Multiscale hierarchical image decomposition and refinements: qualitative and quantitative results

Wen Li, Elena Resmerita* and Luminita Vese†

April 6, 2021

Abstract

The multiscale hierarchical decomposition method (MHDM) proposed in [37] has been proven very appropriate for denoising images with features at different scales and for scale separation. Extensions of it to image deblurring or to time-dependent settings [36]) have also been considered, showing convergence properties and more applications. The recent paper [27] fills in further qualitative results even for nonlinear problems and introduces a tighter version of MHDM with better convergence properties. The contribution of the present work is as follows. First, we derive novel error estimates for MHDM and its tighter version. Second, we provide rules for early stopping of the algorithms in the case of perturbed data, while still ensuring stable approximations of the true image. Last but not least, we propose a refined version of the tighter MHDM, which allows recovering structured images and promotes different features of the components, as compared to the entire image. The theoretical results are validated by numerous numerical experiments for image denoising and deblurring, which also assess the analyzed methods in terms of rate of convergence, stopping rule, and quality of restoration.

AMS Subject classification: 26B30, 46N10, 68U10

Key words: Natural images, multiscale expansion, ill-posed problem, image restoration.

1 Introduction

Decomposing an image in its structure and texture has been challenging both from mathematical modeling and computational viewpoints. Regarding modeling, the reader is first referred to the book [26, Ch. 1], where the advantage of using wavelet expansions versus Fourier series in image processing is clearly explained. The book also points out the limitations of some function spaces of positive differentiability when reconstructing natural images and proposes interesting alternatives, thus opening a door to numerous approaches.

1.1 Related work

The ROF total variation minimization model in [32], [33], so-called (BV, L^2), is the simplest and one of handiest decomposition method of an image $f \in L^2(\Omega)$, aiming to find $u_\lambda \in BV(\Omega)$ and $v_\lambda \in L^2(\Omega)$ that satisfy

$$(u_\lambda, v_\lambda) = \arg \min_{u+v=f} \left\{ \lambda \|v\|^2 + |u|_{BV(\Omega)} \right\}, \tag{1.1}$$

where $\Omega \subseteq \mathbb{R}^2$ is a bounded and open set. While this approach is known to recover edges in images well, it is however not suitable for denoising (possibly blurry) natural images containing texture at different scales.

Department of Mathematics, University of California at Los Angeles (UCLA), Los Angeles, CA 90095 (wenli@math.ucla.edu)

*Institute of Mathematics, Alpen-Adria Universität Klagenfurt, Universitätsstrasse 65–67, 9020 Klagenfurt, Austria (elena.resmerita@aau.at)

†Department of Mathematics, University of California at Los Angeles (UCLA), Los Angeles, CA 90095 (lvese@math.ucla.edu)

During the last two decades, finer and finer techniques have been developed to cope with the structure complexity of natural images. An essential step forward in this direction has been made with the (BV, G), (BV, F), (BV, E) models in [26, Ch. 1] that propose more appropriate norms for quantifying the oscillatory components of an image, in the sense that functions which may have large oscillations still have small norm (which was not the case when measuring their L^2 norms). This finding, although brilliant as a concept, was numerically burdensome because of the involved G, F, E spaces. Thus, it was naturally followed by various strategies to ease the computational effort - see the series of papers [6–10], [42], [29], [20], [25], where, e.g., helpful approximations of the above spaces were introduced. We would like to mention also several interesting approaches for image decomposition based on wavelets, such as [35] that deals with separating the texture from the piecewise smooth part by employing two suitable dictionaries, and [15, 16], which use the Besov space $B_1^1(L^1(\Omega))$ in the models [42], [29], instead of the BV space. More recent models based on higher order total variation functionals can be found in [40] and [34], the latter performing a decomposition into components with different degrees of temporal regularity. Actually, there is a rich literature on compelling variational image decomposition methods that separate cartoon and texture in clean images, or cartoon, texture and noise in noisy images, while iterative methods (e.g., the proximal type method [24]) seem to be seriously outnumbered. In this work, we restrict ourselves to the references that are closely related to the current investigation.

When it comes to more complex images, one could opt for a multiscale decomposition as in [31] or a decomposition of hierarchical nature as introduced in [37] - details will be given later in this section. Separating cartoon and texture is highly dependent on the scale λ from (1.1), in the sense that details in an image (usually part of the texture) can be seen as cartoon at a refined scale, such as 2λ. Based on this idea, the authors of [37] propose a representation of the input image as a sum of simpler images at different scales by the multiscale hierarchical method (MHDM). An analogous process can be done for blurred images, according to [38]. We would also like to recall the related prior work [36], were the authors expressed MHDM using a small parameter τ which goes to 0, and the BV components provided by the MHDM converge to the solution of an integro-differential equation. Unlike the "zooming in" concept mentioned above, the work [5] considers the reversed procedure in the context of weighted total variation flow, that is "zooming-out": One can first identify the smallest components in an image, with a fast denoising as a by-product, and gradually recover also the large features.

A recent development of the method in [37] can be found in the paper [39], having as starting model the variational decomposition involving additionally the $\|\cdot\|_*$ norm (G norm) for textures,

$$(u_\lambda, v_\lambda) = \arg\min_{u,v} \left\{ \lambda \|f - u - v\|^2 + |u|_{BV(\Omega)} + \|v\|_* \right\}, \tag{1.2}$$

rather than (1.1) which is the basis for the approaches in our work. Thus, the method of [39] aims at capturing both structure and texture at different scales.

Multiscale hierarchical decomposition techniques have been successfully applied to a plethora of problems related to graph theory [23], image registration [30], [27], compressed sensing, deconvolution of the Helmholtz filter and linear regression - see the PhD thesis [43].

1.2 Preliminaries

Let us focus on the multiscale hierarchical decomposition method from [37, 38], to get closer to the methods analyzed in this paper in the context of inverse problems arising in image processing. In fact, the framework is more general and can be employed for other inverse problems as well.

Let $T : L^2(\Omega) \to L^2(\Omega)$ be a bounded linear blurring operator which is ill-posed, where $\Omega \subseteq \mathbb{R}^2$ is a bounded and open set. Consider $f \in L^2(\Omega)$ a blurred image. Recovering the true image u reduces usually to solving the ill-posed equation

$$Tu = f. \tag{1.3}$$

Assume that $J : L^2(\Omega) \to [0, \infty]$ is a seminorm which is finite on a Banach space $X \subseteq L^2(\Omega)$. We can think of $X = BV(\Omega)$ and of J as the total variation seminorm, but the results hold also for more general spaces X and functions J. Let $\|\cdot\|$ denote the $L^2(\Omega)$ norm and $\langle \cdot, \cdot \rangle$ the $L^2(\Omega)$ inner product. In this context, solving the ill-posed problem (1.3) by regularization amounts to imposing a-priori constraints on the unknown u as a minimizer of

$$\min_{u \in X} \{\lambda \|Tu - f\|^2 + J(u)\}, \tag{1.4}$$

where $\lambda > 0$.

In the current framework of $X = BV(\Omega)$ and $J(u) = |u|_{BV(\Omega)}$, (1.4) is the classical total variation minimization problem. Existence of minimizers of (1.4) is guaranteed if T does not annihilate constants, as shown in [1].

Let T^* denote the adjoint operator of T and $\|\cdot\|_*$ the dual of the functional J with respect to the $L^2(\Omega)$ scalar product, defined by

$$\|g\|_* = \sup_{\substack{\varphi \in BV(\Omega) \\ J(\varphi) \neq 0}} \frac{\langle g, \varphi \rangle}{J(\varphi)} \in [0, \infty], \quad \forall g \in L^2(\Omega),$$

so that the usual duality holds:

$$\langle g, \varphi \rangle \leq J(\varphi) \|g\|_*.$$

It is possible to show that a minimizer of (1.4) can be characterized in terms of the dual norm $\|\cdot\|_*$ (see [38], inspired from [26] and [3]). Indeed, a minimizer u of (1.4) satisfies:

(i) $\|T^* f\|_* \leq \frac{1}{2\lambda}$ if and only if u is the trivial minimizer.

(ii) If $\frac{1}{2\lambda} < \|T^* f\|_* < \infty$ then: u is a minimizer if and only if $\|T^*(f - Tu)\|_* = \frac{1}{2\lambda}$ and $\langle u, T^*(f - Tu) \rangle = J(u) \frac{1}{2\lambda}$.

In any case, we have that $\|T^*(f - Tu)\|_*$ is finite for a minimizer u.

The multiscale hierarchical decomposition method proposed in [37, 38] starts with the model (1.4) and constructs a multiscale image decomposition in a hierarchical way, as described below.

Let λ_0 be a positive number and u_0 be a solution of

$$\min_{u \in X} \{\lambda_0 \|Tu - f\|^2 + J(u)\}.$$

For $k \geq 1$, define the sequence $(u_k) \subset X$ with u_k as a solution of

$$\min_{u \in X} \{\lambda_k \|Tu - v_{k-1}\|^2 + J(u)\}, \tag{1.5}$$

with $\lambda_k = 2^k \lambda_0$ and $v_{k-1} = f - \sum_{j=0}^{k-1} Tu_j$, and thus $f = Tu_0 + Tu_1 + \cdots + Tu_{k-1} + v_{k-1}$. By denoting $x_{k-1} = \sum_{j=0}^{k-1} u_j$, one can rewrite procedure MHDM (see also [27]) as

$$u_k \in \arg\min_{u \in X} \{\lambda_k \|T(u + x_{k-1}) - f\|^2 + J(u)\}, \; k \geq 0, \tag{1.6}$$

where $x_{-1} = 0$. It was shown in [38] that (Tx_k) converges to f in a weak sense as below,

$$\lim_{k \to \infty} \|T^*(f - Tx_k)\|_* = 0.$$

In the special case of $T = I$ (identity operator), the sequence (x_k) with $x_k = \sum_{j=0}^{k} u_j$ converges even strongly to the true image f in the $L^2(\Omega)$ norm if f lies in $X = BV(\Omega)$, as shown in [37].

One can actually obtain strong convergence of (x_k) to f in $L^2(\Omega)$ even when dropping the $BV(\Omega)$ regularity for f, as proven in the interesting and inspiring work [27], where very fine analysis results are shown in a more general setting. Moreover, [27] shows that (Tx_k) converges strongly to f also in the case $T \neq I$ and discusses the issue of convergence of the iterates x_k to a solution of $Tu = f$ when $T \neq I$. Furthermore, it introduces a tighter hierarchical decomposition method, where a penalization on $x_k = u_0 + \cdots + u_k$ is also enforced and as a consequence, convergence of (x_k) with respect to a natural metric in $BV(\Omega)$ is guaranteed. Namely, the tighter MHDM generates a sequence $(u_k) \subset X$ with

$$u_k \in \arg\min_{u \in X} \{\lambda_k \|T(u + x_{k-1}) - f\|^2 + \lambda_k a_k J(u + x_{k-1}) + J(u)\}, \; k \geq 0, \tag{1.7}$$

where $x_{k-1} = \sum_{j=0}^{k-1} u_j$ and $x_{-1} = 0$. Clear improvements over the original MHDM are highlighted in Subsections 7.1, 7.2 and 7.3.

