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

Jinjun Xu and Stanley Osher

***Abstract*— In this paper we generalize the iterative regularization method and the inverse scale space method, recently developed for total variation-based image restoration, to wavelet-based image restoration. This continues our earlier joint work with others where we applied these techniques to variational based image restoration, obtaining significant improvement over the Rudin-Osher-Fatemi total variation based restoration. Here we apply these techniques to soft shrinkage and obtain the somewhat surprising result that (a) the iterative procedure applied to soft shrinkage gives firm shrinkage and converges to hard shrinkage and (b) that these procedures enhance the noise removal capability both theoretically, in the sense of generalized Bregman distance, and for some examples, experimentally in terms of SNR, leaving less signal in the residual.**

***Index Terms*— Image restoration, iterative regularization method, inverse scale space methods, total variation, Bregman distance, wavelet, wavelet shrinkage.**

I. Introduction

TOTAL variation (TV) regularization (cf., e.g., [1], [2]) and wavelet shrinkage (cf., e.g., [3]–[5]) are among the most useful techniques for signal and image denoising. The relations between them have been studied by several authors, e.g., in [6]–[8].

In recent work, the authors and colleagues developed an iterative regularization method (IRM) and applied it to variational-based image restoration [9]. Significant improvements were obtained in both theoretical and numerical results. Later in [10], [11] the discrete refinement procedure in this method was successfully generalized to a time-continuous inverse scale space (ISS) formulation.

In this paper, we start by reviewing TV-based denoising methods and their recent developments, IRM and ISS, and the relation between TV regularization and one important technique in wavelet based denoising - soft shrinkage (cf. [4]). Then we generalize the iterative regularization idea to the latter and obtain another type of wavelet shrinkage - firm shrinkage (cf. [12]). After taking a limit of the discrete iteration, we show that iterated soft shrinkage becomes hard shrinkage (cf. [4]). Finally we generalize the inverse scale space idea to soft shrinkage and present some numerical examples.

EDICS: RST-DNOI Denoising

Preprint. Submitted to IEEE Trans. Image Proc. on March 13, 2006, revised on June 26, 2006. This work was partially supported by NSF grants DMS-0312222, ACI-0321917 and DMI-0327077, and NIH grant U54RR021813.

J. Xu and S. Osher are with the Department of Mathematics, University of California, Los Angeles, Box 951555, Los Angeles, CA 90095-1555, USA. email: {jjxu, sjo}@math.ucla.edu. Phone: (+1)3108251758. Fax: (+1)3102066673.

II. Total Variation Based Methods, Iterative Regularization and Inverse Scale Space

Given a noisy signal (one-dimensional, 1D, $d = 1$) or image (two-dimensional, 2D, $d = 2$) data, denoted by $f : \Omega \subset \mathbb{R}^d \to \mathbb{R}$, which is corrupted by additive noise from the unknown original data g, the task of denoising is to look for a signal or image u which is close to f and "clean" in some sense. Without confusion, we will use the word "image" in all dimensions.

Variational methods solve this problem via the following minimization

$$u^* = \operatorname*{argmin}_u \Big\{ J(u) + \lambda H(u, f) \Big\}, \tag{1}$$

where $J(u)$ is a regularization term which characterizes some features of the desired solution u, $H(u, f)$ is a fidelity term which measures the difference between u and f and usually is a nonnegative functional, and $\lambda > 0$ is a scale parameter tuning the weight between the regularization term and the fidelity term. TV-based methods use $J(u)$ as the BV-seminorm of u:

$$J(u) = |u|_{BV(\Omega)} = \int_\Omega |\nabla u| \, dxdy, \tag{2}$$

which is also referred to as the *total variation* of u, and then look for solutions u in $BV(\Omega)$ (cf., eg., [13, Chapter 5]). One important feature of using the BV-seminorm as a regularization term is that it helps to recover clean functions u having sharp edges. TV-based methods were introduced to image processing by Rudin-Osher-Fatemi [1], with the following ROF denoising model:

$$u^* = \operatorname*{arg\,min}_{u \in BV(\Omega)} \left\{ |u|_{BV(\Omega)} + \frac{\lambda}{2} \|f - u\|^2_{L^2} \right\}. \tag{3}$$

The solutions of (1) and (3) are often obtained by solving the corresponding Euler-Lagrange equations through, e.g., gradient descent or a fixed point method.

Due to its simple formula and edge preserving property, ROF is one of the most popular TV-based image denoising techniques. A defect involving loss of contrast was observed and analyzed by, e.g., [14], [15]. To solve this problem, in [9] the authors and colleagues developed an iterative regularization method (IRM), which replaces the regularization term in (1) by a generalized Bregman distance. For $p \in \partial J(w)$, which is the subgradient of the weakly convex functional J at w (cf., e.g., [16]), we define the (nonnegative) quantity

$$D^p_J(u, w) \equiv J(u) - J(w) - \langle u - w, p \rangle, \tag{4}$$

which is known as the generalized Bregman distance asso-

ciated with $J(\cdot)$ and p (cf. [17]–[19] for an extension to nonsmooth functionals J). $\langle\cdot,\cdot\rangle$ is the usual L^2 inner product on Ω. We note that if $J(\cdot)$ is not strictly convex, then $\partial J(\cdot)$ may contain more than one element. For a uniquely selected $p \in \partial J(\cdot)$, which is the case for our iterative regularization procedure defined in [9] and also in the following sections of this paper, (4) is well-defined.

To summarize, instead of solving (1), IRM in [9] solves the following sequence of variational problems

$$u^{(k)} = \operatorname*{argmin}_u \left\{ D_J^{p^{(k-1)}}(u, u^{(k-1)}) + \lambda H(u^{(k-1)}, f) \right\}, \quad (5)$$

$$p^{(k)} = p^{(k-1)} - \lambda \partial_u H(u^{(k-1)}, f) \in \partial J(u^{(k)}), \quad (6)$$

where $k \geq 1, u^{(0)} = p^{(0)} = 0$. The sequence $\{u^{(k)}\}$ satisfies: $H(u^{(k)}, f)$ monotonically decreases with respect to k; $D(g, u^{(k)})$ monotonically decreases as long as $H(u^{(k)}, f) \geq H(g, f)$, where g is the unknown clean image. From these two facts, if we have the value or an estimate of $H(g, f)$ (which is usually related to the noise level), then the iteration (5) and (6) will be stopped at the last k such that $H(u^{(k)}, f) \geq H(g, f)$. The numerical examples in [9] show significant improvements of IRM over standard models (1). For rigorous analysis and details, see [9].

Later in [10], [11], the authors and colleagues generalized the discrete IRM to a time-continuous nonlinear inverse scale space (ISS) flow. The idea can be briefly described as follows: rewrite (6) as

$$\frac{p^{(k)} - p^{(k-1)}}{\lambda} = -\partial_u H(u^{(k-1)}, f), \quad (7)$$

then interprete λ as a timestep Δt and let it go to 0, define $p^{(k)} = p(k\Delta t), u^{(k)} = u(k\Delta t)$. After dropping the superscript k and $k-1$ and letting $\Delta t \to 0, k\Delta t \to t$ we obtain

$$\frac{dp}{dt} = -\partial_u H(u, f), \quad (8)$$

with $u(0) = p(0) = 0$ and $u \in \partial J^*(p)$, where J^* is the dual functional of J (cf., e.g., [16]). We obtained similar numerical properties for $u(t)$ in ISS as for $u^{(k)}$ in IRM. Therefore the flow can be stopped at $\bar{t}$ when $H(u(\bar{t}), f) = H(g, f)$. For details see [10], [11].

