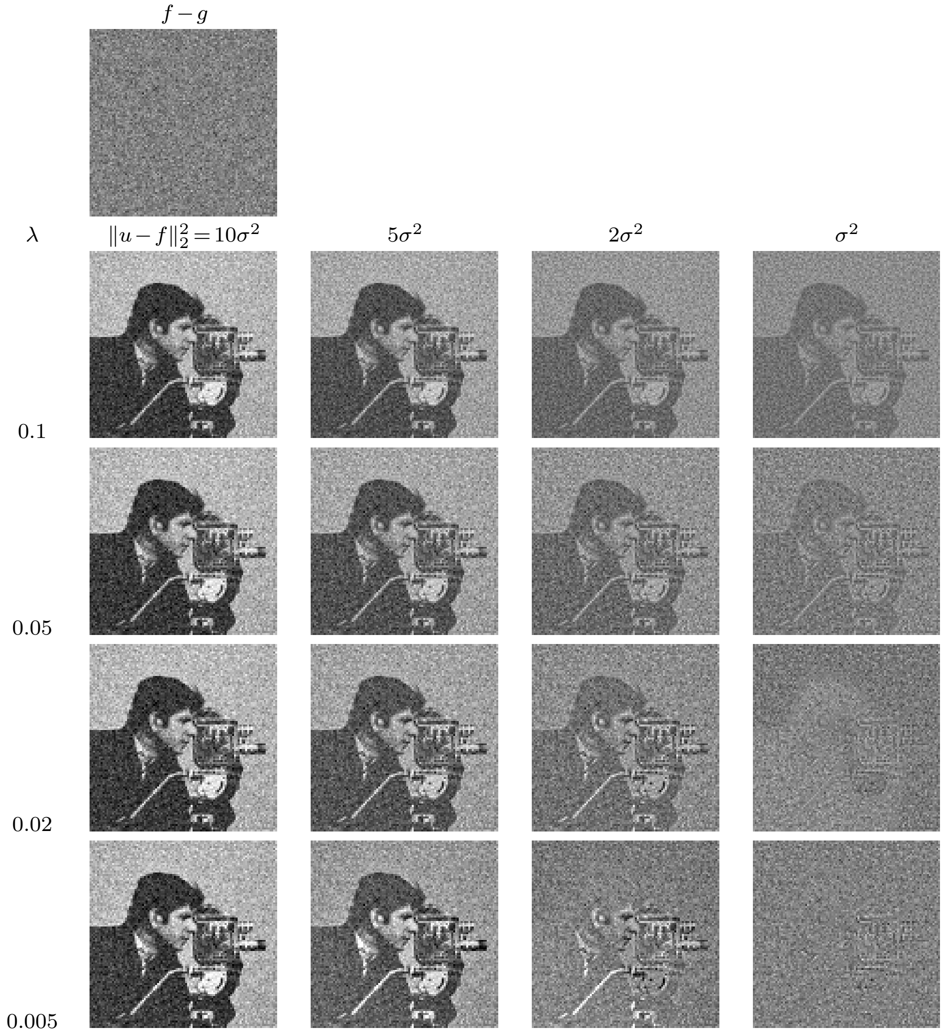


FIG. 6.6. *Denoising with different values of* λ. *Top row:* $f-g$, *an instance of white Gaussian noise* ($\sigma=20$) *which was added to the clean image. Second to fifth rows:* $f-u$ *for the following values of* λ: 0.1, 0.05, 0.02, 0.005, *respectively. Each column, from left to right, depicts the following* L^2 *norm of the residual part* $\|u-f\|_2^2$: $10\sigma^2, 5\sigma^2, 2\sigma^2, \sigma^2$, *respectively.*

cases allows a direct computation of a solution (cf. e.g. [21] 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

$$f(x) = \begin{cases} h\left(1 - \dfrac{R^2}{R_0^2}\right) & \text{if } |x| < R, \\ -h\dfrac{R^2}{R_0^2} & \text{else,} \end{cases} \tag{6.7}$$

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

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

PROPOSITION 6.2. *Let $J : BV(\Omega) \to \mathbb{R}$ be the total variation seminorm and let f be defined via* (6.7). *Then the function p defined via*

$$p(x) = \begin{cases} \dfrac{2}{R} & \textit{if } |x| < R, \\ -\dfrac{2R}{c_0 R_0^2} & \textit{else,} \end{cases} \tag{6.8}$$

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

Proof. It is straight-forward to compute

$$\int_\Omega p \; dx = 0, \qquad \int_\Omega pf \; dx = J(f) = 2\pi R h.$$

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

$$-\Delta q = p \quad \text{in } \Omega, \qquad \partial_n q = 0, \quad \text{on } \partial\Omega.$$