In order to enable more details in the final restored image, we introduce a refinement of the latter method by choosing a variable R_k functional in the last term of (1.7), which is weaker than J. Thus, the refined tighter MHDM is formulated as

$$u_k \in \arg\min_{u\in X}\{\lambda_k\|T(u+x_{k-1})-f\|^2+\lambda_k a_k J(u+x_{k-1})+R_k(u)\},\quad k\geq 0, \tag{1.8}$$

where the parameters λ_k and a_k are chosen in a suitable way (see details in Section 5). As detailed theoretically in Section 5 and numerically in Subsection 7.4 respectively, this method also converges to the true image and provides pretty sharp reconstructions of edges and fine texture.

1.3 Contributions and outline of the current work

A first contribution of our work is providing convergence rates for the data-fitting term $\|Tx_k-f\|$, and thus for $\|x_k-f\|$ in the case of denoising, both for MHDM and for the tighter methods. It remains an open problem to establish convergence rates for $\|x_k-f\|$ for the deblurring case. In such a situation, one needs to employ a so-called source condition, that is a smoothness assumption on the solution (see, e.g. [11] and [18] for the Bregman iteration method).

As a second contribution, we analyze MHDM and its tighter version also in the case of noisy data, when a typical behavior for ill-posed problems is observed due to the data error propagation, namely the so-called semiconvergence: some distance between the iterates x_k and the solution has an initial decay, and then increases - compare to the Landweber (steepest descent) method at [17, pag. 156-157] or to the Bregman iteration (augmented Lagrangian method) in [28]. Therefore, in order to balance between accuracy and noise amplification, we propose earlier stopping of the multiscale decompositions by a posteriori rules (Morozov discrepancy principle) and by a priori rules.

Last but not least, we introduce and investigate a refinement of the tighter MHDM, in the sense that the penalizations on the components u_k and on the sums x_k might be allowed to be different, according to the features that need to be highlighted in each case.

The structure of this paper is as follows. After deriving error estimates for MHDM in Section 2, we provide in Section 3 rules for early stopping of this algorithm in the case of perturbed data, while still ensuring stable approximations of the true image (in the denoising case). Section 4 deals with similar aspects for the tighter version of MHDM, while the refined MHDM is presented in Section 5. Discretizations of the Euler-Lagrange equations corresponding to the tight and refined versions are proposed in Section 6. All theoretical results are validated in Section 7 by several numerical experiments for image denoising and deblurring.

2 Error estimates for MHDM

In this section we establish an error estimate for the residual $\|Tx_k-f\|$ occurring in MHDM formulated at (1.6), that is $\|Tx_k-f\|=O(1/\sqrt{k+1})$. Clearly, an immediate consequence is the strong convergence of (Tx_k) to f in $L^2(\Omega)$, thus providing an alternative proof to the one by contradiction from [27, Theorem 2.1].

Throughout this study, we assume that there is a solution $z\in X$ of (1.3).

It is known that (u_k) is well-defined (see [38]), although u_k might not be a unique solution of the corresponding optimization problem.

By the definition of u_k, one can write for any $k\geq 0$,

$$\lambda_k\|T(u_k+x_{k-1})-f\|^2+J(u_k)\leq\lambda_k\|T(u+x_{k-1})-f\|^2+J(u),\quad\forall u\in BV(\Omega). \tag{2.1}$$

Note that the following convention is used here: $x_{-1}=0$. By taking $u=0$ and then $u=z-x_{k-1}$, one obtains for any $k\geq 0$

$$\lambda_k\|Tx_k-f\|^2+J(u_k)\leq\lambda_k\|Tx_{k-1}-f\|^2, \tag{2.2}$$

$$\lambda_k\|Tx_k-f\|^2+J(u_k)\leq J(z-x_{k-1}). \tag{2.3}$$

Note that $J(0)=0$ yielded (2.2). Thus, one deduces from (2.2) the decreasing monotonicity of the residual $\|Tx_k-f\|^2$.

In the sequel, we show the announced error estimate.

Proposition 2.1. *The following estimate holds:*

$$\|Tx_k - f\|^2 \leq \frac{2J(z)}{(k+1)\lambda_0}, \quad \forall k \geq 0. \tag{2.4}$$

Consequently, $\lim_{k\to\infty} Tx_k = f$ *in the* $L^2(\Omega)$ *norm.*

Proof: Since J is subadditive and $u_k = x_k - x_{k-1}$, one has

$$J(z - x_k) - J(z - x_{k-1}) \leq J(u_k), \quad \forall k \geq 0.$$

By using this inequality in (2.3), one obtains $\lambda_k\|Tx_k - f\|^2 + J(z - x_k) \leq 2J(z - x_{k-1})$, that is

$$\|Tx_k - f\|^2 + \frac{1}{\lambda_k}J(z - x_k) \leq \frac{2}{\lambda_k}J(z - x_{k-1}) = \frac{1}{\lambda_{k-1}}J(z - x_{k-1}), \tag{2.5}$$

where $\dfrac{2}{\lambda_k} = \dfrac{1}{\lambda_{k-1}}$ if $k \geq 1$. By writing (2.5) for indices $0, 1, ..., k$ and summing up, one has for any $k \geq 0$

$$(k+1)\|Tx_k - f\|^2 + \frac{1}{\lambda_k}J(z - x_k) \leq \sum_{j=0}^{k}\|Tx_j - f\|^2 + \frac{1}{\lambda_k}J(z - x_k) \leq \frac{2}{\lambda_0}J(z), \tag{2.6}$$

where the left inequality follows from the monotonicity of the data fidelity term and the right one follows from $x_{-1} = 0$. This yields (2.4). □

Error estimates can be derived in the denoising case under more regularity on the true image.

Corollary 2.2. *Assume that* $T = I$ *and* $f \in BV(\Omega)$. *Then*

$$\|x_k - f\| \leq \sqrt{\frac{2J(f)}{(k+1)\lambda_0}} = O(1/\sqrt{k+1}), \quad \forall k \geq 0, \tag{2.7}$$

and thus, $\lim_{k\to\infty} x_k = f$ *in* $L^2(\Omega)$.

Proof: One applies Proposition 2.1, where f plays the role of the solution $z \in BV(\Omega)$. □

Note that the upper bounds in (2.4) and (2.7) depend on the initial parameter λ_0, implying that too small values of λ_0 yield quite large upper bounds, which are not desirable. This fits the discussions in [37, Section 2] and [38, Section 2] that recommend to choose rather a larger starting parameter, since it will generate proper decompositions of the original image, as opposed to the case of a small λ_0 that will not be effective for a number of iterations, until a large enough value for $\lambda_k = 2^k\lambda_0$ will be reached. The reader is referred also to the comments on the parameters λ_k in Subsections 7.1 and 7.2.

Remark 2.3. *For the sake of completeness, we prove below strong convergence of* (Tx_k) *to* f *in the case of deblurring* ($T \neq I$) *by using the techniques developed in the original paper [37] when* $T = I$.

First, one can check that

$$\sum_{j=0}^{k}\left[\|Tu_j\|^2 + \frac{1}{\lambda_j}J(u_j)\right] = \|f\|^2 - \|v_k\|^2, \quad k \geq 0 \tag{2.8}$$

Note that $v_{2k-1} = Tu_{2k} + v_{2k}$ *with* $Tu_{2k} \to 0$ *as* $k \to \infty$ *due to convergence of the series with term* $\|Tu_j\|^2$ *- see* (2.8). *Thus, in order to show that* $v_k \to 0$ *as* $k \to \infty$, *it is enough proving that* $v_{2k} \to 0$ *as* $k \to \infty$, *as this would imply that also* $v_{2k-1} \to 0$.

It can be readily seen that the following holds:

Since $v_k = \sum_{j=k+1}^{2k} Tu_j + v_{2k}$, $k \geq 1$, *one has* $(v_{2k}, v_{2k}) = (v_{2k}, v_k) - \sum_{j=k+1}^{2k}(v_{2k}, Tu_j)$. *Thus,*

$$\begin{aligned} |(v_{2k}, v_k)| = |(v_{2k}, f - Tx_k)| &= |(T^*v_{2k}, z - x_k)| \\ &\leq \frac{1}{\lambda_{2k}}[J(z) + J(x_k)] \\ &\leq \frac{1}{\lambda_{2k}}J(z) + \frac{\lambda_k}{\lambda_{2k}}\sum_{j=k+1}^{2k}\frac{1}{\lambda_j}J(u_j). \end{aligned}$$

The last sum converges to zero for $k \to \infty$ due to convergence of the series with the corresponding term (see (2.8)). Now,

$$\left| \sum_{j=k+1}^{2k} (v_{2k}, Tu_j) \right| \leq \sum_{j=k+1}^{2k} |(T^* v_{2k}, u_j)| \leq \frac{1}{\lambda_{2k}} \sum_{j=k+1}^{2k} J(u_j) \leq \sum_{j=k+1}^{2k} \frac{1}{\lambda_j} J(u_j),$$

which concludes the proof.

Remark 2.4. *Problem (1.6) can be equivalently written as*

$$\min_{u \in X} \{\lambda_k \|Tu - f\|^2 + \lambda_k \langle u, T^* T x_{k-1} \rangle + J(u)\}, \tag{2.9}$$

which involves a Tikhonov regularization functional and a linear term depending on all the previous iterates. Thus, the computational effort for this problem is similar to the one for the corresponding variational regularization with no additional linear term.

3 MHDM with perturbed data

In the sequel we shall consider MHDM with noisy data. As mentioned in the Introduction, one has to stop early the algorithm, due to data error propagation. First, we shall derive error estimates and then propose stopping rules which ensure convergence of the residual to zero, when the noise level tends to zero. The reader is referred also to [5] regarding a stopping rule for the continuous time, by means of the peak signal-to-noise ratio.