III. Wavelet Shrinkage and the Relation with TV Regularization

Given an orthonormal wavelet basis $\{\psi_{\mathbf{j}}(x)\}$, $\mathbf{j} = (j_1, j_2, j_3)$, which is generated, e.g., by $\{\psi^{(j_3)}(x)\}_{j_3=1}^{2^d-1}$ with $\psi_{\mathbf{j}}(x) = 2^{j_1}\psi^{(j_3)}(2^{j_1}x - j_2)$, $j_1 \in \mathbb{Z}, j_2 \in \mathbb{Z}^d, x \in \mathbb{R}^d$, the wavelet transform of an image f can be represented as (cf., e.g., [20]–[22]):

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

We denote $\tilde{f} = \{\tilde{f}_{\mathbf{j}}\} = \{\langle f, \psi_{\mathbf{j}}\rangle\}$.

In general, wavelet shrinkage attempts to denoise images via the following three steps (cf., e.g., [3], [4]):

(1) *Analysis*. Transform the noisy image f to the wavelet coefficients $\tilde{f} = \{\tilde{f}_{\mathbf{j}}\}$;
(2) *Shrinkage*. Apply a shrinkage operator $\mathcal{T}$ with a threshold parameter τ related to the noise level to the wavelet coefficient $\tilde{f}$: $\tilde{u} := \mathcal{T}_\tau(\tilde{f})$;
(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 (e.g., [7]) the wavelet basis is often divided into two parts: lowpass scaling functions $\varphi(x)$ and bandpass wavelet functions $\psi(x)$. 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 either on all wavelet coefficients or only on the detail coefficients. In our discussion in this paper we will consider shrinkage on all the wavelet coefficients. Since our models are separable, our discussion can be easily generalized to the case of shrinkage on detail coefficients only.*

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

- Soft shrinkage (cf. [4]), for $\tau > 0$,

$$\mathcal{S}_\tau(w) = \begin{cases} w - \tau \text{sign}(w), & \text{if } |w| > \tau, \\ 0, & \text{if } |w| \leq \tau. \end{cases} \quad (9)$$

- Hard shrinkage (cf. [4]), for $\tau > 0$,

$$\mathcal{H}_\tau(w) = \begin{cases} w, & \text{if } |w| > \tau, \\ 0, & \text{if } |w| \leq \tau. \end{cases} \quad (10)$$

- Firm shrinkage (cf. [12]), for $\tau_2 > \tau_1 > 0$,

$$\mathcal{F}_{\tau_1,\tau_2}(w) = \begin{cases} w, & \text{if } |w| > \tau_2, \\ c(w, \tau_1, \tau_2), & \text{if } \tau_1 < |w| \leq \tau_2, \\ 0, & \text{if } |w| \leq \tau_1, \end{cases} \quad (11)$$

where $c(w, \tau_1, \tau_2) = \frac{\tau_2}{\tau_2 - \tau_1}(w - \tau_1 \text{sign}(w))$ is a value between 0 and w.

Now we consider the Besov space $B_1^{1,1}(\Omega)$, which contains, roughly speaking, functions with first order derivatives in $L^1(\Omega)$. (For the formal definition of Besov spaces $B_\alpha^{p,q}$, cf., e.g., [23]). One important fact is that the discrete l^1-norm of the wavelet coefficients is equivalent to the norm in $B_1^{1,1}(\Omega)$, which is a subset of $BV(\Omega)$ for $\Omega \subset \mathbb{R}^2$ (cf., e.g., [6], [8], [24]–[27]). We replace the BV-seminorm (2) by the l^1-norm of wavelet coefficients

$$J(\tilde{u}) = \sum_j |\tilde{u}_{\mathbf{j}}| \approx \|u\|_{B_1^{1,1}}. \quad (12)$$

Here '$\approx$' is used to represent the equivalence between the two norms. (We still use the notation J although it has different meaning now.) From Parseval's identity we have

$$\|f - u\|_{L^2}^2 = \|\tilde{f} - \tilde{u}\|_{L^2}^2 = \sum_{\mathbf{j}} |\tilde{f}_{\mathbf{j}} - \tilde{u}_{\mathbf{j}}|^2. \quad (13)$$

We then approximate the TV-based ROF model (3) by using

the following wavelet-based method:

$$\begin{aligned}\tilde{u}^* &= \underset{\tilde{u}}{\operatorname{argmin}}\left\{ J(\tilde{u}) + \frac{\lambda}{2}\|\tilde{f} - \tilde{u}\|_{L^2}^2 \right\} \\ &= \underset{\tilde{u}}{\operatorname{argmin}}\left\{ \sum_{\mathbf{j}} |\tilde{u}_{\mathbf{j}}| + \frac{\lambda}{2}\sum_{\mathbf{j}} |\tilde{f}_{\mathbf{j}} - \tilde{u}_{\mathbf{j}}|^2 \right\}, \quad (14)\end{aligned}$$

where $\tilde{f}_{\mathbf{j}}$ are the wavelet coefficients of the noisy image f and the restored image u^* is the wavelet reconstruction of $\tilde{u}^*$. To simplify the notation from now on we drop the superscript $*$ if not otherwise specified.

Note that because the summation in (14) is separable, it suffices to solve a sequence of scalar minimization problems $\min_{\tilde{u}_{\mathbf{j}}} \phi_{\tilde{f}_{\mathbf{j}}}(\tilde{u}_{\mathbf{j}})$ for all $\mathbf{j}$, where $\phi_{\tilde{f}_{\mathbf{j}}}(\tilde{u}_{\mathbf{j}}) = |\tilde{u}_{\mathbf{j}}| + \frac{\lambda}{2}(\tilde{f}_{\mathbf{j}} - \tilde{u}_{\mathbf{j}})^2$.

The minimizer of (14) is

$$\tilde{u}_{\mathbf{j}} = \begin{cases} \tilde{f}_{\mathbf{j}} - \frac{1}{\lambda}\operatorname{sign}(\tilde{f}_{\mathbf{j}}), & \text{if } |\tilde{f}_{\mathbf{j}}| > \frac{1}{\lambda}, \\ 0, & \text{if } |\tilde{f}_{\mathbf{j}}| \le \frac{1}{\lambda}, \end{cases} \quad (15)$$

for all $\mathbf{j}$, which is precisely the soft shrinkage algorithm (9) with threshold $\tau = \frac{1}{\lambda}$:

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

The above connection between ROF and soft shrinkage was observed by Chambolle et. al. in [6].