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

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

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

$$\begin{aligned} \int_\Omega p(u-f) \; dx &= \int_\Omega \nabla \cdot (-\nabla q) u \; dx - J(f) \\ &\le \sup_{\mathbf{g}, \|\mathbf{g}\|_\infty \le 1} \int_\Omega \nabla \cdot \mathbf{g} \; u \; dx - J(f) = J(u) - J(f), \end{aligned}$$

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 us to look for solutions of the form

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

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 6.2 it is easy to see that for $|p_1| \le \frac{2}{Rc_0}$, $p \in \partial J(0)$. 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 $|p_1(t)| \le \frac{2}{Rc_0}$ for $t \le t_1$, this yields

$$t_1 = \frac{2}{Rc_0 h}, \tag{6.9}$$

and by a simple integration

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

From Proposition 6.2 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 $u(.,t) = f$) for $t > t_1$, which is clearly a solution since

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

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

$$u(.,t) = \begin{cases} 0 & t < t_1, \\ f & t > t_1. \end{cases} \tag{6.10}$$

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 multicale 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 straight-forward 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, $v_1(t)=\alpha ht$ 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

$$t_1=\begin{cases}\frac{2-Rc_0h\lambda}{\alpha\lambda Rhc_0} & \lambda\leq\frac{2}{Rhc_0},\\ 0 & \text{otherwise.}\end{cases} \tag{6.11}$$

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. (6.9) (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

$$\frac{dr}{dt}(t)=q(t)-\lambda r(t), \quad \frac{dq}{dt}(t)=-\alpha\lambda r(t) \tag{6.12}$$

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

$$u(x,t)=\begin{cases}0, & 0\leq t\leq t_1\\ \left(-\left(\frac{\lambda h}{2}(t-t_1)+h\right)e^{-(t-t_1)\lambda/2}+h\right)f(x), & t_1<t<\infty\end{cases} \tag{6.13}$$

with t_1 as defined in (6.11). We will use this simple equation as a reference in the following numerical experiment.

6.2.6. 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 6.7 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 6.8). In Figure 6.8, left, the values of a cross-section of the disk are plotted for 40 equally spaced time points. On the right of Figure 6.8 the theoretical model of equation (6.13) (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. (6.13). Apart from that, the evolution is quite faithful to the model.

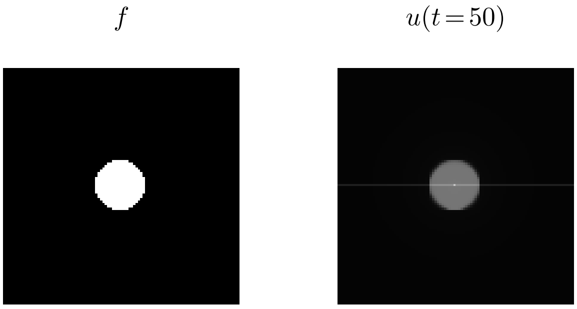


FIG. 6.7. *Disk evolution: f (left) and u (right). The brighter line and point in u show which values of u are plotted in Figure2.*

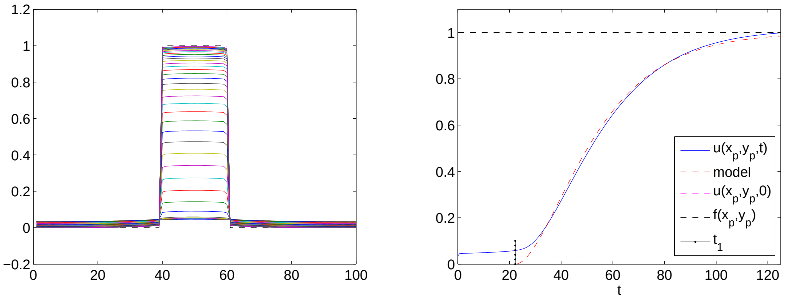


FIG. 6.8. *Disk evolution. Left: plots of a cross section of u at equally spaced time points. Dashed - cross section of f. Right: comparison between the simulation (solid, blue) and the theoretical model (dashed, red).*

6.2.7. 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 [21] we define the G-norm (the dual of the total variation norm for functions with mean zero) as

$$\|u\|_G := \inf_{\nabla\cdot\mathbf{g}=u} \||\mathbf{g}|_{\ell^2}\|_\infty.$$

Then we can use a characterization of subgradients obtained in [21] namely that $p \in \partial J(0)$ if and only if $\|p\|_G \leq 1$. Thus, we immediately obtain the following generalization of the behaviour in the initial phase

THEOREM 6.1. *Let $f \neq 0$ be as above and let $t_1 = \frac{1}{\|f\|_G}$. Then each solution (u,p) of* (3.3) *satisfies*

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

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

Proof. For $t < t_1$ we obtain that $p(.,t) \in \partial J(0)$ from the above reasoning and one easily checks that (u,p) defined as above is a solution of (3.3). 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 (3.3) we obtain $p = tf$, but $p(.,t) \notin J(0)$ 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 $\|f - u(t)\|$ is non-increasing in time (see Section 3.2), 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 $t > t_1$. □

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

We can also give a multiscale interpretation of Theorem 6.1. 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 $t_1 = \frac{\|f\|_{L^2}}{\|f\|_G}$ which actually can be considered as a definition of scale.