While f denotes the exact data (the blurred image without noise), let $f^\delta \in L^2(\Omega)$ stand for the noisy data (the blurred image with noise). That is, the following model is considered

$$f^\delta = Tz + \eta \tag{3.1}$$

where η is additive Gaussian noise, z is the image to be recovered and

$$\|f - f^\delta\| \leq \delta, \quad \delta > 0. \tag{3.2}$$

The method under investigation generates a sequence $(u_k) \subset X$, where

$$u_k \in \arg\min_{u \in X} \{\lambda_k \|Tu - v_{k-1}\|^2 + J(u)\}, \tag{3.3}$$

with $\lambda_k = 2^k \lambda_0$ and $v_{k-1} = f^\delta - \sum_{j=0}^{k-1} Tu_j$, and $f^\delta = Tu_0 + Tu_1 + \cdots + Tu_{k-1} + v_{k-1}$. By denoting again $x_{k-1} = \sum_{j=0}^{k-1} u_j$, one has to find $u_k \in BV(\Omega)$ as a minimizer of

$$\lambda_k \|T(u + x_{k-1}) - f^\delta\|^2 + J(u). \tag{3.4}$$

We shall consider first an a posteriori stopping criterion (that is, depending on the noise level δ and the measured data f^δ), according to which the iterative procedure is stopped when the norm of the misfit is compatible with the noise level. Secondly, an a priori choice (depending only on δ) will be presented. The former type of rule is usually preferred, since it takes into account the knowledge on the given data.

3.1 Discrepancy principle stopping rule

Let the following index depending on δ and f^δ be defined as

$$k^*(\delta) := \max\{k \in \mathbb{N} : \|Tx_k - f^\delta\|^2 \geq \tau \delta^2\}, \quad \text{for some } \tau > 1. \tag{3.5}$$

We will see below that $k^*(\delta)$ is well-defined.

Proposition 3.1. *If* (3.2) *is satisfied, then the following estimate holds*

$$\|Tx_k - f^\delta\|^2 \leq \delta^2 + \frac{2J(z)}{(k+1)\lambda_0}, \quad \forall k \geq 0 \tag{3.6}$$

and the stopping index defined by (3.5) *is finite. If* $(k^*(\delta))$ *is unbounded as* $\delta \to 0$*, then one has* $\lim_{\delta\to 0} Tx_{k^*(\delta)} = f$ *in the* $L^2(\Omega)$ *norm.*

Proof: One has for any $k \geq 0$,

$$\lambda_k\|T(u_k + x_{k-1}) - f^\delta\|^2 + J(u_k) \leq \lambda_k\|T(u + x_{k-1}) - f^\delta\|^2 + J(u), \quad \forall u \in BV(\Omega). \tag{3.7}$$

Again, by setting $u = 0$ and then $u = z - x_{k-1}$, it follows that $\|Tx_k - f^\delta\|^2$ decreases when k increases and

$$\|Tx_k - f^\delta\|^2 + \frac{1}{\lambda_k}J(u_k) \leq \|f - f^\delta\|^2 + \frac{1}{\lambda_k}J(z - x_{k-1}) \leq \delta^2 + \frac{1}{\lambda_k}J(z - x_{k-1}). \tag{3.8}$$

Based on the summing up technique from the proof of Proposition 2.1, one obtains for any $k \geq 0$,

$$(k+1)\|Tx_k - f^\delta\|^2 + \frac{1}{\lambda_k}J(z - x_k) \leq (k+1)\delta^2 + \frac{2}{\lambda_0}J(z), \tag{3.9}$$

which yields (3.6).

Note that $k^*(\delta)$ is well-defined, since $\|Tx_k - f^\delta\|^2$ decreases. Moreover, it is finite, as shown next. By writing the last inequality for $k = k^*(\delta)$, using (3.5) and neglecting a nonnegative term, it follows that

$$(k^*(\delta) + 1)\tau\delta^2 \leq (k^*(\delta) + 1)\delta^2 + \frac{2}{\lambda_0}J(z) \tag{3.10}$$

and thus, the stopping index is finite:

$$k^*(\delta) \leq \frac{2J(z)}{\lambda_0(\tau - 1)\delta^2} - 1. \tag{3.11}$$

If $(k^*(\delta))$ is unbounded, then (3.6) written for $k = k^*(\delta)$ implies $\lim_{\delta\to 0} Tx_{k^*(\delta)} = f$. □

An immediate consequence of the proposition above is convergence of the reconstructions to the original image, when the noise level tends to zero.

Corollary 3.2. *Assume that* $T = I$ *and* $f \in BV(\Omega)$*. Then* $\lim_{\delta\to 0} x_{k^*(\delta)} = f$.

3.2 A priori stopping rule

In the inverse problems literature, a priori rules are of interest in contexts where they yield better convergence rates than the a posteriori rules. Although it is not clear if this is the case for MHDM, we also present an a priori stopping rule next, for the sake of completeness.

Corollary 3.3. *If the stopping index is chosen as* $k^*(\delta) \sim \frac{1}{\delta^2}$*, then* $\|Tx_{k^*(\delta)} - f\| = O(\delta)$*. Consequently,* $\|x_{k^*(\delta)} - f\| = O(\delta)$ *in the denoising case, if* $f \in BV(\Omega)$*.*

Proof: One writes (3.6) for $k = k^*(\delta)$ and applies the choice $k^*(\delta) \sim \frac{1}{\delta^2}$. □

4 A tighter multiscale hierarchical decomposition method

This section is devoted to the tighter MHDM defined at (1.7). Error estimates and convergence results in both the exact and the noisy data cases are established in the spirit of the previous two sections, that is, concerning the residual. An additional opportunity arises regarding the tighter method, namely showing convergence of the sums x_k of the components u_j to the true image in the general context of deblurring.

The paper [27] emphasizes that, besides some special situations (e.g. $T = I$), one usually has only convergence of type $Tx_k \to f$, while convergence of the iterates x_k to a solution of $Tu = f$ is not guaranteed. This motivates the authors of [27] to propose the tighter MHDM with the merit that the sum of the generated components does converge to such a solution. The reader is referred also to [43] regarding upper bounds for the error between the solution of the operator equation and the components sum x_k, when (x_k) does not necessarily converge to that solution.

Let (a_k) be a sequence of nonnegative numbers such that

$$\lim_{k\to\infty} a_k = 0 \quad \text{and} \quad a_k \le a_{k-1},\ \forall k \ge 1. \tag{4.1}$$

Let λ_0 be a positive number and u_0 be a solution of

$$\min_{u\in X}\{\lambda_0\|Tu - f\|^2 + \lambda_0 a_0 J(u) + J(u)\},$$

with J and X as above.

Construct a sequence $(u_k) \subset X$ with u_k as a solution of

$$\min_{u\in X}\{\lambda_k\|T(u + x_{k-1}) - f\|^2 + \lambda_k a_k J(u + x_{k-1}) + J(u)\}. \tag{4.2}$$

where $x_{k-1} = \sum_{j=0}^{k-1} u_j$ and $(\lambda_k) \subset (0, \infty)$ satisfies the relaxed inequality

$$2\lambda_{k-1} \le \lambda_k, \quad \forall k \ge 0, \tag{4.3}$$

rather than the equality $2\lambda_{k-1} = \lambda_k$. We denote again $v_{k-1} = f - \sum_{j=0}^{k-1} Tu_j$, and thus $f = Tu_0 + Tu_1 + \cdots + Tu_{k-1} + v_{k-1}$.

According to [27], the sequence (Tx_k) converges to f if

$$\limsup_{k\to\infty} \frac{2^k}{\lambda_k} < \infty. \tag{4.4}$$

Note that the following inequalities hold $\forall k \ge 0$ (similarly to the case of the original method when $a_k = 0$),

$$\lambda_k\|Tx_k - f\|^2 + \lambda_k a_k J(x_k) + J(u_k) \le \lambda_k\|Tx_{k-1} - f\|^2 + \lambda_k a_k J(x_{k-1}), \tag{4.5}$$

$$\lambda_k\|Tx_k - f\|^2 + \lambda_k a_k J(x_k) + J(u_k) \le \lambda_k a_k J(z) + J(z - x_{k-1}). \tag{4.6}$$

It is interesting to note that inequality (4.5) implies the decreasing monotonicity of the residual-like $\|Tx_k - f\|^2 + a_k J(x_k)$ rather than of the residual $\|Tx_k - f\|^2$, since $a_k \le a_{k-1}$ and $J(u_k) \ge 0$ for any $k \ge 0$.

We show below an error estimate for the residual and hence, for $\|x_k - f\|$ in the case of denoising, based on the summability of (a_k) and on (4.3), the latter implying (4.4). First, denote

$$\sum_{k=0}^{\infty} a_k < \infty. \tag{4.7}$$

Proposition 4.1. *Let conditions* (4.1) *and* (4.3) *be satisfied. Then the following estimate holds:*

$$\|Tx_k - f\|^2 + a_k J(x_k) \le \left(\sum_{j=0}^{k} a_j\right) \frac{J(z)}{k+1} + \frac{2J(z)}{(k+1)\lambda_0}, \quad \forall k \ge 0. \tag{4.8}$$

Moreover, if (4.7) *is verified, then* $\lim_{k\to\infty} Tx_k = f$ *holds. If additionally* $f \in BV(\Omega)$ *and* $T = I$, *then* $\lim_{k\to\infty} x_k = f$ *in* $L^2(\Omega)$.

Proof: The proof resembles the one for Proposition 2.1, by taking into account that $\frac{2}{\lambda_k} \leq \frac{1}{\lambda_{k-1}}$. □

Recall that $z \in X$ is a J minimizing solution of $Tu = f$ if

$$z \in \arg\min_{Tx=f} J(x).$$

Additionally, recall that the metric $d : BV(\Omega) \times BV(\Omega) \to [0, \infty)$ given by

$$d(u, v) = \|u - v\|_{L^1} + |J(u) - J(v)| \tag{4.9}$$

provides the so-called strict convergence for images in the $BV(\Omega)$ space cf. [2, p. 125], which incorporates both the weak* convergence and the one with respect to the $|J(u) - J(v)|$ term in (4.9) - see also [18]. Namely, the following holds:

$$d(u_k, u) \to 0 \quad \text{if and only if} \quad (u_k \overset{w^*}{\to} u \ \text{ and } \ J(u_k) \to J(u)).$$

A few considerations concerning the parameters involved in the tighter version and their role in deriving convergence of the iterates are in order. Estimate (4.8) shows that

$$J(x_k) \leq \left(\sum_{j=0}^{k} a_j \right) \frac{J(z)}{a_k(k+1)} + \frac{2J(z)}{a_k(k+1)\lambda_0}, \quad k \geq 0,$$

when $a_k > 0$, $\forall k \geq 0$, where z solves $Tu = f$. In order for $(J(x_k))$ to be bounded, one needs not only $\sum_{k=0}^{\infty} a_k < \infty$, but also boundedness of the sequence with term $\frac{1}{a_k(k+1)}$, that is $\frac{1}{k+1} = O(a_k)$, which yields divergence of the series with term a_k (not desired). This suggests that a stronger condition on the parameters is indeed necessary to ensure boundedness of $(J(x_k))$ and convergence of (x_k) on subsequences.