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

$$\partial F(\tilde{u}_{\mathbf{j}}) = \begin{cases} sign(\tilde{u}_{\mathbf{j}}), & if\ \ \tilde{u}_{\mathbf{j}} \neq 0, \\ [-1, 1], & if\ \ \tilde{u}_{\mathbf{j}} = 0. \end{cases} \quad (16)$$

The Euler-Lagrange equation of (14) *is*

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

Denote $\tilde{p}_{\mathbf{j}} = \lambda(\tilde{f}_{\mathbf{j}} - \tilde{u}_{\mathbf{j}}) \in \partial F(\tilde{u}_{\mathbf{j}})$, $\tilde{v}_{\mathbf{j}} = \frac{\tilde{p}_{\mathbf{j}}}{\lambda}$, then $\tilde{v}_{\mathbf{j}} = \tilde{f}_{\mathbf{j}} - \tilde{u}_{\mathbf{j}}$ and we have a decomposition $\tilde{f}_{\mathbf{j}} = \tilde{u}_{\mathbf{j}} + \tilde{v}_{\mathbf{j}}$. From (15) *we have*

$$\tilde{p}_{\mathbf{j}} = \begin{cases} sign(\tilde{f}_{\mathbf{j}}), & if\ |\tilde{f}_{\mathbf{j}}| > \frac{1}{\lambda}, \\ \lambda\tilde{f}_{\mathbf{j}}, & if\ |\tilde{f}_{\mathbf{j}}| \le \frac{1}{\lambda}, \end{cases} \quad (17)$$

and

$$\tilde{v}_{\mathbf{j}} = \begin{cases} \frac{1}{\lambda}sign(\tilde{f}_{\mathbf{j}}), & if\ |\tilde{f}_{\mathbf{j}}| > \frac{1}{\lambda}, \\ \tilde{f}_{\mathbf{j}}, & if\ |\tilde{f}_{\mathbf{j}}| \le \frac{1}{\lambda}, \end{cases} \quad (18)$$

for all $\mathbf{j}$. Note that although $\partial F(0) = [-1, 1]$ is a multivalued set, $\tilde{p}_{\mathbf{j}}$ defined above is unique. We will use $\tilde{p}_{\mathbf{j}}$ and $\tilde{v}_{\mathbf{j}}$ later to define the iterative regularization procedure.

IV. Iterative Regularization Applied to Wavelet Shrinkage

Now we generalize IRM and ISS in [9]–[11] to wavelet shrinkage. The generalized Bregman distance associated with $J(\tilde{u})$ in (12) and $\tilde{p} \in \partial J(\tilde{w})$ can be defined as

$$D_J^{\tilde{p}}(\tilde{u}, \tilde{w}) \equiv J(\tilde{u}) - J(\tilde{w}) - \langle \tilde{u} - \tilde{w}, \tilde{p} \rangle. \quad (19)$$

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 $\tilde{p} \in \partial J(\tilde{w})$. Without confusion, we will omit the word "generalized" in the following discussion.

Following the same idea as in [9], we replace $J(\tilde{u})$ in (14) by the Bregman distance (19) 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

$$\tilde{u}^{(k)} = \underset{\tilde{u}}{\operatorname{argmin}}\left\{ D_J^{\tilde{p}^{(k-1)}}(\tilde{u}, \tilde{u}^{(k-1)}) + \frac{\lambda}{2}\|\tilde{f} - \tilde{u}\|_{L^2}^2 \right\}, \quad (20)$$

$$\tilde{p}^{(k)} = \tilde{p}^{(k-1)} + \lambda(\tilde{f} - \tilde{u}^{(k)}), \quad (21)$$

with $k \ge 1, \tilde{u}^{(0)} = 0, \tilde{p}^{(0)} = 0$. We shall show later that such $\tilde{p}^{(k)} \in \partial J(\tilde{u}^{(k)})$. We call this as a wavelet-based iterative regularization method (W-IRM).

If we denote $\tilde{v}^{(k)} = \frac{\tilde{p}^{(k)}}{\lambda}$, then $\tilde{v}^{(0)} = 0$, by plugging (19) into (20) and dropping the constant terms from the minimization, after some simplification we can rewrite (20) as

$$\tilde{u}^{(k)} = \underset{\tilde{u}}{\operatorname{argmin}}\left\{ J(\tilde{u}) + \frac{\lambda}{2}\left\|(\tilde{f} + \tilde{v}^{(k-1)}) - \tilde{u}\right\|_{L^2}^2 \right\}. \quad (22)$$

Note that at the k^{th} iteration we simply replace the wavelet coefficients $\tilde{f}$ in the original minimization (14) by $\tilde{f} + \tilde{v}^{(k-1)}$, and proceed to solve the same minimization procedure as for (14). Therefore the minimizer of (22) is

$$\tilde{u}_{\mathbf{j}}^{(k)} = \begin{cases} (\tilde{f}_{\mathbf{j}} + \tilde{v}_{\mathbf{j}}^{(k-1)}) - \frac{1}{\lambda}\operatorname{sign}(\tilde{f}_{\mathbf{j}} + \tilde{v}_{\mathbf{j}}^{(k-1)}), \\ \qquad\qquad \text{if } |\tilde{f}_{\mathbf{j}} + \tilde{v}_{\mathbf{j}}^{(k-1)}| > \frac{1}{\lambda}, \\ 0, \\ \qquad\qquad \text{if } |\tilde{f}_{\mathbf{j}} + \tilde{v}_{\mathbf{j}}^{(k-1)}| \le \frac{1}{\lambda}, \end{cases} \quad (23)$$

or simply,

$$\tilde{u}_{\mathbf{j}}^{(k)} = \mathcal{S}_{\frac{1}{\lambda}}(\tilde{f}_{\mathbf{j}} + \tilde{v}_{\mathbf{j}}^{(k-1)}), \quad (24)$$

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

$$\tilde{v}_{\mathbf{j}}^{(k)} = \tilde{f}_{\mathbf{j}} + \tilde{v}_{\mathbf{j}}^{(k-1)} - \tilde{u}_{\mathbf{j}}^{(k)}. \quad (25)$$

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

Theorem 1. *For the solutions $\tilde{u}_{\mathbf{j}}^{(k)}$ and $\tilde{v}_{\mathbf{j}}^{(k)}$ defined in the updates* (23) *and* (25)*, $k \ge 1$, we have*

(1)

$$\tilde{v}_{\mathbf{j}}^{(k)} = \begin{cases} \frac{1}{\lambda}sign(\tilde{f}_{\mathbf{j}}), & if\ \ |\tilde{f}_{\mathbf{j}}| > \frac{1}{k\lambda}, \\ k\tilde{f}_{\mathbf{j}}, & if\ \ |\tilde{f}_{\mathbf{j}}| \le \frac{1}{k\lambda}, \end{cases} \quad (26)$$

and $sign(\tilde{v}_{\mathbf{j}}^{(k)}) \equiv sign(\tilde{f}_{\mathbf{j}})$;

(2)

$$\tilde{u}_{\mathbf{j}}^{(k)} = \begin{cases} \tilde{f}_{\mathbf{j}}, & if\ \ |\tilde{f}_{\mathbf{j}}| > \frac{1}{(k-1)\lambda}, \\ k\tilde{f}_{\mathbf{j}} - \frac{1}{\lambda}sign(\tilde{f}_{\mathbf{j}}), & if\ \ \frac{1}{k\lambda} < |\tilde{f}_{\mathbf{j}}| \le \frac{1}{(k-1)\lambda}, \\ 0, & if\ \ |\tilde{f}_{\mathbf{j}}| \le \frac{1}{k\lambda}, \end{cases} \quad (27)$$

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

(3) *for $\tilde{p}_{\mathbf{j}}^{(k)} \equiv \lambda\tilde{v}_{\mathbf{j}}^{(k)}$, $\tilde{p}_{\mathbf{j}}^{(k)} \in \partial F(\tilde{u}_{\mathbf{j}}^{(k)})$.*

The proof of Theorem 1 will be shown in the appendix.