Since the L^2-norm is stronger than the G-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. [21, p.32] or [3]) we have for "clean images" $f \in BV(\Omega)$

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

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 6.3. *Let* $f \neq 0$, $\int_\Omega f d\Omega = 0$, $\lambda \|f\|_G \leq 1$ *and let* $t_1 = \frac{1 - \lambda\|f\|_G}{\alpha\lambda\|f\|_G}$. *Then each solution* (u,v) *of* (5.14) *satisfies*

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

Proof. Let us define the following energy $E(t) := J(u) + \frac{\lambda}{2}\|f + v - u\|^2_{L^2}$. Then the flow of u in (5.14) can be viewed at each time point as a steepest descent of this energy.

Using the decomposition result of [21] (p. 32) and the condition $\lambda\|f\|_G \leq 1$ we can verify that the initial condition $u(0) = 0, v(0) = 0$ at time $t = 0$ is a stationary point, where the energy $E(0) = J(0) + \frac{\lambda}{2}\|f\|^2_{L^2}$ 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 $\lambda\|f + v\|_G \leq 1$. 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

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

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

Note that as in the disk example, we can obtain the time of appearance t_1 of the direct solution (stated in Theorem 6.1) by multiplying the expression for t_1 of the relaxed flow by $\alpha\lambda$ and letting $\lambda \to 0$.

7. Numerical results

In this section we present some numerical examples. We show an example of a 1D problem solved by the direct inverse flow (Section 4) and by the relaxed inverse flow (Section 5, Eq. (5.14)) in order to test and compare their behavior. 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.

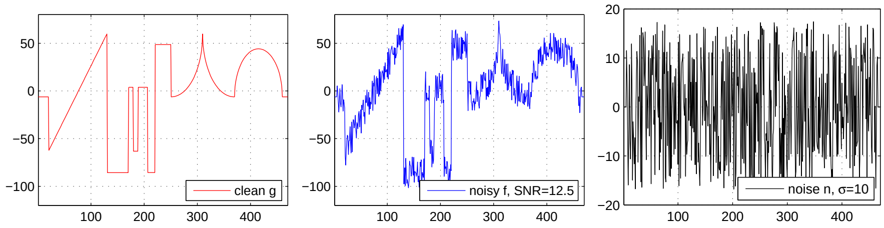


FIG. 7.1. *1D signal denoising. From left to right: clean signal g, noisy signal f ($SNR = 12.5$); Gaussian noise n, $\sigma = \|n\|_{L^2} = 10$.*

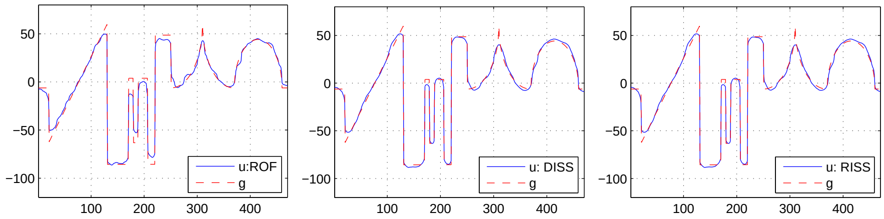


FIG. 7.2. *1D signal denoising. From left to right: restored signal u from ROF, the direct inverse flow (DISS) and the relaxed inverse flow (RISS). SNR: ROF(u)=17.73, DISS(u)=21.94, RISS(u)=21.95.*

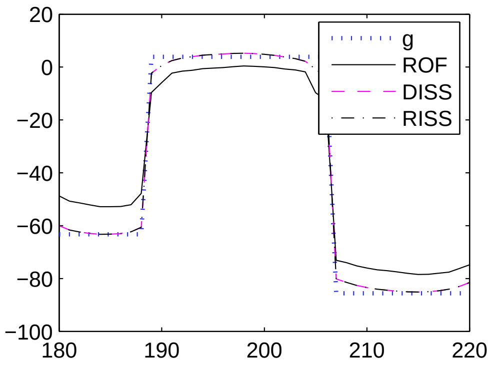


FIG. 7.3. *1D signal denoising. Comparison between the solutions u from ROF, DISS, RISS (part, at $(180, 220)$).*

We focus on denoising by ROF-based flows. A single example of image decomposition by the $TV-L^1$-based flow is also given (Eq. (5.15)). Further generalizations and experiments with other J and H functionals are currently studied and will appear elsewhere.