As shown in [27, Theorem 2.5], the stronger condition

$$\limsup_{k\to\infty} \frac{2^k}{\lambda_k a_k} = 0, \tag{4.10}$$

with $(a_k) \subset (0, \infty)$, guarantees weak convergence of (x_k) (on subsequences) to a solution $\hat{z}$ of $Tu = f$, which minimizes J over the set of solutions. Moreover, $J(x_k) \to J(\hat{z})$ holds on subsequences.

For the sake of completeness, we formulate in the sequel the above mentioned convergence result for the iterates (x_k).

Proposition 4.2. *Let conditions* (4.1), (4.3) *and* (4.10) *be satisfied. Then the sequence* (x_k) *defined above converges on subsequences to a* J*- minimizing solution of* $Tu = f$ *in the sense of the metric* (4.9).

Proof: Recall the main argument of the proof for [27, Theorem 2.5],

$$J(z - x_k) \leq 2^{k+1} J(z), \quad k \geq 0. \tag{4.11}$$

Using (4.11) in (4.6) yields

$$J(x_k) \leq J(z) + \frac{J(z - x_{k-1})}{\lambda_k a_k} \leq J(z) + \frac{2^k J(z)}{\lambda_k a_k}, \quad k \geq 0, \tag{4.12}$$

which implies $\limsup_{k\to\infty} J(x_k) \leq J(z)$. Now boundedness of $J(x_k)$ and (Tx_k) in the total variation setting implies boundedness of (x_k) in $BV(\Omega)$, and thus, existence of a subsequence (denoted again (x_k)) which converges weakly* to some $\bar{z} \in BV(\Omega)$. With standard $BV(\Omega)$ arguments, one can assume that the subsequence converges also weakly in $L^2(\Omega)$. Since T is linear and bounded, it is also weakly continuous. Therefore, (Tx_k) converges weakly to $T\bar{z}$. This must be equal to f, as $Tx_k \to f$. Thus, $\bar{z}$ is also a solution of $Tu = f$. Note that the inequality $\liminf_{k\to\infty} J(x_k) \geq J(\bar{z})$ follows from the lower semicontinuity of J. Since $\limsup_{k\to\infty} J(x_k) \leq J(\bar{z})$ (by setting the solution $z = \bar{z}$ in the corresponding inequality above), one has $\lim_{k\to\infty} J(x_k) = J(\bar{z})$ on a subsequence of (x_k). Actually, $\bar{z}$ is a J minimizing solution of $Tu = f$, since $J(\bar{z}) \leq \liminf_{k\to\infty} J(x_k) \leq \limsup_{k\to\infty} J(x_k) \leq J(z)$, whenever $Tz = f$. Thus, a subsequence of (x_k) converges strictly in $BV(\Omega)$ to a J minimizing solution of $Tu = f$. □

4.1 The tighter version of the MHDM with perturbed data

Suppose that noisy data f^δ satisfying (3.2) are given.

Thus, one aims at finding minimizers $u_k \in BV(\Omega)$ of

$$\lambda_k \|T(u + x_{k-1}) - f^\delta\|^2 + \lambda_k a_k J(u + x_{k-1}) + J(u) \tag{4.13}$$

and at investigating convergence properties for (Tx_k) and (x_k). Consider $a_k > 0$, for any $k \geq 0$.

From (4.13) it follows that

$$\lambda_k \|Tx_k - f^\delta\|^2 + \lambda_k a_k J(x_k) + J(u_k) \leq \lambda_k \|Tx_{k-1} - f^\delta\|^2 + \lambda_k a_k J(x_{k-1}), \tag{4.14}$$

$$\lambda_k \|Tx_k - f^\delta\|^2 + \lambda_k a_k J(x_k) + J(u_k) \leq \lambda_k \delta^2 + \lambda_k a_k J(z) + J(z - x_{k-1}). \tag{4.15}$$

Note that $\|Tx_k - f^\delta\|^2 + a_k J(x_k)$ is decreasing due to (4.14) and to $a_k \leq a_{k-1}$.

Consider the stopping index by a discrepancy rule of the form

$$k^*(\delta) := \max\{k \in \mathbb{N} : \|Tx_k - f^\delta\|^2 + a_k J(x_k) \geq \tau \delta^2\}, \quad \text{for some } \tau > 1. \tag{4.16}$$

One can formulate and prove a convergence result as for the MHDM.

Proposition 4.3. *Let conditions* (3.2), (4.1) *and* (4.3) *be satisfied. Then the following estimate holds*

$$\|Tx_k - f^\delta\|^2 + a_k J(x_k) \leq \delta^2 + \left(\sum_{j=0}^{k} a_j\right) \frac{J(z)}{k+1} + \frac{2J(z)}{(k+1)\lambda_0}, \quad \forall k \geq 0. \tag{4.17}$$

Moreover, if (4.7) *is verified, then the stopping index defined by* (4.16) *is finite. Additionally,*

1. *If* $(k^*(\delta))$ *is unbounded, then* $\lim_{\delta \to 0} Tx_{k^*(\delta)} = f$ *and* $\lim_{\delta \to 0} a_{k^*(\delta)} J x_{k^*(\delta)} = 0$.

2. *If the stopping index is chosen as* $k^*(\delta) \sim \dfrac{1}{\delta^2}$, *then*

$$\|Tx_{k^*(\delta)} - f\|^2 + a_{k^*(\delta)} J(x_{k^*(\delta)}) = O(\delta^2).$$

Proof: One uses $J(z - x_k) - J(z - x_{k-1}) \leq J(u_k)$, $\forall k \geq 0$, in (4.6) to obtain

$$\lambda_k \|Tx_k - f^\delta\|^2 + J(z - x_k) + \lambda_k a_k J(x_k) \leq \lambda_k \delta^2 + \lambda_k a_k J(z) + 2J(z - x_{k-1}),$$

which yields

$$\|Tx_k - f^\delta\|^2 + a_k J(x_k) + \frac{1}{\lambda_k} J(z - x_k) \leq \delta^2 + a_k J(z) + \frac{2}{\lambda_k} J(z - x_{k-1}) \leq \delta^2 + a_k J(z) + \frac{1}{\lambda_{k-1}} J(z - x_{k-1}), \tag{4.18}$$

as $\dfrac{2}{\lambda_k} \leq \dfrac{1}{\lambda_{k-1}}$ if $k \geq 1$. By writing (4.18) for indices $0, 1, ..., k$ and summing up, one has for any $k \geq 0$

$$\begin{aligned}
(k+1)\|Tx_k - f^\delta\|^2 + (k+1)a_k J(x_k) + \frac{1}{\lambda_k} J(z - x_k) &\leq \sum_{j=0}^{k} \|Tx_j - f^\delta\|^2 + \sum_{j=0}^{k} a_j J(x_j) \\
&+ \frac{1}{\lambda_k} J(z - x_k) \\
&\leq (k+1)\delta^2 + \left(\sum_{j=0}^{k} a_j\right) J(z) + \frac{2}{\lambda_0} J(z),
\end{aligned}$$

where the first inequality is based on the monotonicity of the term $\|Tx_k - f^\delta\|^2 + a_k J(x_k)$. This yields (4.17) when dividing by $k + 1$.

In the case of the discrepancy principle rule, by replacing $k = k^*(\delta)$ in the last inequality and using (4.16) it follows that

$$k^*(\delta) \leq \frac{1}{(\tau-1)\delta^2}\left[\left(\sum_{j=0}^{k} a_j\right) J(z) + \frac{2J(z)}{\lambda_0}\right] - 1. \tag{4.19}$$

Thus, the stopping index $k^*(\delta)$ is finite for any arbitrary but fixed δ.

If $(k^*(\delta))$ is unbounded, then (4.17) written for $k = k^*(\delta)$ implies $\lim_{\delta\to 0} Tx_{k^*(\delta)} = f$, that is 1. holds.

Statement 2. follows from (4.17) written for $k^*(\delta)$ chosen a priori as above. □

4.2 Convergence of the iterates in the case of noisy data

In the sequel, we focus on convergence of the iterates (x_k) when $T \neq I$. To this end, we would like to show that

$$\limsup_{\delta\to 0} J(x_{k(\delta)}) \leq J(z) \tag{4.20}$$

for some stopping index $k(\delta)$.

From (4.15) one has

$$J(x_k) \leq \frac{\delta^2}{a_k} + J(z) + \frac{J(z - x_{k-1})}{\lambda_k a_k}, \quad \forall k \geq 0. \tag{4.21}$$

First, we propose upper bounds for the term $J(z - x_{k-1})$ as follows.

Proposition 4.4. *Let conditions* (3.2)*,* (4.1) *and* (4.3) *hold. Then*

$$J(z - x_k) \leq 2^{k+1}\left(\frac{\delta^2}{a_{k+1}} + J(z)\right), \quad \forall k \geq 0. \tag{4.22}$$

Proof: By letting $k = 0$ in (4.15) and taking into account that $x_0 = u_0$, $x_{-1} = 0$, one has

$$J(x_0) \leq \frac{\delta^2\lambda_0}{1 + \lambda_0 a_0} + J(z) \leq \frac{\delta^2}{a_0} + J(z).$$

which yields (4.22) in this particular instance since $J(z - x_0) \leq J(z) + J(x_0)$ and $a_0 \geq a_1$.

Now fix $k \geq 1$. If

$$J(x_k) \leq \frac{\delta^2}{a_k} + J(z), \tag{4.23}$$

then (4.22) is verified since

$$J(z - x_k) \leq J(z) + J(x_k) \leq \frac{\delta^2}{a_k} + 2J(z) \leq \frac{\delta^2}{a_{k+1}} + 2J(z) \leq 2^{k+1}\left(\frac{\delta^2}{a_{k+1}} + J(z)\right).$$

Otherwise, let k_0 with $k > k_0 \geq 0$ be the largest natural number for which

$$J(x_{k_0}) \leq \frac{\delta^2}{a_{k_0}} + J(z) \tag{4.24}$$

holds. By using

$$J(x_k) > \frac{\delta^2}{a_k} + J(z)$$

in (4.15), one derives $J(u_k) \leq J(z - x_{k-1})$. This and $J(z - x_k) \leq J(u_k) + J(z - x_{k-1})$ yield

$$J(z - x_k) \leq 2J(z - x_{k-1}), \quad \forall k \geq k_0 + 1.$$

Thus,

$$\begin{aligned} J(z-x_k) &\leq 2J(z-x_{k-1}) \leq \cdots \leq 2^{k-k_0} J(z-x_{k_0}) \\ &\leq 2^{k-k_0}\left(J(x_{k_0})+J(z)\right) \\ &\leq 2^{k-k_0}\left(\frac{\delta^2}{a_{k_0}}+2J(z)\right) \\ &\leq 2^{k+1-k_0}\left(\frac{\delta^2}{a_{k_0}}+J(z)\right) \\ &\leq 2^{k+1}\left(\frac{\delta^2}{a_{k+1}}+J(z)\right), \end{aligned}$$

where $a_{k_0} \geq a_{k+1} \geq 0$, $k-k_0+1 \leq k+1$ and (4.24) were employed to obtain the last two inequalities. □

We show now (4.20).

