Commun. Comput. Phys.
doi: 10.4208/cicp.210709.180310a

Vol. **8**, No. 5, pp. 976-994
November 2010

Alternating Minimization Method for Total Variation Based Wavelet Shrinkage Model

Tieyong Zeng[1], Xiaolong Li[2] and Michael Ng[1,*]

[1] *Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Hong Kong, China.*

[2] *Institute of Computer Science & Technology, Peking University, Beijing 100871, China.*

Received 21 July 2009; Accepted (in revised version) 18 March 2010

Available online 31 May 2010

Abstract. In this paper, we introduce a novel hybrid variational model which generalizes the classical total variation method and the wavelet shrinkage method. An alternating minimization direction algorithm is then employed. We also prove that it converges strongly to the minimizer of the proposed hybrid model. Finally, some numerical examples illustrate clearly that the new model outperforms the standard total variation method and wavelet shrinkage method as it recovers better image details and avoids the Gibbs oscillations.

AMS subject classifications: 90C25, 52A41, 28E99, 62P35

Key words: Alternating minimization, convergence, Gibbs oscillation, wavelet shrinkage, total variation.

1 Introduction

Digital image denoising plays an important role in numerous areas of applied sciences such as medical and astronomical imaging, film restoration, and image/video coding. Throughout this paper, we suppose that Ω is an open bounded set of $\mathbb{R}^2$ with Lipschitz boundary and all the images are regarded as elements in a classical space

$$\mathcal{H} := L^2(\Omega),$$

*Corresponding author. *Email addresses:* zeng@hkbu.edu.hk (T. Zeng), lixiaolong@icst.pku.edu.cn (X. Li), mng@math.hkbu.edu.hk (M. Ng)

a separable infinite-dimensional real Hilbert space with usual inner product $\langle\cdot,\cdot\rangle$, norm $\|\cdot\|_2$. Note that as every element in $L^2(\Omega)$ can be regarded as a continuous linear functional which maps every test function to their inner product, here we consider $\mathcal{H}$ as a distributional space for convenience. Moreover, we focus on a common noisy model: an ideal image $u\in\mathcal{H}$ is observed in the presence of an additive zero-mean Gaussian noise $b\in\mathcal{H}$ of standard derivation σ. Thus the observed image $f\in\mathcal{H}$ is obtained by

$$f=u+b. \tag{1.1}$$

In the past decades, many denoising approaches have been proposed to handle this ill-posed problem. One of the widely studied techniques is the wavelet shrinkage method, which acknowledges that by applying a wavelet transform on a noisy image, random noise will contribute mainly as small coefficients in the high frequencies. Therefore, theoretically one can remove much of the noise in the image by setting these small coefficients to zero. The wavelet hard shrinkage method, which shrinks the wavelet coefficients smaller than some predefined threshold in magnitude to zero, is extremely easy and rapid to implement. Depending on the threshold, it can reduce noise rather effectively. However, it also revokes unpleasant artifacts around discontinuities as a result of Gibbs phenomenon. As artifacts in some image processing tasks may lead to great inconveniences, the wavelet hard shrinkage can not be used in these tasks without extra efforts. A development over the wavelet hard shrinkage is the wavelet soft shrinkage [15, 16], which diminishes significantly the Gibbs oscillation. Usually, the potential of wavelet shrinkage methods is rather promising when they are combined with other complex techniques which often try to take advantage of geometric information by applying wavelet-like bases better characterizing discontinuities, such as curvelets [5, 17] which can be regarded as one of the best methods among this direction. However, none of them can entirely efface the Gibbs oscillation.

Another important approach adopts regularization techniques and variational principles. Usually this approach is to determine the denoised image by minimizing a cost function consisting a data-fitting term and a regularization term

$$\min_{w\in\mathcal{H}}\frac{1}{2}\|f-w\|_2^2+\beta\mathcal{R}(w), \tag{1.2}$$

where $\mathcal{R}$ is the regularization functional and β is a positive parameter. Various possibilities for $\mathcal{R}(w)$ have been proposed in literature and earlier efforts concentrated on least squares based functionals such as $\|\Delta w\|_2^2$, $\|\nabla w\|_2^2$ and others. Though noise can be adequately reduced, these regularization functionals also impose penalty on discontinuity, conducting to rather smooth restoration images, with subtle details disappeared.

A better choice for $\mathcal{R}(w)$ was developed in [24], in which $\mathcal{R}(w)$ is the total variation of $w\in\mathcal{H}$ commonly defined by

$$\mathcal{R}(w)=TV(w):=\int_\Omega|Dw|, \tag{1.3}$$

where Dw is the distributional gradient of w.

Remarkable investigations have demonstrated that the total variation permits sharper reconstruction and better preserves edges in image. Moreover, it is capable of removing artifacts revoked by other methods since these artifacts usually have rather high total variation value. Therefore, current researches are increasingly focusing on the combinations of total variation and wavelet-like methods. In [23, 25], in order to reduce the Gibbs artifacts, a set of wavelet coefficients was interpolated according to a total variation criterion. This was close to the approach of [9] where PDE techniques were used to reduce edge artifacts generated by wavelet shrinkage. In [10, 19] the authors proposed to determine the most meaningful wavelet coefficients, also via a total variation criterion. In [18, 34], another algorithm was introduced where information was removed from the residual image $w-v$ by using a wavelet analysis, thus allowing for the composition of wavelet techniques with any other method such as variational one. In [20, 26], a hybrid approach by minimizing the total variation on a convex set controlled by a wavelet basis (or other dictionary) was investigated. In [30], the authors analyzed the possibilities to avoid Gibbs-like artifacts for multiscale Haar wavelet shrinkage by combining the advantages of wavelet and PDE/variational approaches.

The goal of this paper is to provide a mixed variational model for total variation and wavelet shrinkage to enrich the hybrid direction. For this, we need to transform the wavelet shrinkage method to variational form. Indeed, it is well known that the wavelet soft shrinkage is equivalent to the following optimization task

$$\min_{w\in\mathcal{H}} \frac{1}{2}\|f-w\|_2^2+\alpha \sum_{\varphi\in\mathcal{D}} |\langle w,\varphi\rangle|,$$

where α is a positive parameter and $\mathcal{D}$ is a orthogonal wavelet basis of $\mathcal{H}$.