Example 1 (1D signal): We first consider a 1D denoising problem. Fig. 7.1 shows the clean signal g, the noisy signal f and the Gaussian noise n ($\sigma = \|n\|_{L^2} = 10 \approx \frac{1}{4}\|g\|_{L^2}$). Fig. 7.2 shows the solutions u obtained by ROF, the direct inverse flow (DISS) and the relaxed inverse flow (RISS). Fig. 7.3 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 3.1 and the following discussion) that the inverse TV flow yields better restorations than the original ROF model.

In Fig. 7.3 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.

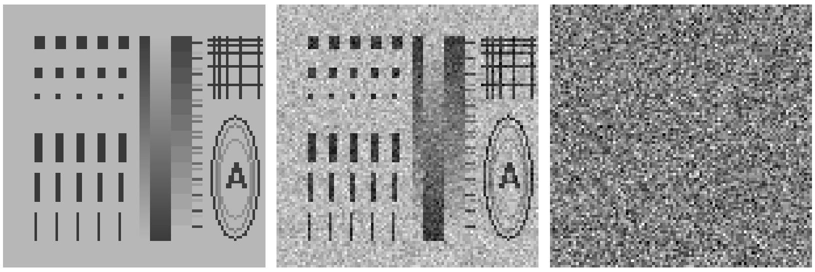


Fig. 7.4. *2D shape image. From left to right: original image g; noisy image f* ($SNR = 7.4$)*; Gaussian noise n* ($\sigma = 40$)*.*

Example 2 (Synthetic test 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 Fig. 7.4. $SNR(f) = 7.4, \sigma = \|f - g\|_{L^2} = 40$. Fig. 7.5 shows the results obtained by ROF, iterated TV refinement (Bregman ROF, cf. [24]) 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. 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. Our proposed method reaches the best SNR results: $SNR(u) = 9.9, 11.8, 12.5$, for ROF, iterated refinements, RISS, respectively.

Natural images are being processed in the next three examples.

Example 3 (Cameraman):. In Fig. 7.6 we compare denoising of the Cameraman image by ROF and RISS. Our proposed method retains great contrast, which is