Corollary 4.5. *Let conditions* (3.2), (4.1) *and* (4.3) *hold. Then*

$$J(x_k) \leq \left(\frac{\delta^2}{a_k}+J(z)\right)\left(\frac{2^k}{\lambda_k a_k}+1\right), \quad \forall k \geq 0. \tag{4.25}$$

Proof: We employ (4.22) (written for $k-1$ instead of k) in inequality (4.21),

$$J(x_k) \leq \frac{\delta^2}{a_k}+J(z)+\frac{J(z-x_{k-1})}{\lambda_k a_k} \leq \frac{\delta^2}{a_k}+J(z)+\frac{2^k}{\lambda_k a_k}\left(\frac{\delta^2}{a_k}+J(z)\right),$$

which reduces to (4.25). □

Remark 4.6. *Letting $\delta = 0$ in* (4.25) *recovers precisely formula* (4.12) *of the exact data case.*

In the sequel, we formulate the result on convergence of the iterates.

Corollary 4.7. *Let conditions* (3.2), (4.1), (4.3), (4.10) *and* (4.7) *hold, and assume that* $\dfrac{\delta^2}{a_{k^*(\delta)}} \to 0$ *as* $\delta \to 0$, *when* $(k^*(\delta))$ *is chosen according to one of the following rules*

1. *a priori as* $k^*(\delta) \sim \dfrac{1}{\delta^2}$.
2. *via the discrepancy principle* (4.16). *Additionally, let* $(k^*(\delta))$ *be unbounded.*

Then (4.20) *holds and* $(x_{(k^*(\delta))})$ *converges strictly on a subsequence to a J minimizing solution of $Tu = f$.*

Proof: We obtain (4.20) by writing (4.25) for $k = k^*(\delta)$ and using the assumptions on the involved sequences. As in Proposition 4.2, one obtains strict convergence of a subsequence of $(x_{k^*(\delta)})$ to a solution of $Tu = f$ which minimizes J. □

Remark 4.8. *A stopping index of the form* $k^*(\delta) \sim \dfrac{1}{\delta^2}$ *yields also* $\lim\limits_{\delta \to 0} Tx_{k^*(\delta)} = f$, *as one can see from* (4.17).

Remark 4.9. *The sequences* (λ_k) *and* (a_k) *should be chosen to satisfy assumptions* (4.1), (4.3), (4.10) *and* (4.7). *Here is an example in this sense:*

$$a_k = \frac{1}{(k+1)^{3/2}}, \quad \lambda_k = 3^k,\ k \geq 0.$$

5 A refinement of the tight MHDM

Recall that z is a solution of $Tu = f$.

One can refine (4.2) by defining a sequence $(u_k) \subset X$ with u_k as a solution of

$$\min_{u\in X}\{\lambda_k\|T(u + x_{k-1}) - f\|^2 + \lambda_k a_k J(u + x_{k-1}) + R_k(u)\}, \quad k \geq 0, \tag{5.1}$$

where $(\lambda_k) \subset (0, \infty)$ satisfies (4.3) and (4.10). That is, the regularizer $J : H \to [0, \infty)$ remains the same in all iterations, while the regularizer on the component u_k will be updated at each iteration. Here J is allowed to be any function which is bounded from below, while each $R_k : H \to \mathbb{R} \cup \{\infty\}$ is a seminorm and is weakly lower semicontinuous. Welldefinedness of minimizers u_k is guaranteed due to the coercivity of the terms containing T and J, as well as due to the lower semicontinuity of the functional that is minimized.

We propose the following class of functions R_k and J that make method (5.1) work:

$$R_k(u) \leq R_{k-1}(u), \quad \forall u \in X, \forall k = 1, 2, ... \tag{5.2}$$

$$R_k(u) \leq cJ(u), \forall u \in X, \forall k \in \mathbb{N}, \tag{5.3}$$

for some positive constant $c > 0$.

Clearly, the case when R is different from the penalty J is the interesting one in this section. Due to condition (5.3), the penalty R_k is a weaker one, thus we allow the new component u_k to be less smooth than x_{k-1}. For instance, choosing for R_k a seminorm that encourages oscillations (such as $\|\cdot\|_*$ which coincides with the seminorm in the Sobolev space $G = \dot{W}^{-1,\infty}(\Omega)$ of negative exponent), we encourage more details or texture components to be recovered in the final restored image.

The following estimate can be derived as in Section 4, implying convergence of (Tx_k) to f when (4.7) is verified.

Proposition 5.1. *Let conditions* (5.2) *and* (5.3) *be satisfied. Then the following estimate holds*

$$\|Tx_k - f\|^2 + a_k J(x_k) \leq \left(\sum_{j=0}^{k} a_j\right) \frac{J(z)}{k+1} + \frac{2R_0(z)}{(k+1)\lambda_0}, \quad \forall k \geq 0. \tag{5.4}$$

More convergence results can be obtained under the additional assumptions on the involved parameters.

Proposition 5.2. *If conditions* (4.10), (4.7), (5.2) *and* (5.3) *hold, then*

$$\limsup_{k\to\infty} J(x_k) \leq J(z).$$

Moreover, (x_k) *converges strictly on subsequences to a solution of* $Tu = f$ *which minimizes* J .

Proof:

By considering (5.1) for $u = z - x_{k-1}$, one has

$$\lambda_k\|Tx_k - f\|^2 + \lambda_k a_k J(x_k) + R_k(u_k) \leq \lambda_k a_k J(z) + R_k(z - x_{k-1}) \quad k = 1, 2, ... \tag{5.5}$$

and thus, according to (5.2),

$$J(x_k) \leq J(z) + \frac{R_k(z - x_{k-1})}{\lambda_k a_k} \leq J(z) + \frac{R_{k-1}(z - x_{k-1})}{\lambda_k a_k}. \tag{5.6}$$

We claim that

$$R_k(z - x_k) \leq 2^{k+1} cJ(z). \tag{5.7}$$

By letting $k = 0$ in (5.5), one obtains $\lambda_0 a_0 J(x_0) + R_0(x_0) \leq \lambda_0 a_0 J(z) + R_0(z)$. Since $u_0 = x_0$ and (5.3) holds, it follows that $(1/c + \lambda_0 a_0)R_0(x_0) \leq (\lambda_0 a_0 + c)J(z)$, that is $R_0(x_0) \leq cJ(z)$, which implies $R_0(z - x_0) \leq R_0(z) + R_0(x_0) \leq 2cJ(z)$.

Now fix $k \geq 1$. If

$$R_k(x_k) \leq cJ(z), \tag{5.8}$$

then this combined with the subadditivity of R_k and (5.3) yields (5.7). Otherwise, let k_0 with $k > k_0 \geq 0$ be the largest positive integer such that

$$R_{k_0}(x_{k_0}) \leq cJ(z).$$

Employing

$$R_k(x_k) > cJ(z)$$

and $R_k(x_k) \leq cJ(x_k)$ in (5.5) yield $R_k(u_k) \leq R_k(z - x_{k-1})$. Combining this with $R_k(z - x_k) \leq R_k(u_k) + R_k(z - x_{k-1})$ and (5.2), one obtains

$$R_k(z - x_k) \leq 2R_k(z - x_{k-1}) \leq 2R_{k-1}(z - x_{k-1}) \cdots \leq 2^{k-k_0} R_{k_0}(z - x_{k_0}) \leq 2^{k-k_0+1} J(z) \leq 2^{k+1} cJ(z).$$

Thus, the claim is proved. Using this in (5.6) and taking into account (4.10) yield $\limsup_{k\to\infty} J(x_k) \leq J(z)$. □

One can define also stopping rules as in the previous sections and obtain similar results. Moreover, Subsection 6.2 contains the discretization of the refined version of tight MHDM while Subsection 7.4 presents numerical results with this refinement version when J is the total variation and $R_k = \|\cdot\|_*$.

6 Discretization of Euler-Lagrange equations of the tighter version and its refinement

6.1 The tight MHDM

The discretization of the Euler-Lagrange equation at each step of MHDM has been given in [38] when $J(u) = |u|_{BV(\Omega)} \approx \int_\Omega \sqrt{\epsilon^2 + |\nabla u|^2} dx$. Thus, J is regularized using a small parameter $\epsilon > 0$, becoming differentiable at points where $|\nabla u| = 0$.

In this section, the discretization of tighter MHDM is elaborated using the same choice for J, inspired from image restoration, and a similar finite difference scheme inspired from [41]. For one step, the Euler-Lagrange equation associated with the minimization from equation (4.13) is

$$T^*(T(u + x_{k-1}) - f^\delta) = \frac{a_k}{2} \mathrm{div}\left(\frac{\nabla(u + x_{k-1})}{\sqrt{\epsilon^2 + |\nabla(u + x_{k-1})|^2}}\right) + \frac{1}{2\lambda_k} \mathrm{div}\left(\frac{\nabla u}{\sqrt{\epsilon^2 + |\nabla u|^2}}\right) \text{ in } \Omega, \tag{6.1}$$

$$\nabla u \cdot \vec{n} = 0 \text{ on } \partial\Omega, \tag{6.2}$$

where $\vec{n}$ is the exterior unit normal to the boundary $\partial\Omega$, $k \geq 0$ with $x_{-1} = 0$. Using the gradient descent scheme to solve the minimization problem, we are solving the following dynamic PDE in $u = u_k$:

$$\begin{aligned} \frac{\partial u}{\partial t} + T^*(T(u + x_{k-1}) - f^\delta) = \frac{a_k}{2} \mathrm{div}\left(\frac{\nabla(u + x_{k-1})}{\sqrt{\epsilon^2 + |\nabla(u + x_{k-1})|^2}}\right) \\ + \frac{1}{2\lambda_k} \mathrm{div}\left(\frac{\nabla u}{\sqrt{\epsilon^2 + |\nabla u|^2}}\right) \text{ in } \Omega \times [0, \infty), \end{aligned} \tag{6.3}$$

$$u = f^\delta - Tx_{k-1} \text{ on } \partial\Omega \text{ and } t = 0. \tag{6.4}$$