The soft-shrinkage can be interpreted from a MAP point of view. Generally, if the wavelet coefficients of the underlying image are assumed independent and Laplace distributed (zero mean) then one can prove that the linear Bayesian estimate is optimal in mean squared error sense and it is just the wavelet shrinkage. However, usually for natural images, the distribution of wavelet coefficients is far from a single statistical model such as Gaussian or Laplace since they are highly correlated. Indeed, it is pointed out in [12] that most natural images have non-stationary properties as they typically consist of regions of smoothness and sharp transitions. These regions of varying characteristics can be well differentiated in the wavelet domain. A common technique to overcome this problem is to apply adaptive shrinkage scheme. In [12], each wavelet coefficient is modeled as a random variable following the so-called Generalized Gaussian distribution with an unknown parameter and a spatially adaptive wavelet shrinkage method based on context modeling is then proposed. In [21], the author examined some non-linear shrinkage methods which take into account the correlation structures of the multi-resolution wavelet coefficients.

In this paper, instead of modeling those non-stationary structures (which are rather complex for natural images), we are interested in a new scheme which is composed of

three steps: first, remove the basic structure out of the noisy image to get a stationary residual; second, apply wavelet shrinkage on the stationary residual to extract extra details; third, add back the extra details to the basic structure to form a result image.

Noting that in many cases, the cartoon part of Rudin-Osher-Fatemi method is rather good to extract the basic structure from the noisy image and the only problem is that it loses some texture information as usually the texture part has a higher total variation value. This also hints us to consider the recompensation idea: if one can extract some extra texture information from the residual and add it back to the cartoon part, then we have a greater chance to obtain a better result. In these regards, we propose to study the following unified minimization model

$$\min_{w,r\in\mathcal{H}} \frac{1}{2}\|f-r-w\|_2^2+\alpha\sum_{\varphi\in\mathcal{D}}|\langle r,\varphi\rangle|+\beta TV(w), \tag{1.4}$$

where α,β are fixed positive parameters and we will take

$$u=w+r,$$

as the denoising result. The physical idea of this model is that the noisy image f can be decomposed into three parts: a Gaussian noise part, a cartoon part characterized by the total variation and a detail part which has sparse representation over the dictionary $\mathcal{D}$. Note that the idea of decomposing image into several rather homogenous components appears numerously in the domain of image processing [3, 7, 13, 29]. In this paper, we choose the infinite convolution of TV and wavelet-type energy to illustrate that simple ideas also derive rather good results. The differences between our model with the previous works are clear. For instance, since none of [9, 10, 18–20, 23, 25, 26, 30, 32–34] has been addressed in the form of decomposition into three different components: cartoon part, texture part, noise part, they are evidently different from ours. Moreover, compared to [7,13] where inf-convolution of two convex potentials are proposed, we concentrate on a specific energy choice which provides promising results; compared to [29], our minimization algorithm on the objective functional itself, not the approximated ones, is casted in a more rigorous theoretical way since we use only one dictionary; compared to [3] where the indeed purpose is for decomposition, we use an energy more suitable for denoising since total variation and wavelet shrinkage have been supported by extensive works in image processing. Moreover, when $\alpha=+\infty$, we need to take $r=0$ and thus this model reduces to the Rudin-Osher-Fatemi (ROF) model. Similarly, if $\beta=+\infty$, we should take w as zero (or other constant), then this model reduces to the wavelet shrinkage.

The outline of the paper is as follows. In Section 2, some preliminaries and discussions on the new model are addressed. In Section 3, we present an alternating minimization direction algorithm to solve the proposed model. In Section 4, the convergence of the proposed algorithm is analyzed. In Section 5, numerical examples are given to illustrate the effectiveness of the proposed model. Finally, some concluding remarks are addressed in Section 6.

2 Preliminaries and discussions

2.1 The BV space

The $BV(\Omega)$ is the subset of functions $u \in L^1(\Omega)$ such that the following quantity is finite

$$J(u) = \sup\Big\{ \int_\Omega u \cdot \mathrm{div}(\xi(x))\mathrm{d}x \big| \xi \in C_0^\infty(\Omega,\mathbb{R}^2),\ \ \|\xi\|_{L^\infty(\Omega,\mathbb{R}^2)} \le 1 \Big\}.$$

$BV(\Omega)$ endowed with the norm $\|u\|_{BV} = \|u\|_{L^1} + J(u)$, is a Banach space. If $u \in BV(\Omega)$, the distributional derivative Du is a bounded Radon measure and the above quantity corresponds to the total variation, i.e.,

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

and we can also define the BV space as

$$BV(\Omega) = \Big\{ w \in L^1(\Omega);\ \int_\Omega |Dw| < +\infty \Big\}.$$

For $\Omega \subset \mathbb{R}^2$, if $1 \le p \le 2$, we have

$$BV(\Omega) \subset L^p(\Omega).$$

Moreover, for $1 \le p < 2$, this embedding is compact. See book [2] for further details about the BV space.

Since $BV(\Omega) \subset L^2(\Omega)$, we can extend the functional J (which we also denote by TV) over $L^2(\Omega)$

$$TV(u) = \begin{cases} \int_\Omega |Du|, & \text{if } u \in BV(\Omega), \\ +\infty, & \text{if } u \in L^2(\Omega) \setminus BV(\Omega). \end{cases} \tag{2.1}$$

Note that the definition of Eq. (2.1) is consistent with Eq. (1.3).

Let μ be the Lebesgue measure. The following Calderón-Zygmund theorem is useful for us.