We note that (27) is firm shrinkage (11) (cf. [12]) with thresholds $\tau^{(k)} = \frac{1}{k\lambda}$ and $\tau^{(k-1)} = \frac{1}{(k-1)\lambda}$. In other words, iterative soft shrinkage gives firm shrinkage! We 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 L^2 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 (20) and (21) 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 standard deviation 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 [11], in typical imaging situations, an estimate for the standard deviation or variance of the noise is known, which yields a bound of the form $\|f - g\|_{L^2} \leq \sigma$.

A. 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 turn to compute the Bregman distance (19) between the wavelet coefficients $\tilde{g}$ and $\tilde{u}^{(k)}$. Using $\tilde{p}^{(k)} = \lambda\tilde{v}^{(k)}$ and (26), we have

$$\begin{aligned} D_J^{\tilde{p}^{(k)}}(\tilde{g}, \tilde{u}^{(k)}) &= J(\tilde{g}) - J(\tilde{u}^{(k)}) - \langle \tilde{g} - \tilde{u}^{(k)}, \tilde{p}^{(k)} \rangle \\ &= J(\tilde{g}) - \langle \tilde{g}, \tilde{p}^{(k)} \rangle = \sum_{\mathbf{j}} |\tilde{g}_{\mathbf{j}}| - \sum_{\mathbf{j}} \tilde{p}_{\mathbf{j}}^{(k)} \tilde{g}_{\mathbf{j}} \\ &= \sum_{\mathbf{j}:|\tilde{f}_{\mathbf{j}}|>1/k\lambda} (|\tilde{g}_{\mathbf{j}}| - \text{sign}(\tilde{f}_{\mathbf{j}})\tilde{g}_{\mathbf{j}}) \\ &\quad + \sum_{\mathbf{j}:|\tilde{f}_{\mathbf{j}}|\leq 1/k\lambda} (|\tilde{g}_{\mathbf{j}}| - k\lambda \tilde{f}_{\mathbf{j}}\tilde{g}_{\mathbf{j}}) \qquad (28) \\ &\geq 0. \end{aligned}$$

And we also have

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

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

Therefore, the Bregman distance (19) monotonically decrease for k less than $\bar{k}$, which is the last iterate such that $\|\tilde{f} - \tilde{u}^{(k)}\|_{L^2} \geq \sigma$ and is also our stopping point as mentioned above. In other words, $\tilde{u}^{(k)}$ monotonically converges to $\tilde{g}$ in the sense of Bregman distance when $k \leq \bar{k}$. This unsurprisingly is the same conclusion that we obtained in the iterative regularization procedure for TV-based models, see [9].

B. Limiting case

If we reinterpret $\lambda = \Delta t$ as a timestep and $k\lambda = t^{(k)}$, dropping superscript k then (27) becomes

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

Let $\Delta t \searrow 0$, then $t - \Delta t \to t$. We have the following solution

$$\tilde{u}_{\mathbf{j}}(t) = \begin{cases} \tilde{f}_{\mathbf{j}}, & |\tilde{f}_{\mathbf{j}}| \geq \frac{1}{t}, \\ 0, & |\tilde{f}_{\mathbf{j}}| < \frac{1}{t}, \end{cases} \qquad (29)$$

which turns out to be hard shrinkage (10) with threshold $\tau = \frac{1}{t}$:

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

The Bregman distance $D_J^{\tilde{p}(t)}(\tilde{g}, \tilde{u}(t))$ is same as the one stated in (28), with $k\lambda$ replaced by t, i.e.,

$$D_J^{\tilde{p}(t)}(\tilde{g}, \tilde{u}(t)) = \sum_{\mathbf{j}:|\tilde{f}_{\mathbf{j}}|>1/t} (|\tilde{g}_{\mathbf{j}}| - \text{sign}(\tilde{f}_{\mathbf{j}})\tilde{g}_{\mathbf{j}}) - \sum_{\mathbf{j}:|\tilde{f}_{\mathbf{j}}|\leq 1/t} (|\tilde{g}_{\mathbf{j}}| - t\tilde{f}_{\mathbf{j}}\tilde{g}_{\mathbf{j}}). \qquad (30)$$

So we have the result that $D_J^{\tilde{p}(t)}(\tilde{g}, \tilde{u}(t))$ monotonically decrease in time as long as $\|\tilde{f} - \tilde{u}(t)\|_{L^2} > \sigma$.

We point out here that the idea of reinterpreting λ as a shrinking time-step Δt was used in [10] to define an inverse scale space model. Therefore, the above formula (29) can also be viewed as a solution to the inverse scale space model derived from wavelet denoising (14). We will introduce another version of inverse scale space below involving a regularized $J_\epsilon(\tilde{u})$.

V. 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 a standard regularization technique from the TV-based imaging community: we approximate $F(\tilde{u}_{\mathbf{j}}) = |\tilde{u}_{\mathbf{j}}|$ as

$$F_\epsilon(\tilde{u}_{\mathbf{j}}) = \sqrt{\tilde{u}_{\mathbf{j}}^2 + \epsilon}, \qquad (31)$$

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

$$\tilde{p}_{\mathbf{j}} = \partial F_\epsilon(\tilde{u}_{\mathbf{j}}) = \frac{\tilde{u}_{\mathbf{j}}}{\sqrt{\tilde{u}_{\mathbf{j}}^2 + \epsilon}} \qquad (32)$$

is well-defined and unique everywhere, 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 (14), we have

$$\tilde{u}_{\mathbf{j}} = \underset{\tilde{u}_{\mathbf{j}}}{\text{argmin}} \left\{ \sqrt{\tilde{u}_{\mathbf{j}}^2 + \epsilon} + \frac{\lambda}{2}(\tilde{f}_{\mathbf{j}} - \tilde{u}_{\mathbf{j}})^2 \right\}, \quad \forall \mathbf{j}. \qquad (33)$$

The corresponding Euler-Lagrange equation now is

$$\frac{\tilde{u}_{\mathbf{j}}}{\sqrt{\tilde{u}_{\mathbf{j}}^2 + \epsilon}} + \lambda(\tilde{u}_{\mathbf{j}} - \tilde{f}_{\mathbf{j}}) = 0. \qquad (34)$$

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 (27).