After discretizing the space domain with grid size h, and the time interval with time step Δt, u_{ij}^n represents the value of u at (ih, jh) and time $n\Delta t$. The finite differences method is applied to solving the above problem. To simplify the expressions, the following notations are used:

$$\nabla_+^x u_{ij} = \frac{u_{i+1,j} - u_{ij}}{h}, \qquad \nabla_-^x u_{ij} = \frac{u_{ij} - u_{i-1,j}}{h},$$

$$\nabla_+^y u_{ij} = \frac{u_{i,j+1} - u_{ij}}{h}, \qquad \nabla_-^y u_{ij} = \frac{u_{ij} - u_{i,j-1}}{h},$$

$$c(u_{ij}) = \frac{1}{\sqrt{\epsilon^2 + (\nabla_+^x u_{ij})^2 + (\nabla_+^y u_{ij})^2}}.$$

Similarly, u can be replaced with $w = u + x_{k-1}$ in the above equations to obtain $\nabla_+^x w_{ij}$, $\nabla_-^x w_{ij}$, $\nabla_+^y w_{ij}$, $\nabla_-^y w_{ij}$ and $c(w_{ij})$. Then equation (6.3) can be solved using the following finite difference method as

$$\begin{aligned}\frac{\partial u_{ij}}{\partial t} &= -T^*(Tw_{ij} - f_{ij}^\delta) + \frac{a_k}{2}[\nabla_-^x(c(w_{ij})\nabla_+^x w_{ij}) + \nabla_-^y(c(w_{ij})\nabla_+^y w_{ij})] \\ &\quad + \frac{1}{2\lambda_k}[\nabla_-^x(c(u_{ij})\nabla_+^x u_{ij}) + \nabla_-^y(c(u_{ij})\nabla_+^y u_{ij})].\end{aligned} \tag{6.5}$$

By applying semi-implicit time marching scheme, we have

$$\begin{aligned}\frac{u_{ij}^{n+1} - u_{ij}^n}{\Delta t} &= -T^*(Tw_{ij}^n - f_{ij}^\delta) \\ &\quad + \frac{a_k}{2h^2}[c(w_{ij}^n)(w_{i+1,j}^n - w_{ij}^{n+1}) - c(w_{i-1,j}^n)(w_{ij}^{n+1} - w_{i-1,j}^n) \\ &\quad + c(w_{ij}^n)(w_{i,j+1}^n - w_{ij}^{n+1}) - c(w_{i,j-1}^n)(w_{ij}^{n+1} - w_{i,j-1}^n)] \\ &\quad + \frac{1}{2\lambda_k h^2}[c(u_{ij}^n)(u_{i+1,j}^n - u_{ij}^{n+1}) - c(u_{i-1,j}^n)(u_{ij}^{n+1} - u_{i-1,j}^n) \\ &\quad + c(u_{ij}^n)(u_{i,j+1}^n - u_{ij}^{n+1}) - c(u_{i,j-1}^n)(u_{ij}^{n+1} - u_{i,j-1}^n)],\end{aligned} \tag{6.6}$$

where $w^n = u^n + x_{k-1}$. Therefore, the discretization becomes

$$\begin{aligned}\frac{u_{ij}^{n+1} - u_{ij}^n}{\Delta t} &= -T^*(Tw_{ij}^n - f_{ij}^\delta) \\ &\quad + \frac{a_k}{2h^2}[c(w_{ij}^n)w_{i+1,j}^n + c(w_{i-1,j}^n)w_{i-1,j}^n + c(w_{ij}^n)(w_{i,j+1}^n + c(w_{i,j-1}^n)w_{i,j-1}^n \\ &\quad - (c(w_{ij}^n) + c(w_{i-1,j}^n) + c(w_{ij}^n) + c(w_{i,j-1}^n))x_{k-1,ij} \\ &\quad - (c(w_{ij}^n) + c(w_{i-1,j}^n) + c(w_{ij}^n) + c(w_{i,j-1}^n))u_{ij}^{n+1}] \\ &\quad + \frac{1}{2\lambda_k h^2}[c(u_{ij}^n)u_{i+1,j}^n + c(u_{i-1,j}^n)u_{i-1,j}^n + c(u_{ij}^n)(u_{i,j+1}^n + c(u_{i,j-1}^n)u_{i,j-1}^n \\ &\quad - (c(u_{ij}^n) + c(u_{i-1,j}^n) + c(u_{ij}^n) + c(u_{i,j-1}^n))u_{ij}^{n+1}].\end{aligned} \tag{6.7}$$

Then u_{ij}^{n+1} can be obtained by:

$$u_{ij}^{n+1} = \frac{u_{ij}^n - \Delta t T^*(Tw_{ij}^n - f_{ij}^\delta) + \frac{\Delta t a_k}{2h^2}[\Xi_w^n - c_w^n x_{k-1,ij}] + \frac{\Delta t}{2\lambda_k h^2}\Xi_u^n}{1 + \frac{\Delta t a_k}{2h^2}c_w^n + \frac{\Delta t}{2\lambda_k h^2}c_u^n} \tag{6.8}$$

where

$$\begin{aligned}\Xi_w^n &= c(w_{ij}^n)w_{i+1,j}^n + c(w_{i-1,j}^n)w_{i-1,j}^n + c(w_{ij}^n)w_{i,j+1}^n + c(w_{i,j-1}^n)w_{i,j-1}^n, \\ \Xi_u^n &= c(u_{ij}^n)u_{i+1,j}^n + c(u_{i-1,j}^n)u_{i-1,j}^n + c(u_{ij}^n)u_{i,j+1}^n + c(u_{i,j-1}^n)u_{i,j-1}^n, \\ c_w^n &= c(w_{ij}^n) + c(w_{i-1,j}^n) + c(w_{ij}^n) + c(w_{i,j-1}^n), \\ c_u^n &= c(u_{ij}^n) + c(u_{i-1,j}^n) + c(u_{ij}^n) + c(u_{i,j-1}^n).\end{aligned}$$

6.2 A refinement of the tight MHDM

In this case, J remains the total variation and we let $R_k(u) = ||u||_*$, i.e. its dual weaker norm, hoping that more details and texture are recovered. Then the refinement of the tight MHDM becomes:

$$\min_{u\in X}\{\lambda_k\|T(u + x_{k-1}) - f\|^2 + \lambda_k a_k J(u + x_{k-1}) + ||u||_*\}, \quad k \geq 0, \tag{6.9}$$

where $||u||_* = \sup_{J(\varphi)\neq 0} \frac{\langle u, \varphi\rangle}{J(\varphi)}$, and $J(\varphi) = |\varphi|_{BV(\Omega)} \approx \int_\Omega \sqrt{\epsilon^2 + |\nabla\varphi|^2}dx$. We employ the numerical scheme from [13] for dealing with the $\|\cdot\|_*$ norm.

Denote $F(\varphi) = \dfrac{\langle u, \varphi\rangle}{J(\varphi)}$. Suppose u is given, then the associated Euler-Lagrange equation associated with $\sup\limits_{J(\varphi)\neq 0} F(\varphi)$ is

$$\frac{u}{\int_\Omega \sqrt{\epsilon^2 + |\nabla\varphi|^2}dx} + \frac{\int_\Omega u\varphi dx}{(\int_\Omega \sqrt{\epsilon^2 + |\nabla\varphi|^2}dx)^2}\mathrm{div}(\frac{\nabla\varphi}{\sqrt{\epsilon^2 + |\nabla\varphi|^2}}) = 0 \qquad \text{in } \Omega, \tag{6.10}$$

$$\nabla\varphi\cdot\vec{n} = 0 \ \text{ on } \partial\Omega. \tag{6.11}$$

Equation (6.10) can also be written as

$$u + \frac{\int_\Omega u\varphi dx}{\int_\Omega \sqrt{\epsilon^2 + |\nabla\varphi|^2}dx}\mathrm{div}(\frac{\nabla\varphi}{\sqrt{\epsilon^2 + |\nabla\varphi|^2}}) = 0 \qquad \text{in } \Omega. \tag{6.12}$$

Using the gradient ascent method to obtain $\{\varphi : \sup\limits_{J(\varphi)\neq 0} F(\varphi)\}$, we are solving

$$\frac{\partial\varphi}{\partial t} = F'(\varphi(t)) \quad t\in[0,\infty), \tag{6.13}$$

which means we are solving the following PDE:

$$\frac{\partial\varphi}{\partial t} = u + \frac{\int_\Omega u\varphi dx}{\int_\Omega \sqrt{\epsilon^2 + |\nabla\varphi|^2}dx}\mathrm{div}(\frac{\nabla\varphi}{\sqrt{\epsilon^2 + |\nabla\varphi|^2}}) \qquad \text{in } \Omega\times[0,\infty). \tag{6.14}$$

Once φ is obtained, the associated Euler-Lagrange equation for minimizing the refinement of the tight MHDM as shown in (6.9) is

$$T^*(T(u+x_{k-1}) - f^\delta) = \frac{a_k}{2}\mathrm{div}(\frac{\nabla(u+x_{k-1})}{\sqrt{\epsilon^2 + |\nabla(u+x_{k-1})|^2}}) - \frac{1}{2\lambda_k}\frac{\varphi}{\int_\Omega \sqrt{\epsilon^2 + |\nabla\varphi|^2}dx} \ \text{ in } \Omega, \tag{6.15}$$

$$\nabla u\cdot\vec{n} = 0 \ \text{ on } \partial\Omega. \tag{6.16}$$

Applying the gradient descent scheme to solving this minimization problem, we have:

$$\begin{aligned}\frac{\partial u}{\partial t} + T^*(T(u+x_{k-1}) - f^\delta) = {} & \frac{a_k}{2}\mathrm{div}(\frac{\nabla(u+x_{k-1})}{\sqrt{\epsilon^2 + |\nabla(u+x_{k-1})|^2}}) \\ & - \frac{1}{2\lambda_k}\frac{\varphi}{\int_\Omega \sqrt{\epsilon^2 + |\nabla\varphi|^2}dx} \qquad \text{in } \Omega\times[0,\infty).\end{aligned} \tag{6.17}$$