Theorem 2.1. (Thm 3.83, [1]) *Any function $u \in BV(\Omega)$ is approximately differentiable at μ-almost every point of Ω. Moreover, the approximate differential ∇u is the density of the absolutely continuous part of Du with respect to μ.*

2.2 Basic properties of the unified model

Now suppose the Hilbert space $(\mathcal{H},\mathcal{H})$ is the product of $\mathcal{H}$ and $\mathcal{H}$ and for any $(r,w) \in (\mathcal{H},\mathcal{H})$, we consider the following functional

$$E(r,w) := \frac{1}{2}\|f - r - w\|_2^2 + \alpha \sum_{\varphi \in \mathcal{D}} |\langle r, \varphi \rangle| + \beta TV(w). \tag{2.2}$$

The Bessel's equality of orthogonal wavelet basis is essential to the proof of the following proposition.

Proposition 2.1. We have:

1. The functional $E(r,w)$ is coercive and convex,
2. For any global minimal point (r^*,w^*) of $E(r,w)$, r^*+w^* is unique,
3. The global minimal point of $E(r,w)$ is unique up to an image of approximate differential zero.

Proof. The convexity of E is obvious. Suppose that $(r,w)\to\infty$. Then one of the following two cases must be true:

(i) $\|r\|_2\to+\infty$; (ii) $\|r\|_2<+\infty$, $\|w\|_2\to+\infty$.

Noting that

$$\Big(\sum_{\varphi\in\mathcal{D}}|\langle r,\varphi\rangle|\Big)^2\geq\sum_{\varphi\in\mathcal{D}}|\langle r,\varphi\rangle|^2=\|r\|_2^2,$$

both cases will lead to

$$E(r,w)\to+\infty.$$

Therefore, $E(r,w)$ is coercive. Moreover, since E is coercive and convex, the global minimal point set of E is non-empty, compact and convex. Furthermore, if we suppose that $(r_1,w_1),(r_2,w_2)\in(\mathcal{H},\mathcal{H})$ are two different global minimal points, then for any $\theta\in(0,1)$, we have

$$\theta E(r_1,w_1)+(1-\theta)E(r_2,w_2)=E\big(\theta r_1+(1-\theta)r_2,\theta w_1+(1-\theta)w_2\big).$$

As each term in $E(r,w)$ is convex, the above equality means that for each term in $E(r,w)$ the equality holds

$$\begin{aligned}&\theta\|f-r_1-w_1\|_2^2+(1-\theta)\|f-r_2-w_2)\|_2^2=\|f-\theta(r_1+w_1)-(1-\theta)(r_2+w_2)\|_2^2,\\&\theta TV(w_1)+(1-\theta)TV(w_2)=TV\big(\theta w_1+(1-\theta)w_2\big).\end{aligned}$$

We thus obtain

$$r_1+w_1=r_2+w_2, \tag{2.3}$$

$$\theta\int_\Omega|Dw_1|+(1-\theta)\int_\Omega|Dw_2|=\int_\Omega|\theta Dw_1+(1-\theta)Dw_2|. \tag{2.4}$$

Eq. (2.3) illustrates the second assertion. Let

$$S=S_1\cup S_2,$$

where S_1,S_2 is respectively the approximately non-differentiable set of w_1,w_2. As $TV(\cdot)$ is indeed convex on S and on $\Omega\setminus S$, by Jensen inequality, we have

$$\begin{aligned}&\theta\int_S|Dw_1|+(1-\theta)\int_S|Dw_2|\geq\int_S|\theta Dw_1+(1-\theta)Dw_2|,\\&\theta\int_{\Omega\setminus S}|Dw_1|+(1-\theta)\int_{\Omega\setminus S}|Dw_2|\geq\int_{\Omega\setminus S}|\theta Dw_1+(1-\theta)Dw_2|.\end{aligned}$$

Together with Eq. (2.4), the above two inequalities should be equalities. Hence, we have

$$\theta\int_{\Omega\setminus S}|\nabla w_1|+(1-\theta)\int_{\Omega\setminus S}|\nabla w_2|=\int_{\Omega\setminus S}|\theta\nabla w_1+(1-\theta)\nabla w_2|. \tag{2.5}$$

By Theorem 2.1, we know that

$$\mu(S)\leq\mu(S_1)+\mu(S_2)=0.$$

Hence, we have

$$\theta\int_{\Omega}|\nabla w_1|+(1-\theta)\int_{\Omega}|\nabla w_2|=\int_{\Omega}|\theta\nabla w_1+(1-\theta)\nabla w_2|. \tag{2.6}$$

By using (2.6), and denoting

$$g=\theta|\nabla w_1|+(1-\theta)|\nabla w_2|-|\theta\nabla w_1+(1-\theta)\nabla w_2|,$$

then we have

$$g\geq 0 \quad \text{and} \quad \int_{\Omega} g=0.$$

Therefore, $g=0$ (in distributional sense, this means that for any test function $\eta\in C_0^\infty(\Omega)$, we have $\langle g,\eta\rangle=0$) and $\nabla w_1=\nabla w_2$. (This means that the approximate differential parts of w_1 and w_2 are the same). This leads to

$$\nabla(w_1-w_2)=0.$$

Hence, the global minimal point of $E(r,w)$ is unique up to an image of approximate differential zero. □

2.3 Characteristics of the two components

Suppose (r^*,w^*) is a solution to (1.4). Then easily we know that

$$TV(w^*)\leq\frac{1}{2}\|f\|_2^2<+\infty.$$

Noting that Ω is an open bounded subset of $\mathbb{R}^2$, we have

$$\mu(\Omega)<+\infty,$$

where $\mu(\Omega)$ is the Lebesgue measure. Using Corollary 14.7 of Book [4], since $w^*\in L^2(\Omega)$, we have

$$w^*\in L^s(\Omega),\quad \forall s\in[1,2).$$

In particular, $w^* \in L^1(\Omega)$ and so w^* is in the BV space. Now for r^*, observing that

$$r^* = \arg\min_{r\in\mathcal{H}} \frac{1}{2}\|f-r-w^*\|_2^2 + \alpha \sum_{\varphi\in\mathcal{D}} |\langle r,\varphi\rangle|,$$

we know that r^* is the wavelet soft-shrinkage result of $f-w^*$:

$$r^* = \sum_{\varphi\in\mathcal{D}} \rho_\alpha(\langle f-w^*,\varphi\rangle)\varphi,$$

where ρ_α is the soft-shrinkage function defined by

$$\rho_\alpha(t) = \begin{cases} t-\alpha, & \text{if } t\geq\alpha, \\ 0, & \text{if } t\in(-\alpha,\alpha), \\ t+\alpha, & \text{otherwise.} \end{cases} \tag{2.7}$$

Hence, every non-zero term appearing in r^* will lead to a wavelet coefficient of $f-w^*$ with modulus bigger than α. However, by Proposition 2.1, we know that $f-w^*$ is bounded. Therefore, r^* is linear combination of finitely many wavelet filter terms.