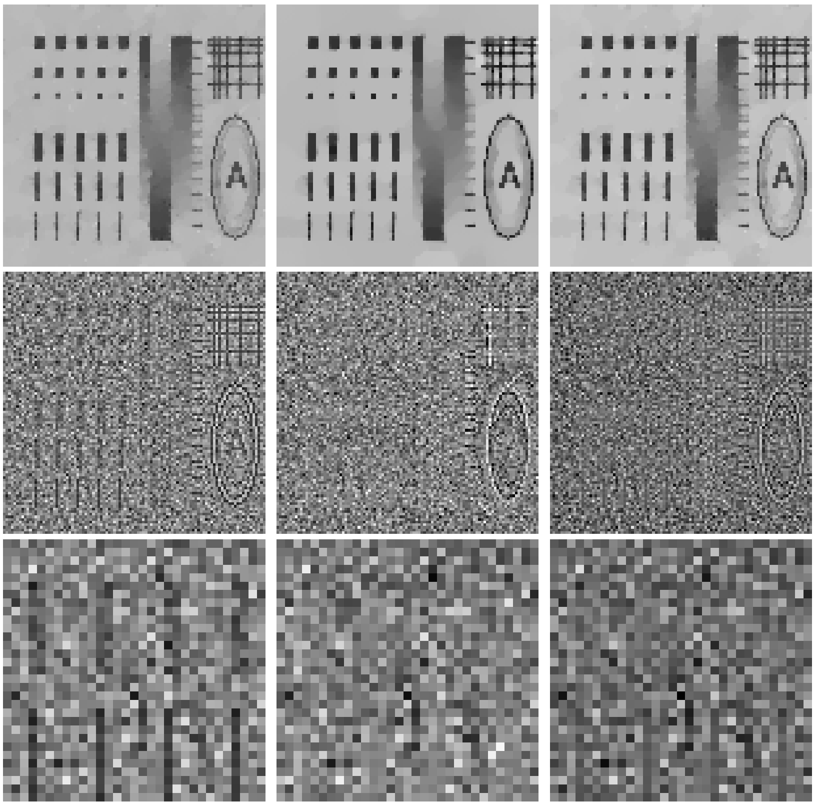


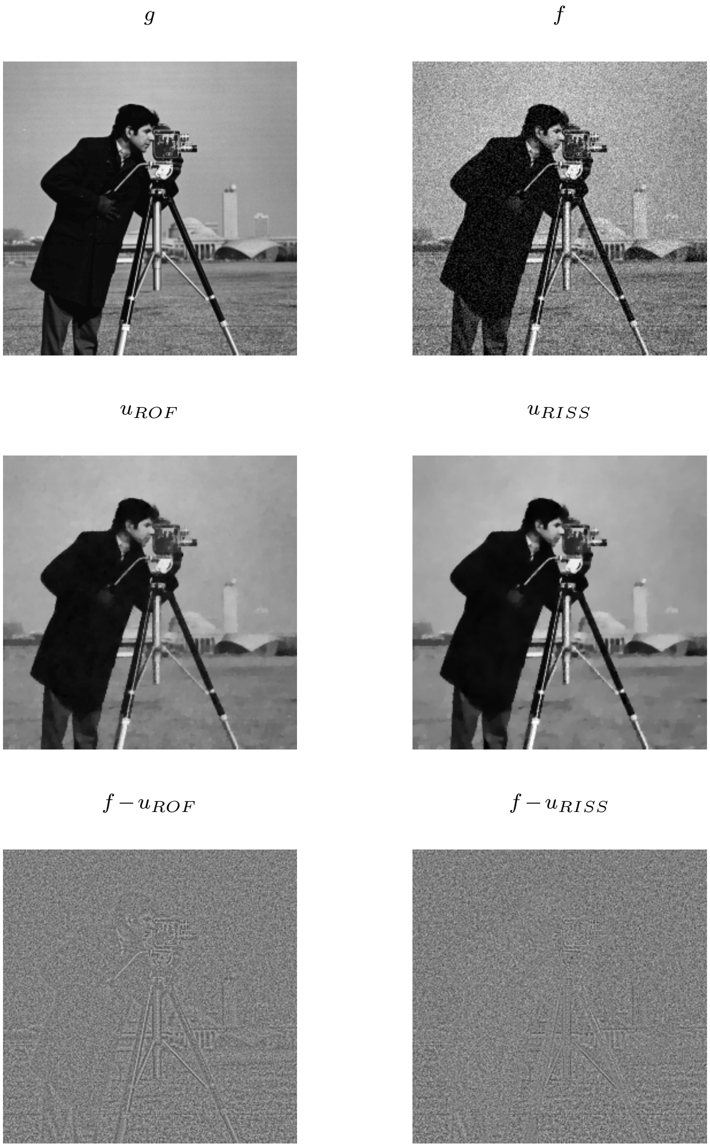
FIG. 7.5. *2D shape image. Denoising results. From top to bottom: denoised u, residual $w = f - u$ and part of w from ROF, Bregman ROF and RISS (column-by-column). SNR: ROF(u)=9.9, Bregman(u)=11.8, RISS(u)=12.5.*

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 Figs. 7.7 and 7.8 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 Fig. 7.9 some more information on the evolution of Examples 3 and 4 is given. Three performance criteria which measure the closeness of u to the clean image g are shown as a function of the evolution time. In the first row the SNR is plotted (which is based on the L^2 distance), the second row depicts the L^1 distance, $\|g - u\|_{L^1}$, and the third row depicts the Bregman distance $D(g,u)$. The SNR and the Bregman distance are both monotonically 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 Fig. 7.10 it is demonstrated that our method can work very well also for decomposition purposes. Eq.



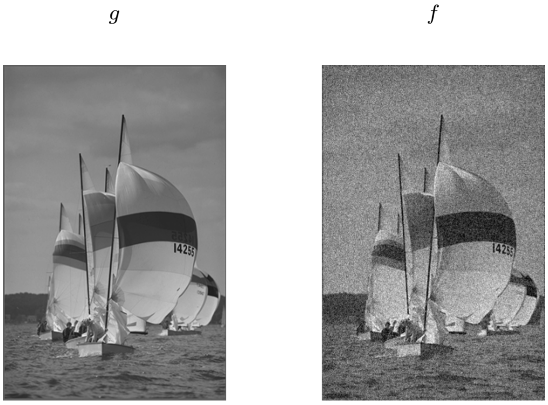
g f

u_{ROF} u_{RISS}

$f-u_{ROF}$ $f-u_{RISS}$

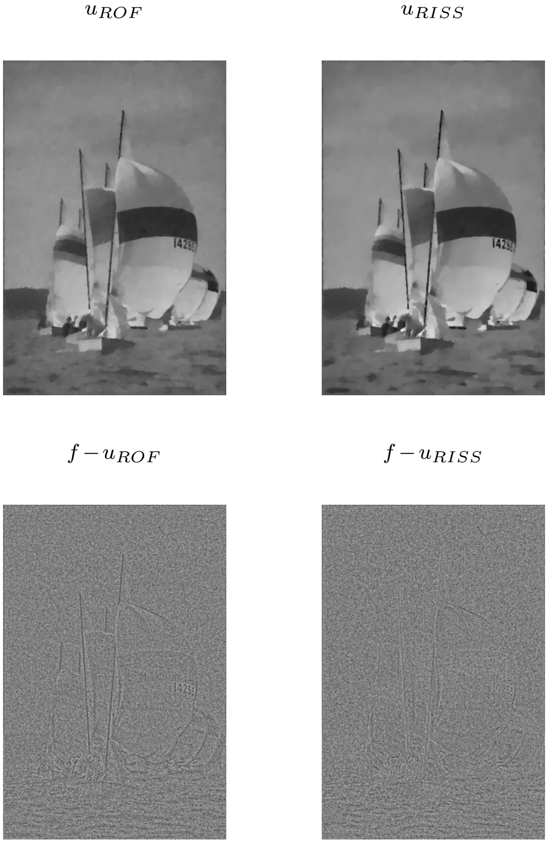
FIG. 7.6. *Cameraman image ($\sigma=20$, SNR=9.89). Top row: clean image g (left), noisy image f (right). Second row: denoised image u by ROF, SNR=15.76 (left) and by the proposed inverse flow, SNR=16.42 (Eq.* (5.14)*, right). Third row: corresponding residual parts $f-u$. [$\lambda=0.01$].*