We use finite differences and semi-implicit time marching to solve equations (6.14) and (6.17) iteratively as follows:

$$\varphi_{ij}^{n+1} = \frac{\varphi_{ij}^n + \Delta t u_{ij}^n + \frac{\Delta t}{h^2}\frac{\int_\Omega u_{ij}^n\varphi_{ij}^n dx}{\int_\Omega \sqrt{\epsilon^2+|\nabla\varphi_{ij}^n|^2}dx}\Xi_\varphi^n}{1 + \frac{\Delta t}{h^2}\frac{\int_\Omega u_{ij}^n\varphi_{ij}^n dx}{\int_\Omega \sqrt{\epsilon^2+|\nabla\varphi_{ij}^n|^2}dx}c_\varphi^n}, \tag{6.18}$$

$$u_{ij}^{n+1} = \frac{u_{ij}^n - \Delta t T^*(Tw_{ij}^n - f_{ij}^\delta) + \frac{\Delta t a_k}{2h^2}[\Xi_w^n - c_w^n x_{k-1,ij}] - \frac{\Delta t}{2\lambda_k}\frac{\varphi_{ij}^{n+1}}{\int_\Omega \sqrt{\epsilon^2+|\nabla\varphi_{ij}^{n+1}|^2}dx}}{1 + \frac{\Delta t a_k}{2h^2}c_w^n} \tag{6.19}$$

where

$$\Xi_\varphi^n = c(\varphi_{ij}^n)\varphi_{i+1,j}^n + c(\varphi_{i-1,j}^n)\varphi_{i-1,j}^n + c(\varphi_{ij}^n)\varphi_{i,j+1}^n + c(\varphi_{i,j-1}^n)\varphi_{i,j-1}^n,$$
$$c_\varphi^n = c(\varphi_{ij}^n) + c(\varphi_{i-1,j}^n) + c(\varphi_{ij}^n) + c(\varphi_{i,j-1}^n).$$

In addition to the procedures above, other algorithms can also be applied to solving the tight MHDM and its refinement, such as the generic proximal algorithm [14], the Douglas-Rachford algorithm and the alternating direction method of multipliers (ADMM) [4, 12, 19, 22]. A clear advantage of the method we propose here is its stability - see Subsection 7.1 and the comments at the end of Section 7.

7 Numerical results

We present in this section various experimental results for image denoising and deblurring using the proposed methods. These help further to: validate the theoretical findings, compare the convergence rates of MHDM and tighter MHDM, and evaluate the stopping criteria. The experiments employ the stopping index $k^*(\delta)$ determined by the discrepancy principle (3.5) for MHDM and (4.16) for tighter MHDM.

7.1 MHDM and tighter MHDM for deblurring

First, MHDM and the tighter version of MHDM are applied to image deblurring. As shown in Fig. 2, f is the blurred version of the original image z obtained by applying the 5×5 Gaussian blur window with variance 2. The root mean square error ($RMSE$) defined as follows is used to evaluate the deblurring results:

$$RMSE = \frac{||x_k - z||}{N}, \tag{7.1}$$

where $||\cdot||$ is the Euclidean norm and N is the total number of pixels in the image.

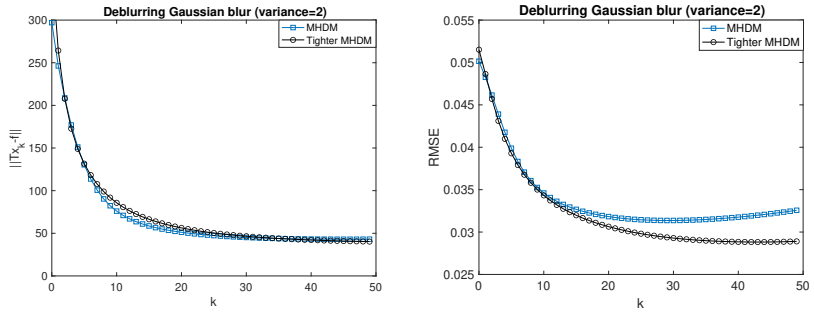

Figure 1: $||Tx_k - f||$ and RMSE using MHDM and tighter MHDM for deblurring.

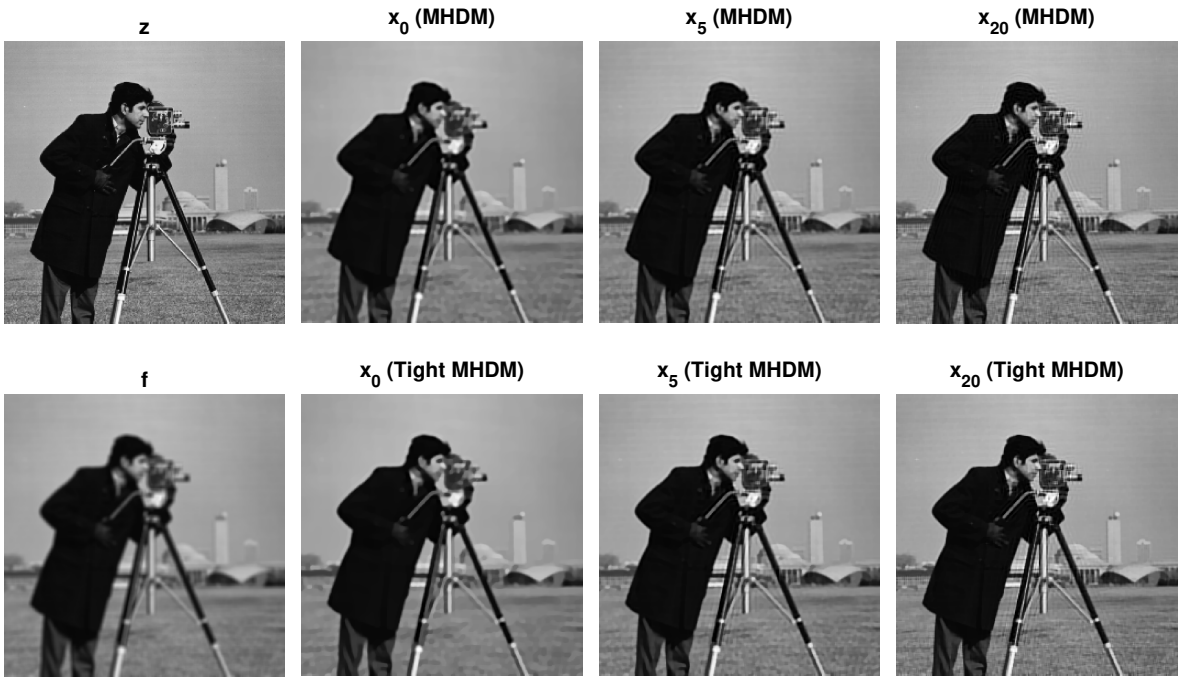

Figure 2: Deblurring results of the image with Gaussian blur (variance=2).

The value of ϵ in total variation is used to avoid the singularity when $|\nabla u| = 0$, so ϵ^2 is chosen as a small number close to 0, namely $\epsilon^2 = 10^{-6}$ here. This numerical experiment shows that the greater the

ϵ, the lower the errors, but the difference is not obvious. As the value of k increases, the difference is getting smaller, so that it can be ignored. The time step size used in this paper is $\Delta t = 0.025$. However, the results are not very sensitive to the choice of Δt. For example, if we change Δt from 0.025 to 0.01, the $RMSE$ decreases around 5×10^{-5} for each k. On the other hand, if Δt increases up to $\Delta t = 2$, the algorithms are still stable and the results still converge. The parameter λ_k can adjust the weights of the fidelity term and the regularizing term. The smaller the λ_k, the more cartoon the u_k preserves. The larger λ_k can keep more texture details of the image in u_k. To restore blurred images without noise, $\lambda_0 = 1$ is used for both algorithms, while for denoising images in subsection 7.2, a smaller λ_0 is used so that the regularizing weight can be increased. Recall that $\lambda_k = 2^k\lambda_0$ for MHDM; $\lambda_k = 3^k\lambda_0$, $a_k = \frac{1}{(k+1)^{3/2}}$, $k \geq 0$ for tighter MHDM, so that equations (4.1), (4.3), (4.10), and (4.7) can be satisfied. The grid size is $h = 1$.

We let $N = 256^2$ for the image used in Fig. 2. As shown in Fig. 1, $||Tx_k - f||$ for deblurring decreases as k increases, which means $\lim_{k\to\infty} Tx_k = f$. However, from the $RMSE$ curves we can see that, as k exceeds 30, the $RMSE$ for MHDM begins to slightly increase. That means (x_k) does not converge to z for MHDM. The deblurred images using both methods as k increases can also be seen in Fig. 2. Overall, the deblurring effects are similar for both methods. However, when k becomes relatively large, compared to its tighter version, MHDM produces more contour-like error fluctuations along the edges of the person and tripod.

7.2 MHDM and tighter MHDM for denoising

Both methods are applied to denoising the noisy images f^δ with zero-mean additive Gaussian noise. In this case, $T = I$, so the original image is $z = f$. Gaussian noise of different variances is added to f.

Recall that the stopping index $k^*(\delta)$ is determined by (3.5) for MHDM and (4.16) for tighter MHDM. Note that for some specific f^δ, the value of k^* decreases as τ increases, which means that larger τ can result in an earlier stopping. In all the numerical experiments, the value of τ is defined by

$$\tau := \min\{\tilde{\tau}, \quad \text{s.t.} \quad \tilde{\tau} > 1 \quad \textit{and} \quad \exists k \in \mathbb{N} : \|Tx_k - f^\delta\|^2 \geq \tilde{\tau}\delta^2\}, \quad \text{for MHDM,} \tag{7.2}$$

and

$$\tau := \min\{\tilde{\tau}, \quad \text{s.t.} \quad \tilde{\tau} > 1 \quad \textit{and} \quad \exists k \in \mathbb{N} : \|Tx_k - f^\delta\|^2 + a_k J(x_k) \geq \tilde{\tau}\delta^2\}, \quad \text{for tighter MHDM.} \tag{7.3}$$

Therefore, by using the definition above, the value of τ depends on the noise level. For example, if the noise variance is 10^{-2}, then $\tau = 1.0157$ for MHDM, and $\tau = 1.0303$ for tighter MHDM; if the noise variance is 10^{-3}, then $\tau = 1.5166$ for MHDM, and $\tau = 1.8064$ for tighter MHDM.

As discussed after (2.7), if a small λ_0 is used, a good restoration cannot be obtained until a large enough value of k is reached. However, if λ_0 is relatively big, it is possible that there doesn't exist a k that satisfies (3.5) or (4.16). On the other hand, as mentioned earlier, bigger λ_0 leads to a more textured restoration image. Therefore, if a large λ_0 is used for denoising, there will be a greater possibility of retaining more noise in initial stages, which is contrary to our intention of denoising. Thus, compared to deblurring, we use a smaller $\lambda_0 = 0.01$ for denoising using MHDM and tighter MHDM. The values of all the other parameters are the same as those in the previous (deblurring) case.