Usually, a natural image is composed of two different components: a smooth part and an oscillating part (or detail part). For the oscillating part, the total variation is rather high, so it will prefer to be represented by r^*; for smooth part, as the TV is rather low, it will be described by w^*. Overall, we obtain a decomposition of the ideal image: cartoon part w^* and oscillating part r^*. The former is in the BV space and the latter is a linear span of finitely many terms of wavelet filters which have rather high frequencies.

3 Alternating minimization direction algorithm

In Eq. (1.4), there are two unknown images, one is the cartoon part w, and the other is the oscillating part r. In this paper, we propose an alternating minimization direction approach to solve the new model:

1. Initialize r_0,
2. Repeat until convergent
   - update

$$w_n = \arg\min_{w} \frac{1}{2}\|(f-r_{n-1})-w\|_2^2 + \beta TV(w), \tag{3.1}$$

$$r_n = \arg\min_{r} \frac{1}{2}\|(f-w_n)-r\|_2^2 + \alpha \sum_{\varphi\in\mathcal{D}} |\langle r,\varphi\rangle|; \tag{3.2}$$

   - take

$$u_n = w_n + r_n.$$

In the upcoming section, we will study the convergence of w_n,r_n. Let us first analyze the computational cost of the alternating minimization algorithm.

The first step of the method is to compute the cartoon part by applying an exact TV denoising scheme to the image without the oscillating part. The minimizer of the optimization problem

$$\min_w \frac{1}{2}\|(f-r_{n-1})-w\|_2^2+\beta TV(w),$$

can be solved by many TV denoising methods such as Chambolle's Projection [6], the semismooth Newton method [22], the multilevel optimization method [11], and the graph-based optimization method [8].

The second step is to calculate the oscillating part from the image without the cartoon part. The minimizer of the optimization problem

$$\min_r \frac{1}{2}\|(f-w_n)-r\|_2^2+\alpha\sum_{\varphi\in\mathcal{D}}|\langle r,\varphi\rangle|,$$

is given by the wavelet soft-shrinkage on $f-w_n$. Therefore, this step has a simple closed form.

Overall, as the solving of Eq. (3.2) has a simple closed form and the minimization of Eq. (3.1) is efficient, the proposed algorithm will be rather effective. Indeed, numerically we observe that for real image restoration task, setting the outer iteration as 10 to 15 has been already enough.

4 Convergence analysis

In this section, we investigate the convergence of the alternating minimization algorithm. We make use of the classical alternating minimization method to show that the algorithm converges weakly to a minimizer of $E(r,w)$ and then prove that the convergence holds in the norm topology as well.

4.1 Lower semi-continuity

Before the convergence analysis, we present some preliminary results on the lower semi-continuity.

Definition 4.1. Let X be a Banach space and $F:X\to\mathbb{R}$ be a functional on X. F is called lower semicontinuous (l.s.c.) for the weak topology if for any sequence (x_n) converging weakly to x_0, we have

$$\liminf_{n\to\infty} F(x_n)\geq F(x_0).$$

The same definition can be given with the strong topology.

Theorem 4.1. (l.s.c. strong and weak, see book [2]) *Let $F \to \mathbb{R}$ be convex. Then F is weakly l.s.c. if and only if F is strongly l.s.c..*

Using this theorem, one can prove:

Proposition 4.1. The following functions are all weakly l.s.c.,

(i) $f_1 : x \mapsto \|x-a\|_2^2$ where $x \in \mathcal{H}$ and a is a fixed element in $\mathcal{H}$;

(ii) $f_2 : x \mapsto \sum_{\varphi \in \mathcal{D}} |\langle x, \varphi \rangle|$ where $x \in \mathcal{H}$ and $\mathcal{D}$ is a countable dictionary in $\mathcal{H}$;

(iii) $f_3 : x \mapsto TV(x)$ where $x \in \mathcal{H}$.

Proof. The weak lower semi-continuity of f_1, f_2 is straightforward. The weak lower semi-continuity of f_3 is a special case of Theorem 2.1.3 of book [2]. □

4.2 Weak convergence

Now we have the result of weak convergence.

Theorem 4.2. *We have:*

(i) the sequence (r_n, w_n) converges weakly to a global minimal point of $E(r,w)$;

(ii) the sequence (u_n) converges weakly to a unique point regardless of the initialization.

Proof. As our alternating minimization algorithm falls into the category of Proximal Forward-Backward Splitting method (see [13]), the weak convergence of (u_n) is direct. Noting for any minimal point (r^*, w^*) of $E(r,w)$, by Proposition 2.1, $r^* + w^*$ is always constant. Therefore, the sequence (u_n) converges weakly to this constant which is then independent of the initialization of (r_0, w_0). □

4.3 Strong convergence

We can also prove the strong convergence of r_n, w_n with the help of the following lemma.

Lemma 4.1. (Lemma 4.3 of [14]) *If the sequence of vectors $(g_k)_{k \in \mathbb{N}}$ converges weakly in $\mathcal{H}$ to g, and*

$$\lim_{k \to \infty} \sum_{\varphi \in \mathcal{D}} |\langle \varphi, g_k \rangle| = \sum_{\varphi \in \mathcal{D}} |\langle \varphi, g \rangle|,$$

then $(g_k)_{k \in \mathbb{N}}$ converges to g in the $\mathcal{H}$-norm, i.e.,

$$\lim_{k \to \infty} \|g - g_k\|_2 = 0.$$

Finally, we can prove our main result.

Theorem 4.3. *We have:*

(i) the sequence (w_n,r_n) converges strongly to a global minimal point of $E(r,w)$;

(ii) the sequence (u_n) converges strongly to a unique point regardless of the initialization.