Remark. *Note that when $\epsilon \to 0$, $F_\epsilon(\tilde{u}_\mathbf{j}) \to F(\tilde{u}_\mathbf{j})$, and for $\tilde{u}_\mathbf{j} \neq 0$ we also have $\partial F_\epsilon(\tilde{u}_\mathbf{j}) \to \partial F(\tilde{u}_\mathbf{j})$. When $\tilde{u}_\mathbf{j} = 0$, $\partial F_\epsilon(0) = 0$, while the original iteration* (21) *gives $\tilde{p}_\mathbf{j}^{(k)}(\tilde{u}_\mathbf{j}^{(k)} = 0) = k\lambda\tilde{f}_\mathbf{j} \in \partial F(0) = [-1, 1]$, which is not identically 0. However, $\tilde{p}_\mathbf{j}$ defined in* (32) *still satisfies $\tilde{p}_\mathbf{j}(\tilde{u}_\mathbf{j} = 0) = \partial F_\epsilon(\tilde{u}_\mathbf{j} = 0) \in [-1, 1] = \partial F(0)$.*

Remark. *The regularization technique used in the approximation* (31) *is similar to the approximation of the BV-seminorm term in many TV-based models (cf., e.g., [1], [15]). The reason we use this regularization here is that for inverse scale space discussed later, we need to invert $\tilde{u}$ from $\tilde{p}$, which requires that $\tilde{p}(\tilde{u})$ to be uniquely determined. The role and choice of the parameter ϵ will be studied in our future work.*

A. Inverse scale space

Now we define an inverse scale space with the above regularized $F_\epsilon(\tilde{u}_\mathbf{j})$. We start from the Bregman iteration involving the dual variable $\tilde{p}_\mathbf{j}$ in (21). For each $\mathbf{j}$ we have,

$$\frac{\tilde{p}_\mathbf{j}^{(k)} - \tilde{p}_\mathbf{j}^{(k-1)}}{\lambda} = (\tilde{f}_\mathbf{j} - \tilde{u}_\mathbf{j}^{(k)}), \quad k \geq 1 \tag{35}$$

$$\tilde{u}_\mathbf{j}^{(0)} = \tilde{p}_\mathbf{j}^{(0)} = 0. \tag{36}$$

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

$$\frac{d\tilde{p}_\mathbf{j}}{dt} = \tilde{f}_\mathbf{j} - \tilde{u}_\mathbf{j}, \quad \tilde{u}_\mathbf{j}(0) = 0. \tag{37}$$

Since $\frac{d\tilde{p}_\mathbf{j}}{d\tilde{u}_\mathbf{j}} = \frac{\epsilon}{(\tilde{u}_\mathbf{j}^2+\epsilon)^{3/2}}$, we have a wavelet-based inverse scale space (W-ISS) flow for each $\tilde{u}_\mathbf{j}$ as follows

$$\frac{d\tilde{u}_\mathbf{j}}{dt} = \frac{(\tilde{u}_\mathbf{j}^2 + \epsilon)^{3/2}}{\epsilon}(\tilde{f}_\mathbf{j} - \tilde{u}_\mathbf{j}), \quad \tilde{u}_\mathbf{j}(0) = 0. \tag{38}$$

This gives us a simple flow involving $\tilde{u}_\mathbf{j}$, instead of $\tilde{p}_\mathbf{j}$.

B. Convergence analysis

We now study the behavior of the above regularized inverse scale space model (38). 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 Gronwall's inequality we have

$$\|\tilde{f} - \tilde{u}(t)\|_{L^2}^2 \leq e^{-2\epsilon^{1/2}(t-s)}\|\tilde{f} - \tilde{u}(s)\|_{L^2}^2, \quad t \geq s.$$

If $f \in L^2$, let $s = 0$, we have $t \to \infty$,

$$\|\tilde{f} - \tilde{u}(t)\|_{L^2}^2 \leq e^{-2\epsilon^{1/2}(t)}\|\tilde{f}\|_{L^2}^2 \searrow 0, \text{ as } t \nearrow \infty.$$

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

$$\begin{aligned}
0 \leq d_\mathbf{j}(\tilde{g}_\mathbf{j}, \tilde{u}_\mathbf{j}) &= F_\epsilon(\tilde{g}_\mathbf{j}) - F_\epsilon(\tilde{u}_\mathbf{j}) - (\tilde{g}_\mathbf{j} - \tilde{u}_\mathbf{j})\frac{\partial F_\epsilon}{\partial \tilde{u}_\mathbf{j}}(\tilde{u}_\mathbf{j}) \\
&= \sqrt{\tilde{g}_\mathbf{j}^2 + \epsilon} - \frac{\epsilon + \tilde{g}_\mathbf{j}\tilde{u}_\mathbf{j}}{\sqrt{\tilde{u}_\mathbf{j}^2 + \epsilon}} \\
&\to |\tilde{g}_\mathbf{j}| - \tilde{g}_\mathbf{j}\,\mathrm{sign}(\tilde{u}_\mathbf{j}) \quad \text{as } \epsilon \searrow 0.
\end{aligned}$$

For any $g \in B_1^1(L^1)$,

$$\begin{aligned}
\frac{d}{dt}D_J^{\tilde{p}}(\tilde{g}, \tilde{u}) &= \sum_\mathbf{j}(\tilde{g}_\mathbf{j} - \tilde{u}_\mathbf{j})\frac{d\tilde{p}_\mathbf{j}}{dt} = -\sum_\mathbf{j}(\tilde{g}_\mathbf{j} - \tilde{u}_\mathbf{j})(\tilde{f}_\mathbf{j} - \tilde{u}_\mathbf{j}) \\
&\leq -\sum_\mathbf{j}\frac{(\tilde{f}_\mathbf{j} - \tilde{u}_\mathbf{j})^2}{2} + \sum_\mathbf{j}\frac{(\tilde{f}_\mathbf{j} - \tilde{g}_\mathbf{j})^2}{2} \\
&< 0,
\end{aligned}$$

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

We may rewrite

$$d_\mathbf{j}(\tilde{g}_\mathbf{j}, \tilde{u}_\mathbf{j}) = \frac{\epsilon(\tilde{u}_\mathbf{j} - \tilde{g}_\mathbf{j})^2}{\sqrt{\tilde{u}_\mathbf{j}^2 + \epsilon}(\sqrt{\tilde{u}_\mathbf{j}^2 + \epsilon}\sqrt{\tilde{g}_\mathbf{j}^2 + \epsilon} + \epsilon + \tilde{g}_\mathbf{j}\tilde{u}_\mathbf{j})} \tag{39}$$

The factor $\epsilon > 0$ can be removed and we may rewrite,

$$\frac{d}{dt}\frac{D_J^{\tilde{p}}(\tilde{g}, \tilde{u})}{\epsilon} < 0, \quad \text{as long as } \|\tilde{f} - \tilde{u}\|_{L^2} > \|\tilde{f} - \tilde{g}\|_{L^2} = \sigma.$$

We also have a stopping criterion which is similar to the one for iterative refinement: we can stop the evolution (38) at $t = \bar{t}$ such that $\|\tilde{f} - \tilde{u}(t)\|_{L^2} = \sigma$.

VI. Numerical Examples

In this section we present two numerical examples of wavelet denoising using soft shrinkage, hard shrinkage and the iterative regularization method (W-IRM) and inverse scale space (W-ISS) flow we introduced above.