(5.15) 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 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 published elsewhere.



g f

FIG. 7.7. *Sailboat image* ($\sigma = 20$, *SNR=4.40). Clean image* g *(left), noisy image* f *(right).*



u_{ROF} u_{RISS}

$f - u_{ROF}$ $f - u_{RISS}$

FIG. 7.8. *Sailboat image (cont'). Top row: denoised image* u *by ROF, SNR=11.43 (left) and by the proposed inverse flow, SNR=11.98 (Eq.* (5.14), *right). Bottom row: corresponding residual parts* $f - u$. *[*$\lambda = 0.01$*].*

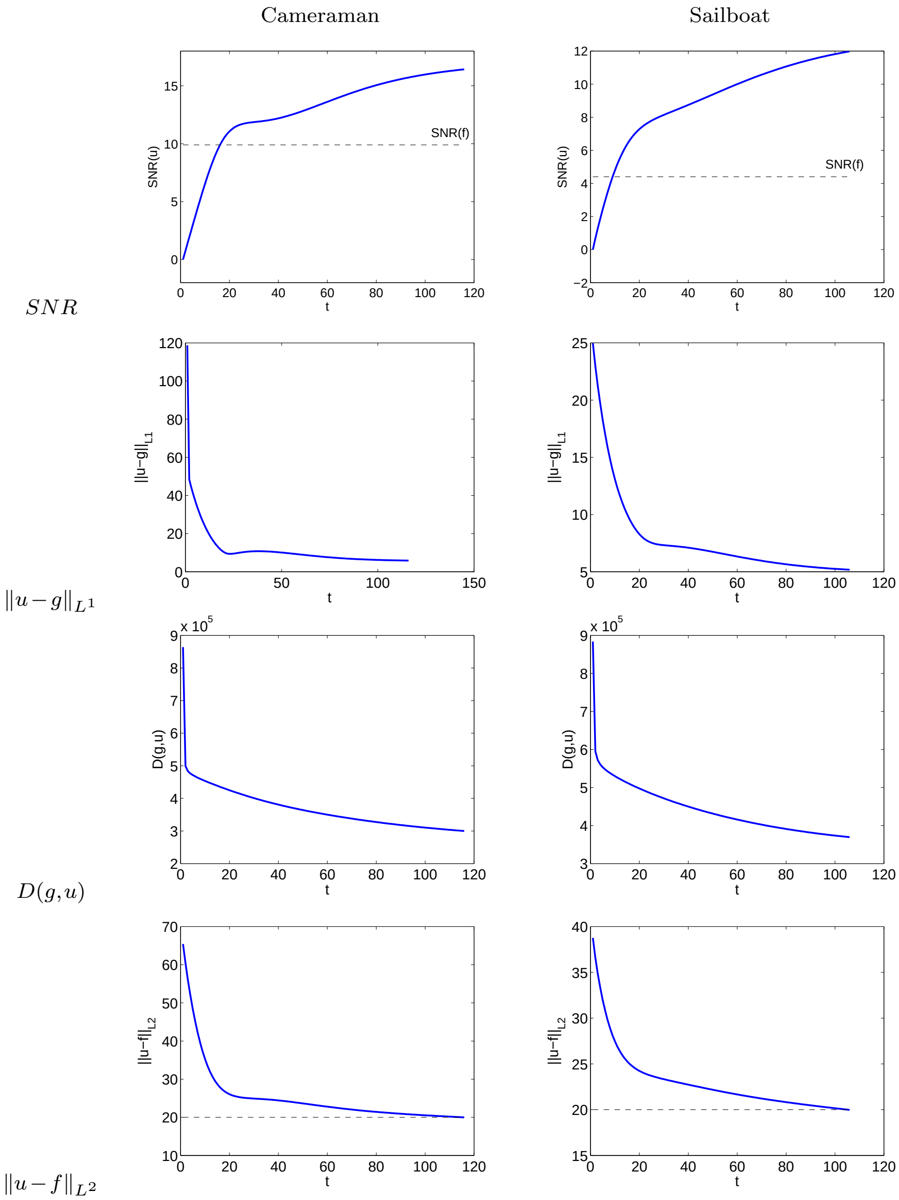


Fig. 7.9. *Performance and convergence plots as a function of time. From top:* $SNR(u)$, $\|u-g\|_{L^1}$, $D(g,u)$, $\|u-f\|_{L^2}$. *Left - Cameraman image, right - Sailboat image.*