Proof. We only need to prove that the convergence of (r_n), (w_n) is strong. Noting that (r_n,w_n) converges weakly to a global minimizer (r^*,w^*) of $E(r,w)$, we know that

$$\lim_{n\to\infty} E(r_n,w_n)=E(r^*,w^*).$$

Therefore,

$$\limsup_{n\to\infty} E(r_n,w_n)=E(r^*,w^*).$$

As $E(r,w)$ contains three convex terms, we thus have

$$\begin{aligned}&\liminf_{n\to\infty}\frac{1}{2}\|f-w_n-r_n\|_2^2+\alpha\limsup_{n\to\infty}\sum_{\varphi\in\mathcal{D}}|\langle r_n,\varphi\rangle|+\beta\liminf_{n\to\infty}TV(w_n)\\ \le&\limsup_{n\to\infty}E(r_n,w_n)=E(r^*,w^*),\end{aligned}$$

where we used twice the propriety that

$$\liminf x_n+\limsup y_n\le\limsup(x_n+y_n).$$

Using Proposition 4.1, we have

$$\begin{aligned}&\liminf_{n\to\infty}\frac{1}{2}\|f-w_n-r_n\|_2^2\ge\frac{1}{2}\|f-w^*-r^*\|_2^2,\\ &\liminf_{n\to\infty}TV(w_n)\ge TV(w^*).\end{aligned}$$

Hence, we must have

$$\limsup_{n\to\infty}\sum_{\varphi\in\mathcal{D}}|\langle r_n,\varphi\rangle|\le\sum_{\varphi\in\mathcal{D}}|\langle r^*,\varphi\rangle|,$$

but still according to Proposition 4.1, we have

$$\liminf_{n\to\infty}\sum_{\varphi\in\mathcal{D}}|\langle r_n,\varphi\rangle|\ge\sum_{\varphi\in\mathcal{D}}|\langle r^*,\varphi\rangle|.$$

Therefore, the only possible choice is that

$$\limsup_{n\to\infty}\sum_{\varphi\in\mathcal{D}}|\langle r_n,\varphi\rangle|=\liminf_{n\to\infty}\sum_{\varphi\in\mathcal{D}}|\langle r_n,\varphi\rangle|=\sum_{\varphi\in\mathcal{D}}|\langle r^*,\varphi\rangle|,$$

i.e.,

$$\lim_{n\to\infty}\sum_{\varphi\in\mathcal{D}}|\langle r_n,\varphi\rangle|=\sum_{\varphi\in\mathcal{D}}|\langle r^*,\varphi\rangle|.$$

By Lemma 4.1, (r_n) converges strongly to r^*. Similarly, we can prove that

$$\lim_{n\to\infty}\|f-w_n-r_n\|_2=\|f-w^*-r^*\|_2.$$

It is well known that in any Hilbert space, if (g_n) converges weakly to g^* and

$$\lim_{n\to\infty}\|g_n\|_2=\|g^*\|,$$

then g_n converges strongly to g^*. Based on this observation, $(f-w_n-r_n)$ converges strongly to $f-w^*-r^*$. Moreover (w_n) converges strongly to w^*. □

5 Experiments

In this section, we discuss the performance of our proposed model for image denoising problems. The signal-to-noise-ratio (SNR) is defined as

$$SNR(g,\tilde{g})=20\log_{10}\Big(\frac{\|g\|_2}{\|g-\tilde{g}\|_2}\Big),$$

where $g,\tilde{g}$ are the original and observed image, together with the visual effect are used to assess the quality of the restored image. The computer that we use is of Linux system (Mandrake 10) with Intel Core Quand CPU 2.66 GHz and 3.48G RAM. We code in C under Megawave2 (see "`http://megawave.cmla.ens-cachan.fr/`"). The outer iteration of our main algorithm is set as 10. We compare the proposed model (hybrid one) with the wavelet shrinkage and the ROF method. In all these experiments, the noise stand deviation σ is fixed as 20 and the SNR of the noisy image varies accordingly. In order to escape from the jungle of parameter tuning, we fix $\beta=50$ which corresponds to $\lambda=1/2\beta=0.01$, when one uses the standard form of the ROF model

$$\min_{w\in\mathcal{H}}TV(w)+\lambda\|f-w\|_2^2.$$

Moreover, we take $\alpha=3\sigma=60$.

In this paper, we use the Daubechies-4 wavelet basis of level 3 due to its simplicity. Note that the tuning of the different wavelet basis may lead better result though this is beyond the scope of the current paper. Roughly speaking, a wavelet basis adaptively (in the sense of sparse representation) to the ideal image is in favoring of better result. We refer the reader to [27, 35, 36] for some discussion related to basis choosing (or in more general settings, dictionary choosing).

The first experiment is for the Barbara image. The leg part of this image and the SNR value for the whole image are shown in Fig. 1. The Gibbs phenomenon of the wavelet shrinkage is rather clear and the ROF model suffers from the washout effect. However, for our new approach, we observe better texture reconstructions and the Gibbs oscillations are totally removed. The CPU time is about 30 seconds for our algorithm.