We add Gaussian i.i.d. noise n to the original clean image g and obtain our noisy image $f = g + n$. For different thresholds and parameters, there are two ways to define "optimal" results numerically: (i) the signal-to-noise-ratio (SNR, $SNR(w) := 20\log_{10}(\|g - \bar{g}\|_{L^2}/\|\eta - \bar{\eta}\|_{L^2})$, where $\eta = w - g$) of the restored image u is the biggest among all; (ii) $\|f - u\|_{L^2} \approx \sigma$. In applications we may have an estimate of σ but obviously no information about g, therefore we use the second criterion in our experiments. Moreover, as indicated in previous sections, (ii) is also our stopping criterion for W-IRM and W-ISS. To compare the numerical results in this section we choose thresholds τ for soft shrinkage and hard shrinkage such that their results satisfy $\|f - u\|_{L^2} = \sigma$ and we use our stopping criterion (ii) for W-IRM and W-ISS.

Figure 1 shows the original image g, which includes different shapes and scales, and the noisy image f with $\sigma = \|f - g\|_{L^2} = 30$ and $SNR(f) = 7.29$.

We choose the Haar basis and level 3 for wavelet decomposition in this example. In Figure 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 W-IRM ($\lambda = 0.001, \bar{k} = 11$) and W-ISS ($dt = 0.001, \bar{t} =$

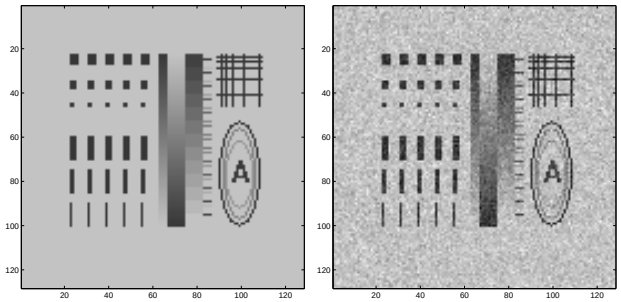


Fig. 1. shape image, 128×128. left: original image; right: noisy image, $\sigma = 30$, $SNR = 7.29$.

$0.014, \epsilon = 0.01$), 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.

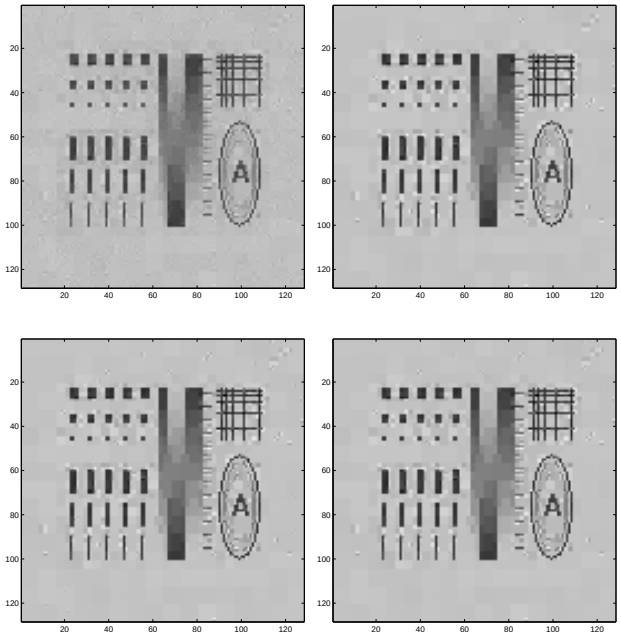


Fig. 2. First row: denoised results from soft shrinkage (left, $SNR = 12.03$) and hard shrinkage (right, $SNR = 13.04$); Second row: denoised results from W-IRM (27) (left, $SNR = 13.56$) and W-ISS (38) (right, $SNR = 13.45$). All $\|f - u\|_{L^2} \approx \sigma = 30$.

In Figure 2 there are some artifacts in the results. This is a common defect of wavelet imaging. We point out here that techniques such as cycle-spinning (cf. [28]), can be easily incorporated into our new methods. Moreover, in the W-ISS method proposed above, we introduced a regularized parameter ϵ. Numerical experiments show that bigger ϵ will make the results smoother and thus this can be used to decrease the artifacts. In Figure 3 we show a result u of W-ISS with $\epsilon = 10$, which has many fewer artifacts than the previous results. The corresponding $SNR = 12.84$ is higher than that of soft thresholding. 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. This is similar to the residual of hard thresholding. In soft thresholding, we removed some signal along with the residual (if we shrink only the detailed coefficients, then the signal loss is less, but still exists due to the shrinkage). In TV-based denoising, a similar defects occurs, as we discussed in Section II.

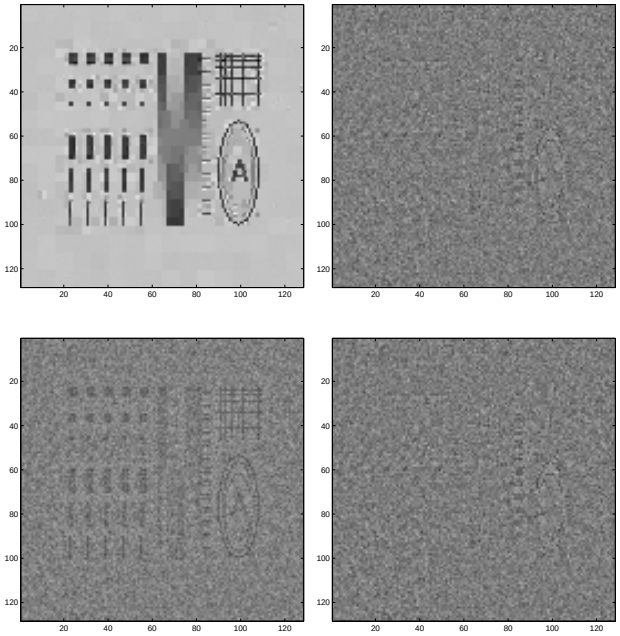


Fig. 3. First row: denoised result u from inverse scale space with $\epsilon = 10$ (left, $SNR = 12.84$) and corresponding residual $v = f - u$ (+128, right); Second row: residuals $v + 128$ of soft shrinkage (left) and hard shrinkage (right) in Figure 2.

In Figure 4 we show a result from TV based relaxed inverse scale space (cf. [11]). 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.

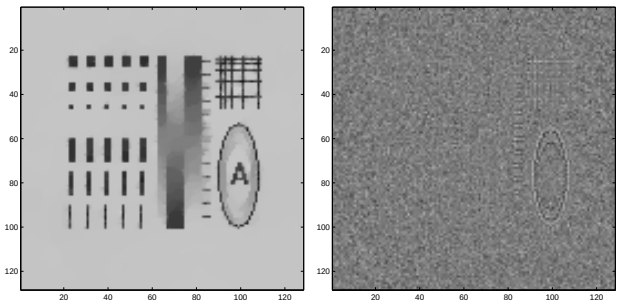


Fig. 4. Result from TV relaxed ISS ([11]). left: denoised u ($SNR = 14.96$, $\|f - u\|_{L^2} = \sigma = 30$); right: residual $v + 128$ ($v = f - u$).

In the second example we denoised an MRI image. Figure 5 shows the original image g and noisy image f with $\sigma = 30$ and $SNR(f) = 4.43$. We use db3 basis and level 3 for wavelet decomposition in this example. Figure 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.