8. Discussion and conclusion

Two new types of nonlinear processes are presented in this paper for image simplification and regularization. Both extend the Bregman iterations procedure introduced in [24] 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. [2, 15, 19, 26, 33, 34]): First, the

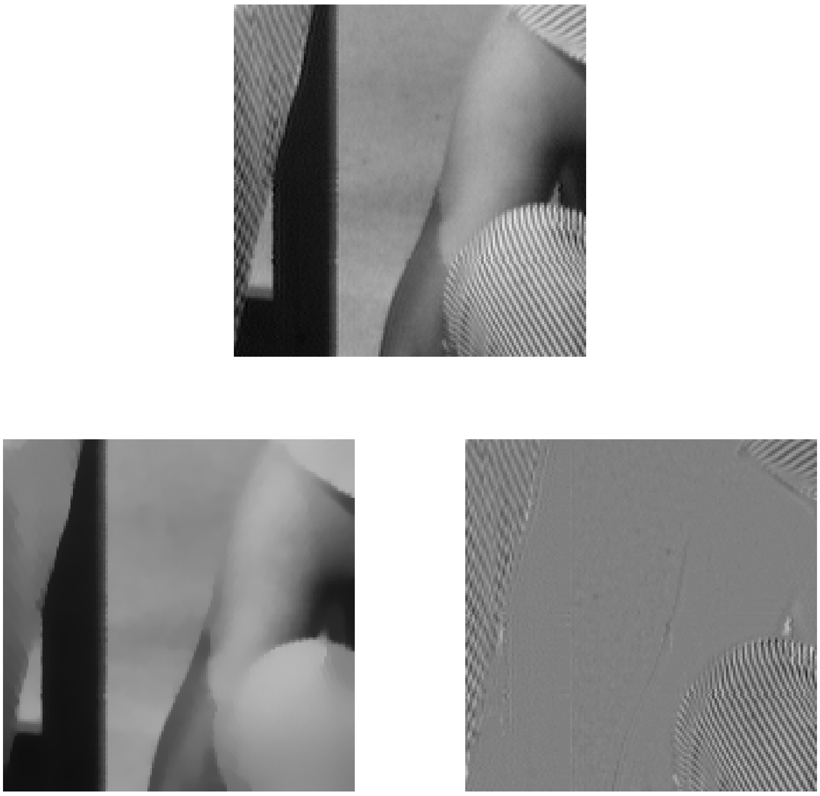


Fig. 7.10. *Decomposition of part of Barbara using* $TV-L1$ *inverse flow, Eq.* (5.15). *Top: original. Bottom: geometric part* u *(left), textural part* $f-u$ *(right).* $[\lambda=0.02]$.

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 $H=\frac{1}{2}\|f-Ku\|_{L^2}^2$) or have a flow based on the L^1 fidelity term for removal of impulsive noise or for structure-texture decomposition [11, 23, 35]. Other types of fidelity terms may be considered in the future, for instance ones based on the G-norm [4, 21, 32], H^{-1} norm [25] or on Gabor functions [5]. 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 colleagues [20]. The flow produces excellent denoising results and retains very good contrast of larger objects (in some cases contrast may be even slightly enhanced). An open question concerning the relaxed flow is whether in some sense u approaches the clean image g (as shown for the direct flow). The numerical indications are promising. We are also investigating other types of inverse scale-space equations.

Acknowledgement. The work of M.B. has been supported by the Austrian National Science Foundation FWF through Project SFB F 013/08 and the Austrian Academy of Sciences ("OAW) through the Johann Radon Institute for Computational and Applied Mathematics (RICAM). The work of G.G, S.O. and J.X. has been supported by the NIH through grant U54RR021813, and the NSF through grants DMS-0312222, ACI-0321917 and DMI-0327077. The authors would like to thank Mark Green and Barry Merriman (both UCLA) for useful and stimulating discussions.

REFERENCES

[1] L. Alvarez, F. Guichard, P-L. Lions and J-M Morel, *Axioms and fundamental equations of image processing*, Arch. Rat. Mech. and Anal., 123(3), 199-257, 1993.

[2] L. Alvarez, P-L. Lions and J-M. Morel, *Image selective smoothing and edge detection by nonlinear diffusion (II)*, SIAM J. Numer. Anal., 29(3), 845-866, 1992.

[3] J-P. Aubin and I. Ekeland, *Applied Nonlinear Analysis*, John Wiley & Sons, New York, 1984.

[4] J-F. Aujol, G. Aubert, L. Blanc-Féraud and A. Chambolle, *Image decomposition into a bounded variation component and an oscillating component*, Journal of Mathematical Imaging and Vision, 22(1), 71-88, 2005.