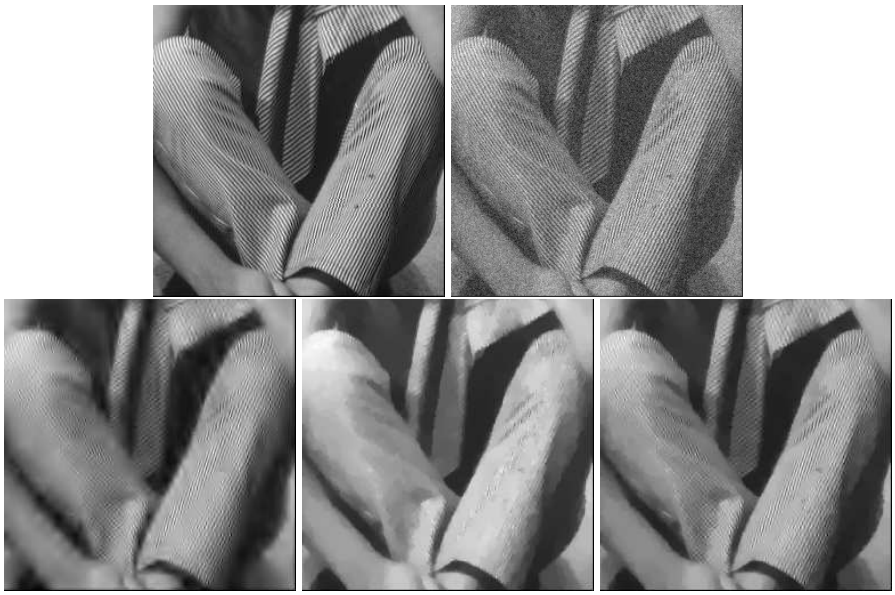

Figure 1: Leg part of Barbara and SNR for the total image of Barabara. Top-left: clean image; top-right: noisy image degraded by Gaussian noise of standard deviation 20, $SNR=8.70$; bottom-left: result of the wavelet soft-shrinkage with $\alpha=50$, $SNR=10.57$; bottom-middle: result of the ROF model with $\lambda=0.01$, $SNR=10.87$; bottom-right: new model with $\alpha=50$, $\beta=60$, $SNR=12.22$.

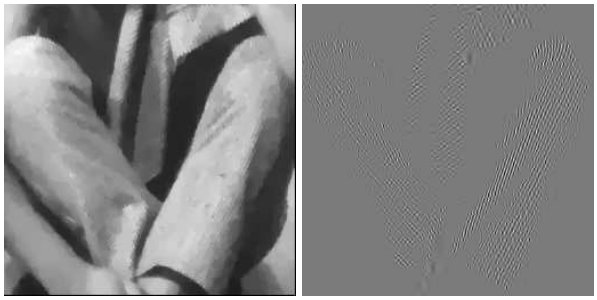

Figure 2: The cartoon part w_n (left) and the texture part r_n (right) for the leg part of the Barbara image.

In order to better understand the basic idea of our hybrid model, we also display the two components w_n,r_n of the result image of our approach in Fig. 2. One can discover that the cartoon part w_n is rather smooth and the detail part r_n is composed of textures captured by the wavelet basis. This is indeed expected by using our model.

Our next example is for the Mandrill image. The original image, the noisy image and the restoration results of the wavelet shrinkage method, the ROF model and the new approach are displayed in Fig. 3. This time, the Gibbs oscillation of the wavelet shrinkage is less prominent than the previous two examples due to the fact that the original image has texture nearly everywhere. But it still loses some important information (see face part, for instance) as well as the ROF model. And it is rather evident that the new approach is

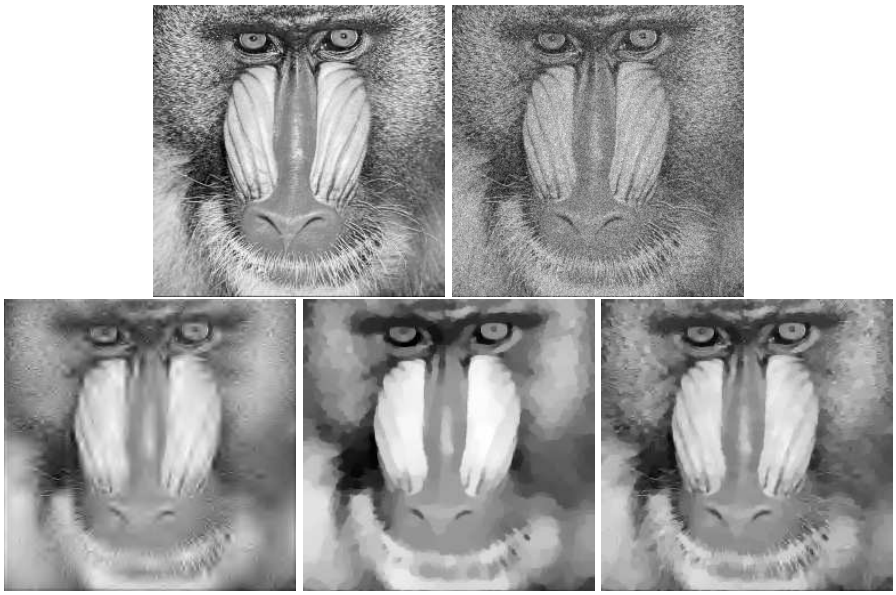

Figure 3: Experiments on the Mandrill image. Top-left: clean image; top-right: noisy image degraded by Gaussian noise of standard derivation 20, $SNR=5.77$; bottom-left: result of wavelet soft-shrinkage with $\alpha=50$, $SNR=5.97$; bottom-middle: result of ROF, $SNR=6.23$ with $\lambda=0.01$; bottom-right: result of new model with $\alpha=50$, $\beta=60$, $SNR=7.45$.

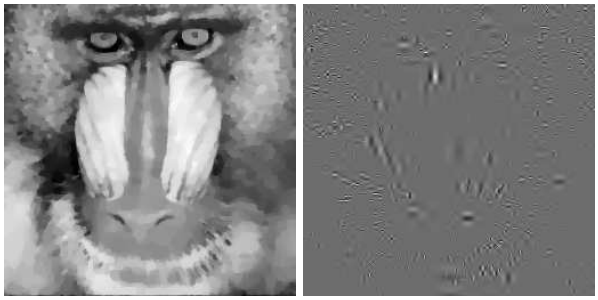

Figure 4: The Mandrill components: cartoon part w_n (left) and detail part r_n (right).

better since it recovers better details (see face part and beard part, for example). The CPU time is about 7 seconds for our algorithm.

The components w_n,r_n of the result image are exposed in Fig. 4. The cartoon part w_n is rather smooth and the texture part r_n contains some significant information which has a rather high total variation value.