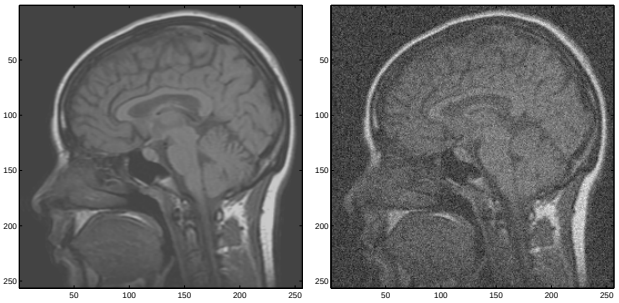


Fig. 5. MRI image, 256×256. left: original image; right: noisy image, $\sigma = 30$, $SNR = 4.43$,

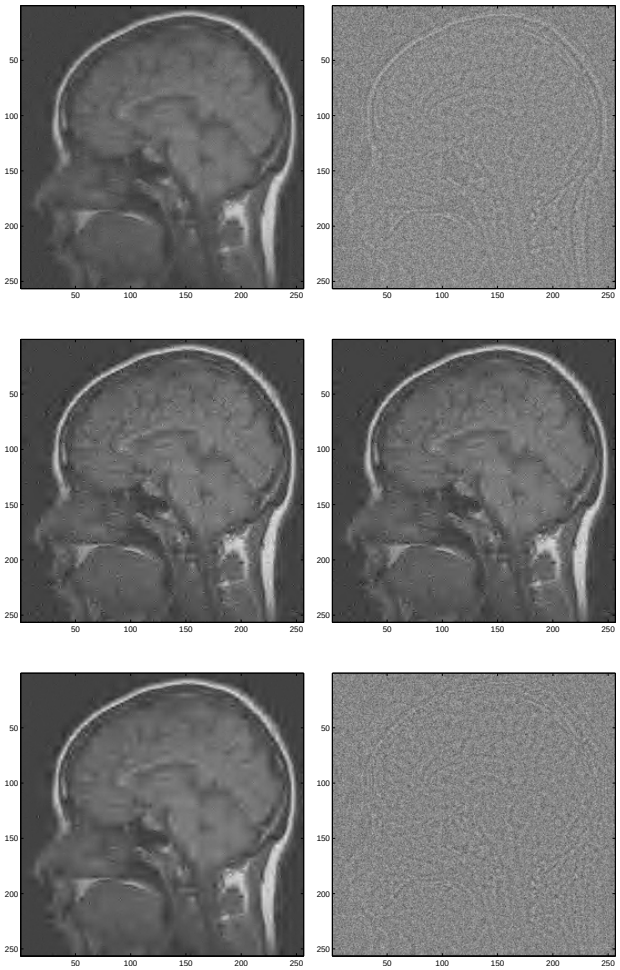


Fig. 6. First row: denoised result from soft shrinkage (left, $SNR = 11.72$) and corresponding residual $v = f - u$ (+128,right); Second row: denoised results from hard shrinkage (left, $SNR = 11.01$) and W-IRM (27) (right, $SNR = 11.01$); Third row: denoised result from W-ISS (38) (left, $\epsilon = 10, SNR = 11.94$) and corresponding residual $v + 128$ (right). All $\|f - u\|_{L^2} \approx \sigma = 30$.

VII. Conclusion

We have presented two alternatives to soft and hard wavelet shrinkage. These involve Bregman iteration and inverse scale space ideas borrowed from TV based restorations ([9]–[11]). The iterative soft shrinkage gives firm shrinkage, with the thresholds dynamically changing in the iteration. It appears that the new methods W-IRM and W-ISS, especially the latter, 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.

As we mentioned at the beginning of Section III, some shrinkage methods apply the shrinkage operator only on the detail coefficients and keep the scaling ones unchanged. Since the summation parts in our models are separable, our discussion in this paper can be easily extended to those methods. Furthermore, W-IRM and W-ISS can be incorporated with some other ideas developed in wavelet denoising, e.g., cycle-spining (cf. [28]) methods which were introduced to reduce the artifacts in the denoising results. The main goal of this paper is to discover the link between TV-based ideas and wavelet-based models, and our results so far are promising.

Our future work will involve the role of the parameter ϵ in the regularization (31), and using our methods on different wavelet bases and different types of noise.

Acknowledgments

We thank Gabriele Steidl (University of Mannheim) for mentioning firm shrinkage. We also thank the unknown reviewers for their useful suggestions in improving the quality of this paper.

Appendix
The Proof of Theorem 1

(1) Plugging (23) into (25) we have

$$\tilde{v}_{\mathbf{j}}^{(k)} = \begin{cases} \frac{1}{\lambda}\text{sign}(\tilde{f}_{\mathbf{j}} + \tilde{v}_{\mathbf{j}}^{(k-1)}), & \text{if } |\tilde{f}_{\mathbf{j}} + \tilde{v}_{\mathbf{j}}^{(k-1)}| > \frac{1}{\lambda}, \\ \tilde{f}_{\mathbf{j}} + \tilde{v}_{\mathbf{j}}^{(k-1)}, & \text{if } |\tilde{f}_{\mathbf{j}} + \tilde{v}_{\mathbf{j}}^{(k-1)}| \le \frac{1}{\lambda}, \end{cases} \tag{40}$$

for $k \ge 1$. Since $\tilde{v}_{\mathbf{j}}^{(0)} = 0$, we have $\text{sign}(\tilde{v}_{\mathbf{j}}^{(1)}) = \text{sign}(\tilde{f}_{\mathbf{j}})$. By induction, $\text{sign}(\tilde{v}_{\mathbf{j}}^{(k)}) = \text{sign}(\tilde{f}_{\mathbf{j}})$ for all $k \ge 1$. Next we also prove (26) by induction. For $k = 1$, we have (26) from (18). For $k \ge 2$,

(i) If $|\tilde{f}_{\mathbf{j}}| > \frac{1}{(k-1)\lambda}$, then $\tilde{v}_{\mathbf{j}}^{(k-1)} = \frac{1}{\lambda}\text{sign}(\tilde{f}_{\mathbf{j}})$, and $|\tilde{f}_{\mathbf{j}} + \tilde{v}_{\mathbf{j}}^{(k-1)}| \ge |\tilde{f}_{\mathbf{j}}| > \frac{1}{(k-1)\lambda} > \frac{1}{k\lambda}$. From (40), $\tilde{v}_{\mathbf{j}}^{(k)} = \frac{1}{\lambda}\text{sign}(\tilde{f}_{\mathbf{j}})$;

(ii) If $|\tilde{f}_{\mathbf{j}}| \le \frac{1}{(k-1)\lambda}$, 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 (40), 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, $|k\tilde{f}_{\mathbf{j}}| \le \frac{1}{\lambda}$, $\tilde{v}_{\mathbf{j}}^{(k)} = \tilde{f}_{\mathbf{j}} + \tilde{v}_{\mathbf{j}}^{(k-1)} = k\tilde{f}_{\mathbf{j}}$.

This validates (26).

(2) Now we prove (27). From (25) we have $\tilde{u}_{\mathbf{j}}^{(k)} = \tilde{f}_{\mathbf{j}} + \tilde{v}_{\mathbf{j}}^{(k-1)} - \tilde{v}_{\mathbf{j}}^{(k)}$. Using (26), for $k > 1$ we have

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

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

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

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

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

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

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 (27) reduces to (15). This validates (27).