[5] J-F. Aujol, G. Gilboa, T. Chan and S. Osher, *Structure-texture image decomposition - modeling, algorithms, and parameter selection*, CAM-Report 05-10, Los Angeles, UCLA, CA, 2005, to appear in International Journal of Computer Vision.

[6] G. Bellettini, V. Caselles and M. Novaga, *The total-variation flow in $\mathcal{R}^n$*, J. Differential Equations, 184, 475-525, 2002.

[7] 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.

[8] M. Burger, D. Goldfarb, S. Osher, J. Xu and W. Yin, *Inverse total variation flow*, in preparation.

[9] M. Burger and S. Osher, *Convergence rates of convex variational regularization*, Inverse Problems, 20, 1411-1422, 2004.

[10] T. Chan and J. Shen, *Image Processing and Analysis*, SIAM, Philadelphia, 2005.

[11] T. F. Chan and S. Esedoglu, *Aspects of total variation regularized L^1 function approximation*, CAM-Report 04-07, UCLA, Los Angeles, CA, 2004, to appear in SIAM J. Appl. Math.

[12] G. Chen and M. Teboulle, *Convergence analysis of a proximal-like minimization algorithm using Bregman functions*, SIAM J. Optim., 3, 538-543, 1993.

[13] I. Ekeland and R. Temam, *Convex Analysis and Variational Problems*, North-Holland Publishers, Amsterdam, 1976.

[14] H. W. Engl, M. Hanke and A. Neubauer, *Regularization of Inverse Problems*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1996.

[15] G. Gilboa, N. Sochen and Y-Y. Zeevi, *Image enhancement and denoising by complex diffusion processes*, IEEE Trans. Pattern Analysis and Machine Intelligence, 25(8), 1020-1036, 2004.

[16] G. Gilboa, N. Sochen and Y-Y. Zeevi, *Estimation of optimal PDE-based denoising in the SNR sense*, CAM-Report 05-48, UCLA, Los Angeles, CA, 2005, to appear in IEEE Trans. Image Proc.

[17] C. Groetsch and O. 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.

[18] L. He, A. Marquina and S. Osher, *Blind deconvolution using TV regularization and Bregman iteration*, International Journal of Imaging Systems and Technology, 5, 74-83, 2005.

[19] C. Kimmel, R. Malladi and N. 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.

[20] J. Lie and J-M. Nordbotten, *Inverse scale spaces for nonlinear filtering*, in preparation, 2005.

[21] Y. Meyer, *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations*, AMS, Providence, RI, 2001.

[22] P. Mrázek and M. Navara, *Selection of optimal stopping time for nonlinear diffusion filtering*, International Journal of Computer Vision, 52(2/3), 189-203, 2003.

[23] M. Nikolova, *Weakly constrained minimization. application to the estimation of images and signals involving constant regions*, J. Math. Imaging Vision, 21, 155-175, 2004.

[24] S. Osher, M. Burger, D. Goldfarb, J. Xu and W. Yin, *An iterative regularization method for total variation based image restoration*, Multiscale Model. and Simul., 4, 460-489, 2005.

[25] S. Osher, A. Solé and L. Vese, *Image decomposition and restoration using total variation minimization and the H^{-1} norm*, Multiscale Model. Simul., 1, 349-370, 2003.

[26] P. Perona and J. Malik, *Scale-space and edge detection using anisotropic diffusion*, IEEE Trans. PAMI, 12(7), 629-639, 1990.

[27] R. Plato, *The discrepancy principle for iterative and parametric methods to solve linear ill-posed problems*, Numer. Math., 75, 99-120, 1996.

[28] L. I. Rudin, S. J. Osher and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Phys. D, 60, 259-268, 1992.

[29] O. Scherzer and C. Groetsch, *Inverse scale space theory for inverse problems*, In M.Kerckhove, editor, Scale-Space and Morphology in Computer Vision, Lecture Notes in Comput. Sci., Springer, New York, 2106, 317-325, 2001.

[30] O. Scherzer and J. Weickert, *Relations between regularization and diffusion filtering*, Journal of Mathematical Imaging and Vision, 12, 43-63, 2000.

[31] E. Tadmor, S. Nezzar and L. Vese, *A multiscale image representation using hierarchical (BV, L^2) decompositions*, Multiscale Model. Simul., 2, 554-579, 2004.

[32] L. Vese and S. Osher, *Modeling textures with total variation minimization and oscillatory patterns in image processing*, J. Sci. Comput., 19, 553-572, 2003.

[33] J. Weickert, *Coherence-enhancing diffusion filtering*, International Journal Computer Vision, 31, 111-127, 1999.

[34] A-P. Witkin, *Scale space filtering*, In Proc. Int. Joint Conf. On Artificial Intelligence, 1019-1023, 1983.

[35] W. Yin, D. Goldfarb and S. 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.