The third example that we want to discuss is to the Boat image. The clean, noisy images and the restoration images by wavelet shrinkage, ROF and the new approaches are exhibited in Fig. 5. The result of wavelet shrinkage has visible Gibbs oscillation particularly in those rather flat regions and the ROF loses some crucial details due to the smoothing effect. However, our new approach can recover more details (see the cords

Figure 5: Experiments on the Boat image. Top-left: clean image; Top-right : noisy image (SNR is 10.16). Bottom-left: result of wavelet shrinkage (SNR is 11.96); Middle: result of ROF (SNR is 13.79); Right: result of our new model: (SNR is 15.67).

Figure 6: The components of the Boat image: the cartoon part (left) and the detail part (right).

and letters for example) without any perceptible Gibbs artifact. The CPU time is about 8 seconds for our algorithm.

The two components of the result image of Boat are presented in Fig. 6. Again, the cartoon part w_n is rather smooth and the detail part r_n captures some higher frequency information missed by the cartoon part.

At last we report the experiment on the Cameraman image. The clean image and noisy image are demonstrated on the top of Fig. 7. The results of the wavelet shrinkage method, the ROF model and the new approach are shown on the bottom of the same figure. One can see that the Gibbs effect of wavelet shrinkage is rather clear; the result of ROF is rather clean but some details are lost. Our new approach avoids entirely the Gibbs oscillation and the details (the ground, for instance) are better recovered. This can also

Figure 7: Experiments on Cameraman. Top-left: clean image; top-right: noisy image degraded by Gaussian noise of standard derivation 20, $SNR=9.87$; bottom-left: wavelet soft-shrinkage with $\alpha=50$, $SNR=11.89$; bottom-middle: ROF, $SNR=13.61$; bottom-right: new model with $\alpha=50$, $\beta=60$, $SNR=15.73$.

be confirmed if one compares the SNR values of these images. The CPU time is about 7 seconds for our algorithm.

The components w_n,r_n of the result image are also reported in Fig. 8. In this case, though the Cameraman image has less texture than the Barbara image, the wavelet filters still can extract some important information into the detail image.

As an interesting example recommended by the anonymous reviewers, here we also compare our approach with Meyer's model to show their difference. This model was first proposed in [28] and later studied in [31] as an important image decomposition method to replace the classical ROF model. The Meyer's model decomposes an image as two components: a cartoon part which has small TV value, and a detail part mainly containing noise and texture information which can be controlled by the so called G-norm (see [28, 31] for details). Theoretically, when the ideal image does not contain too much texture information, this model can also be used as an image denoising model since the detail part will mainly contain noise information. Hence, we use the Cameraman image for comparison which is indeed in favoring of Meyer's model. The noisy image to be decomposed/denoised is the same as top-right of Fig. 7. We tune the parameter of Meyer's model to obtain a reasonable separation. The cartoon part and detail part of the Meyer's model are shown in top of Fig. 9; the result of our model and its residual are shown in bottom of Fig. 9. Clearly, some high frequency information is decomposed into detail part of Meyer's model (this is expected by the model) and our denoising result contains more information (see house part for instance). This fact is further confirmed by the comparison of SNR values.

Figure 8: The cartoon part w_n (left) and the detail part r_n (right) for the Boat image.

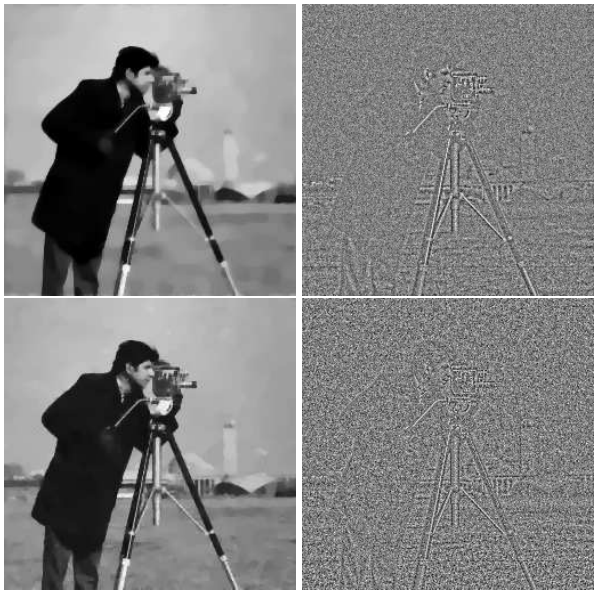

Figure 9: Decomposition experiments on Cameraman. Top-left: cartoon part of Meyer's model, $SNR=13.24$; texture and noise part of Meyer's model; bottom: our model with $\alpha=50$, $\beta=60$, $SNR=15.73$ and its residual.

6 Conclusions

In this paper, we proposed a hybrid variational model which takes advantages of the wavelet shrinkage and the total variation model. An iterative algorithm based on the alternating minimization direction idea was then presented to minimize the hybrid model. The strong convergence of the proposed algorithm was carefully proved. Finally, various numerical results were reported to exhibit the performance of the new model and it is rather clear that this model can recover better details than the usual ROF model and avoids efficiently the Gibbs oscillation of the wavelet shrinkage method. Moreover, this model might be extended to the de-convolution case though one needs extra efforts on the numerical optimization issue. We leave this for interested readers.

Acknowledgments

This work is supported by RGC 203109, RGC 201508, the FRGs of Hong Kong Baptist University, and the PROCORE- France/Hong Kong Joint Research Scheme sponsored by the Research Grant Council of Hong Kong and the Consulate General of France in Hong Kong F-HK05/08T. The authors would like to thank the helpful comments of the anonymous reviewers and the discussion with Prof. Alain Trouve and Dr. Pierre Weiss.