(3) From (26) we have

$$\tilde{p}_{\mathbf{j}}^{(k)} = \begin{cases} \text{sign}(\tilde{f}_{\mathbf{j}}), & \text{if } |\tilde{f}_{\mathbf{j}}| > \frac{1}{k\lambda}, \\ k\lambda\tilde{f}_{\mathbf{j}}, & \text{if } |\tilde{f}_{\mathbf{j}}| \leq \frac{1}{k\lambda}, \end{cases} \tag{41}$$

For the first part, we have $\tilde{u}_{\mathbf{j}}^{(k)} \neq 0$ and $\partial F(\tilde{u}_{\mathbf{j}}^{(k)}) = \text{sign}(\tilde{u}_{\mathbf{j}}^{(k)}) = \tilde{p}_{\mathbf{j}}^{(k)}$. For the second part, we have $\tilde{u}_{\mathbf{j}}^{(k)} = 0$, and $|\tilde{p}_{\mathbf{j}}| \leq 1$, $\tilde{p}_{\mathbf{j}} \in \partial F(\tilde{u}_{\mathbf{j}}^{(k)})$.

Overall we have proved Theorem 1. This also shows that the dual variable $\tilde{p}^{(k)}$ updated via (21) is automatically a subgradient of $J(\tilde{u}^{(k)})$.

REFERENCES

[1] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D*, vol. 60, pp. 259–268, 1992.

[2] G. Aubert and L. Vese, "A variational method in image recovery," *SIAM J. Numer. Anal.*, vol. 34, pp. 1948–1979, 1997.

[3] D. Donoho, "Denoising by soft thresholding," *IEEE. Trans. Inform. Theory*, vol. 41, no. 3, pp. 613–627, 1995.

[4] D. Donoho and I. Johnstone, "Ideal spatial adaptation via wavelet shrinkage," *Biometrika*, vol. 81, pp. 425–455, 1994.

[5] ——, "Adapting to unknown smoothness via wavelet shrinkage," *J. Amer. Stat. Assoc.*, vol. 90, pp. 1200–1224, 1995.

[6] A. Chambolle, R. DeVore, N.-Y. Lee, and B. Lucier, "Nonlinear wavelet image processing: Variational problems, compression, and noise removal through wavelet shrinkage," *IEEE Trans. Image Proc.*, vol. 7, no. 3, pp. 319–335, 1998.

[7] G. Steidl, J. Weickert, T. Brox, P. MráZek, and M. Welk, "On the equvalence of soft wavelet shrinkage, total variation diffusion, total variation regularization, and SIDEs," *SIAM J. Numer. Anal.*, vol. 42, no. 2, pp. 686–713, 2004.

[8] I. Daubechies and G. Teschke, "Variational image restoration by means of wavelets: Simultaneous decomposition, deblurring and denoising," *Applied and Computational Harmonic Analysis*, vol. 19, no. 1, pp. 1–16, 2005.

[9] 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.*, vol. 4, pp. 460–489, 2005.

[10] M. Burger, S. Osher, J. Xu, and G. Gilboa, "Nonlinear inverse scale space methods for image restoration," *Lecture Notes in Computer Science*, vol. 3752, pp. 25–36, 2005.

[11] M. Burger, G. Gilboa, S. Osher, and J. Xu, "Nonlinear inverse scale space methods," *Comm. Math. Sci.*, vol. 4(1), pp. 175–208, 2006.

[12] H.-Y. Gao and A. G. Bruce, "WaveShrink with firm shrinkage," *Statist. Sinica*, vol. 7, no. 4, pp. 855–874, 1997.

[13] W. P. Ziemer, *Weakly Differentiable Functions*. New York: Springer-Verlag, 1989.

[14] D. Strong and T. Chan, "Edge-preserving and scale-dependent properties of total variation regularization," *Inverse Problems*, vol. 19, pp. S165–S187, 2003.

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

[16] I. Ekeland and R. Téman, *Convex Analysis and Variational Problems*. Philadelphia: Classics Appl. Math. 28, SIAM, 1999.

[17] 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.*, vol. 7, pp. 200–217, 1967.

[18] G. Chen and M. Teboulle, "Convergence analysis of a proximal-like minimization algorithm using bregman functions," *SIAM J. Optim.*, vol. 3, pp. 538–543, 1993.

[19] K. Kiwiel, "Proximal minimization methods with generalized bregman functions," *SIAM J. Control Optim.*, vol. 35, pp. 1142–1168, 1997.

[20] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia: SIAM, 1992.

[21] Y. Meyer, *Wavelets and Operators*. New York: Cambridge Univ. Press, 1992.

[22] A. Cohen, I. Daubechies, and P. Vial, "Wavelets on the interval and fast wavelet transforms," *Appl. Comput. Harmon. Anal.*, vol. 1, pp. 54–81, 1993.

[23] R. DeVore and B. J. Lucier, "Wavelets," *Acta Numerica*, vol. 1, pp. 1–56, 1992.

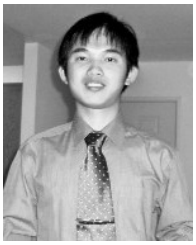
[24] Y. Meyer, *Ondelettes et Opérateurs I: Ondelettes*. Paris: Hermann, 1990.

[25] R. DeVore and B. Lucier, "Fast wavelet techniques for near-optimal image processing," in *IEEE Military Commun. Conf. Rec.* San Diego, CA: IEEE Press, 1992, pp. 1129–1135.

[26] A. Cohen, R. DeVore, P. Petrushev, and H. Xu, "Nonlinear approximation and the space $BV(\mathbb{R}^2)$," *American J. of Math.*, vol. 121, no. 3, pp. 587–628, 1999.

[27] A. Cohen, W. Dahmen, I. Daubechies, and R. DeVore, "Harmonic analysis of the space BV," *Rev. Mat. Iberoamericana*, vol. 19, pp. 235–263, 2003.

[28] D. Donoho and R. Coifman, "Translation-invariant de-noising," in *Wavelets and Statistics*, A. Antoniadis and G. Oppenheim, Eds. New York: Springer-Verlag, 1995, pp. 125–150.



Jinjun Xu received the B.S. and M.S. degrees in applied mathematics from Beijing University in 1998 and 2001 respectively, the M.A. and the Ph.D. degree in mathematics from University of California, Los Angeles in 2002 and 2006 respectively. He is currently a postdoctoral scholar in the Department of Mathematics, University of California, Los Angeles.

Stanley Osher received the B.S. degree in mathematics from Brooklyn College in 1962, the M.S. and Ph.D. degrees in mathematics from New York University in 1964 and 1996 respectively.

He has been with the University of California, Los Angeles, since 1977. He is currently a Professor in the Department of Mathematics, and the Director of Special Projects in the Institute for Pure and Applied Mathematics.

Dr. Osher is the founder and CEO of Level Set Systems, Inc. and has cofounded two other companies. He was an invited speaker at the 1994 International Congress of Mathematics and an ISI Original Highly Cited Researcher. He received the 2002 Computational Mechanics Award from Japan Society of Mechanical Engineers, the 2003 ICIAM Pioneer Prize and the 2005 SIAM Kleinman Prize. He was elected to the National Academy of Sciences in 2005.