References

[1] Luigi Ambrosio, Nicola Fusco and Diego Pallara, Functions of Bounded Variation and Free Discontinuity Problems, Oxford University Press, 2000.
[2] Gilles Aubert and Pierre Kornprobst, Mathematical Problems in Image Processing, Springer, 2002.
[3] J. F. Aujol, G. Aubert, L. Blanc-Ferraud and A. Chambolle, Image decomposition into a bounded variation component and an oscillating component, J. Math. Imaging. Vision., 22 (1) (2005), 71–88.
[4] Heinz Bauer and Robert B. Burckel (Translator), Measure and Integration Theory, Volume 26 of De Gruyter Studies in Mathematics, Walter de Gruyter, Berlin, New York, 2001.
[5] E. Candes and D. L. Donoho, Curvelets-a surprisingly effective nonadaptive respresentation for objects with edges, In Curves and surfaces, pages 105–120, Vanderbilt University Press, Nashville, TN, US, 2000.
[6] A. Chambolle, An algorithm for total variation minimization and applications, J. Math. Imaging. Vision., 20 (1-2) (2004), 89–97.
[7] A. Chambolle and P.-L. Lions, Image recovery via total minimization and related problems, Numer. Math., 76 (1997), 167–188.
[8] Antonin Chambolle, Total variation minimization and a class of binary MRF models, in Energy Minimization Methods in Computer Vision and Pattern Recognition, Lecture Notes in Computer Science, 3757 (2005), 136–152. Springer Berlin/Heidelberg.
[9] T. Chan and H. M. Zhou, Total variation improved wavelet thresholding in image compression, In Proceedings to the International Conference on Image Processing, Vancouver, BC, Canada, Sept. 10-13, 2000.
[10] T. F. Chan and H. M. Zhou, Optimal construction of wavelet coefficients using total variation regularization in image compression, Technical Report CAM 00-27, University of California Los Angeles, July 2000.
[11] Tony F. Chan and Ke Chen, An optimization-based multilevel algorithm for total variation image denoising, Multiscale. Model. Sim., 5 (2) (2006), 615–645.
[12] S. Chang, B. Yu and M. Vetterli, Spatially adaptive wavelet thresholding with context modeling forimage denoising, IEEE. Trans. Image. Process., 9 (9) (2000), 1522–1531.
[13] Patrick L. Combettes and Valérie R. Wajs, Signal recovery by proximal forward-backward splitting, Multiscale. Model. Sim., 4 (4) (2005), 1168–1200.
[14] I. Daubechies, M. Defrise and C. De Mol, An iterative thresholding algorithm for linear inverse problems with a sparsity constraint, Comm. Pure. Appl. Math., 57 (11) (2004), 1413–1457.
[15] D. L. Donoho, De-noising by soft-thresholding, IEEE. Trans. Inf. Theor., 41 (3) (1995), 613–627.

[16] D. L. Donoho and I. M. Johnstone, Ideal spatial adaptation by wavelet shrinkage, Biometrika., 81 (3) (1994), 425–455.
[17] E. J. Candes, L. Demanet, D. L. Donoho and L. Ying, Fast discrete curvelet transforms, Multiscale. Model. Sim., 5 (2005), 861–899.
[18] F. Malgouyres, A noise selection approach of image restoration, SPIE, Internatinal Conference on Wavelets IX, Unser Laine and Aldroubi, Eds., San Diego, 4473 (2001), 34–41.
[19] A. Jalobeanu, L. Blanc-féraud and J. Zerubia, Satellite image deblurring using complex wavelet packets, Int. J. Comput. Vision., 51 (3) (2003), 205–217.
[20] F. Malgouyres, Minimizing the total variation under a general convex constraint for image restoration, IEEE, Trans. Image. Process., 11 (12) (2002), 1450–1456.
[21] M. Amiri, Z. Azimifar and P. Fieguth, Correlated non-linear wavelet shrinkage, Proceedings to the International Conference on Image Processing, 2008, 2348–2351.
[22] M. Ng, L. Qi, Y. Yang and Y. Huang, On semismooth Newton's methods for total variation minimization, J. Math. Imaging. Vision., 27 (2007), 265–276.
[23] R. R. Coifman and A. Sowa, Combining the calculus of variations and wavelets for image enhancement, Appl. Comput. Harmon. Anal., 9 (2000), 1–18.
[24] L. Rudin, S. Osher and E. Fatemi, Nonlinear total variation based noise removal algorithms, Phys. D., 60 (1992), 259–268.
[25] S. Durand and J. Froment, Reconstruction of wavelet coefficients using total varaition minimization, SIAM. J. Sci. Comput., 24 (2003), 1754–1767.
[26] S. Lintner and F.Malgouyres, Solving a variational image restoration model which involves l^∞ contraints, Inverse. Probl., 20 (3) (2004), 815–831.
[27] S. Mallat, A Wavelet Tour of Signal Processing, Academic Press, 1999 (2nd ed.).
[28] Yves Meyer, Oscillating Patterns in Image Processing and Nonlinear Evolution Equations, University Lecture Series (Vol. 22), American Mathematical Society, Providence, RI, 2001. Note: The fifteenth Dean Jacqueline B. Lewis memorial lectures.
[29] J.-L. Starck, M. Elad and D. L. Donoho, Image decomposition via the combination of sparse representation and a variational approach, IEEE Trans. Image Process., 14 (10) (2005), 1570–1582.
[30] Gabriele Steidl, Joachim Weickert, Thomas Brox, Pavel Mráze and Martin Welk, On the equivalence of soft wavelet shrinkage, total variation diffusion, total variation regularization, and sides, SIAM J. Numer. Anal., 42 (2) (2004), 686–713.
[31] Yang Wang and Haomin Zhou, Modeling textures with total variation minimization and oscillating patterns in image processing, J. Sci. Comput., 19 (1-3) (2003), 553–572.
[32] Yang Wang and Haomin Zhou, Total variation wavelet-based medical image denoising, Int. J. Biomedical. Imaging., 2006 (2006), 1–6.
[33] You-Wei Wen, Michael. Ng and Wai-Ki Ching, Iterative algorithms based on decoupling of deblurring and denoising for image restoration, SIAM J. Sci. Comput., 30 (5) (2008), 2655–2674.
[34] L. Woog, Adapted Waveform Algorithms for Denoising, PhD thesis, Yale University, April 1996. Available at `http://www.cs.yale.edu/homes/woog.html`.
[35] T. Zeng, Incorporating known features into a total variation dictionary model for source separation, Proceedings of IEEE ICIP'08., pp. 577–580, 2008, San Diego, California, U.S.A.
[36] T. Zeng and M. Ng, On the total variation dictionary model, IEEE Trans. Image Process., 19 (3) (2010), 821–825.