7.2.1 MHDM for denoising

We first denoise the image f^δ which is obtained by adding Gaussian noise to the original image z from Fig. 2. The results are shown in Fig. 4. The error $||x_k - f||$ of MHDM versus k with different noise levels is shown in Fig. 3. To determine $k^*(\delta)$ we use equation (3.5), and δ satisfying (3.2) is calculated as $\delta = ||f - f^\delta||$ in all the numerical experiments in this section. Noise variance values 10^{-2}, 10^{-3}, 10^{-4}, and 10^{-5} correspond to $\delta = 6238$, 2045, 656, 220 respectively.

From Fig. 3 we can see that, as the value of k increases, the errors first decrease and then increase, which is especially obvious when the noise variance is relatively big. This illustrates the importance of setting up the stopping rule ($k^*(\delta)$ is marked in red asterisk). Fig. 3 shows that the value of k^* increases as δ decreases. This shows that an early stopping criterion is necessary to prevent error propagation.

The semiconvergence behavior mentioned in the Introduction is clearly seen in some figures, thus justifying the earlier stopping. Note that the stopping index is pretty close to the index $k_{min} :=$

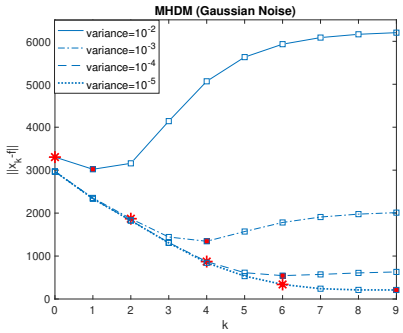

Figure 3: Errors of denoising results of the Cameraman image with different levels of Gaussian noise using MHDM. k^* is marked in red asterisk, k_{min} is marked in red squares.

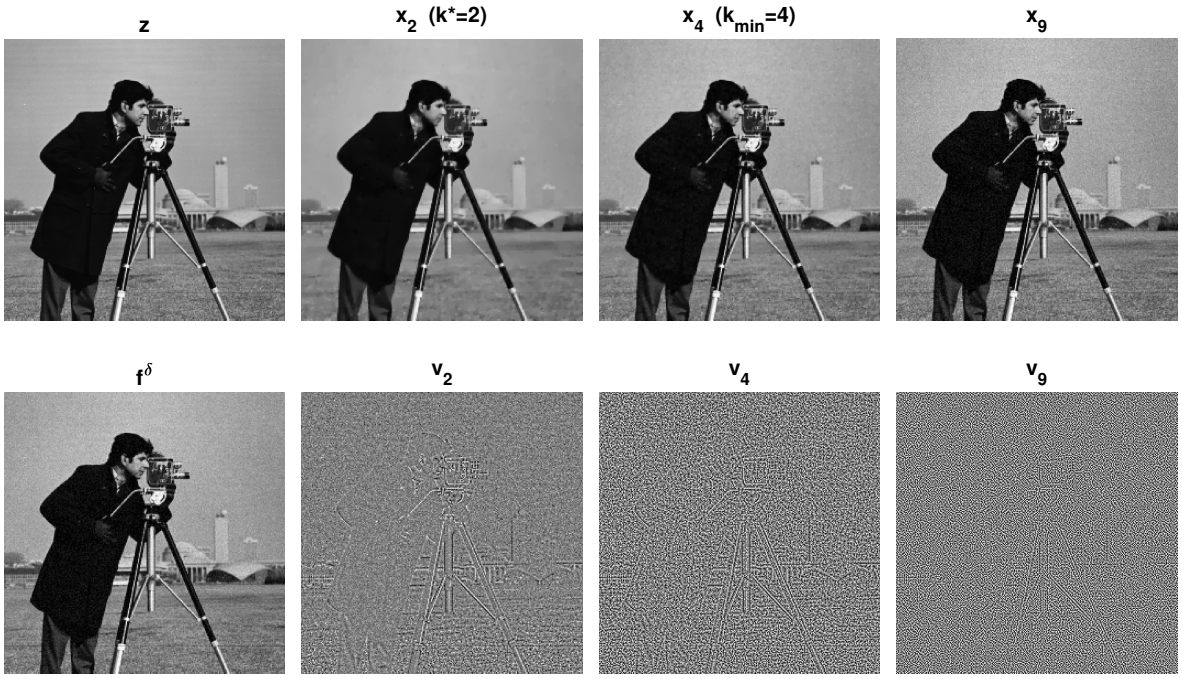

Figure 4: MHDM denoising results of the Cameraman image with Gaussian noise (variance = 10^{-3}), $\delta = 2045$.

$\arg\min\{\|x_k - f\|, k \in \mathbb{N}\}$ for MHDM and tight MHDM for the denoising problem - see subsection 7.2.2. Although we did not prove the semiconvergence with respect to this distance, we observe graphically that the analyzed procedures produce for a while better and better approximations of the solution, up to some index, when the iterates start to be less and less meaningful.

In this example, for certain δ, $k^* < k_{min}$. Even though k^* is not always equal to k_{min}, the difference between $||x_{k^*} - f||$ and $||x_{k_{min}} - f||$ is acceptable - the details of the difference between x_{k^*} and $x_{k_{min}}$, as well as v_{k^*} and $v_{k_{min}}$ with $\delta = 2045$, can be seen in Fig. 4. In order to show fine features more clearly, all residual images in this paper are scaled to the full range of the intensity. We can also see that if the MHDM does not stop earlier at $k^* = 2$, then e.g. the denoised image x_9 has more noise than x_{k^*}. The decomposition components u_k, $k = 0, 1, 2, 3, 4$ are also shown in Fig. 5. Since $x_k = \sum_{j=0}^{k} u_j$, in this case, u_1 and u_2 clearly reveal the difference between x_0 and x_{k^*}. Similarly, u_3 and u_4 reveal the difference between x_{k^*} and $x_{k_{min}}$. Therefore, by comparing x_{k^*} with x_0 and $x_{k_{min}}$, it can be seen that the MHDM can sharpen edges while denoising, and the early stopping rule can keep most edge details while removing most of the noise.

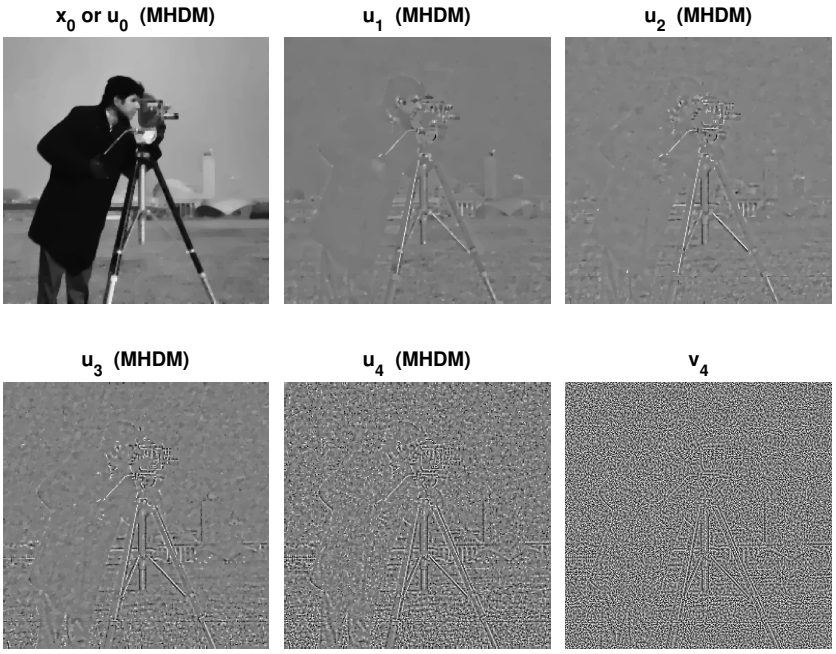

Figure 5: MHDM decomposition components for denoising the image with Gaussian noise (variance = 10^{-3}), $\delta = 2045$.

7.2.2 Tighter MHDM for denoising

The tighter version of MHDM is first used to denoise the same noisy images as we used in subsection 7.2.1. As shown in Fig. 6, the tighter MHDM gives similar results as MHDM. In this case, k^* is determined by equation (4.16). Compared with k_{min}, k^* is very well predicted. Here $k^* = k_{min} - 1$ for variance $= 10^{-2}, 10^{-3}, 10^{-4}$, and $k^* = k_{min}$ for variance $= 10^{-5}$.

The convergence behaviors of MHDM and tighter MHDM with respect to δ are compared in Fig. 7. The $RMSE_{k^*(\delta)}$ is defined as follows:

$$RMSE_{k^*(\delta)} = \frac{||x_{k^*(\delta)} - z||}{N} \tag{7.4}$$

where $z = f$ when $T = I$. The strict convergence is measured by the metric

$$d(x_{k^*(\delta)}, z) = ||x_{k^*(\delta)} - z||_1 + |J(x_{k^*(\delta))} - J(z)|. \tag{7.5}$$

From Fig. 7 we can see that the RMSE convergence rate of tighter MHDM is a little bit higher than MHDM. Fig. 7 also shows that tighter MHDM has much better strict convergence rate than MHDM.

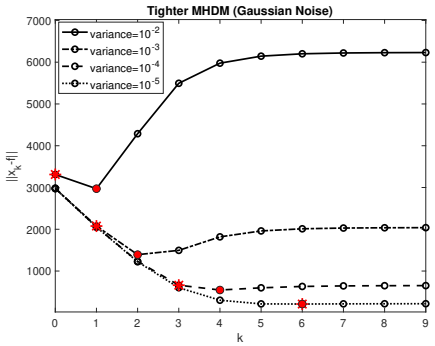

Figure 6: Errors of denoising results of the Cameraman image with different levels of Gaussian noise using the tighter MHDM. k^* is marked by red asterisk, k_{min} is marked by red circle.

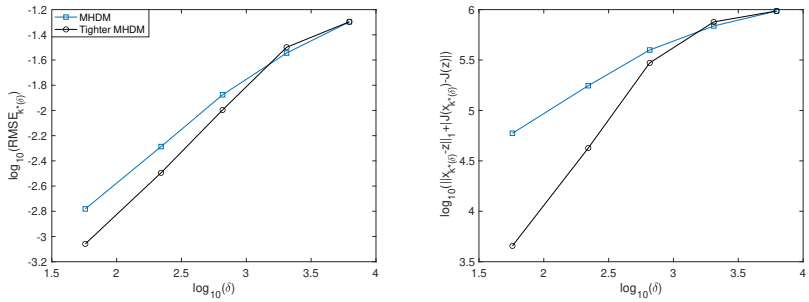

Figure 7: Comparison of convergence rates and strict convergence rates of MHDM and tighter MHDM for denoising based on the index stopping rule.

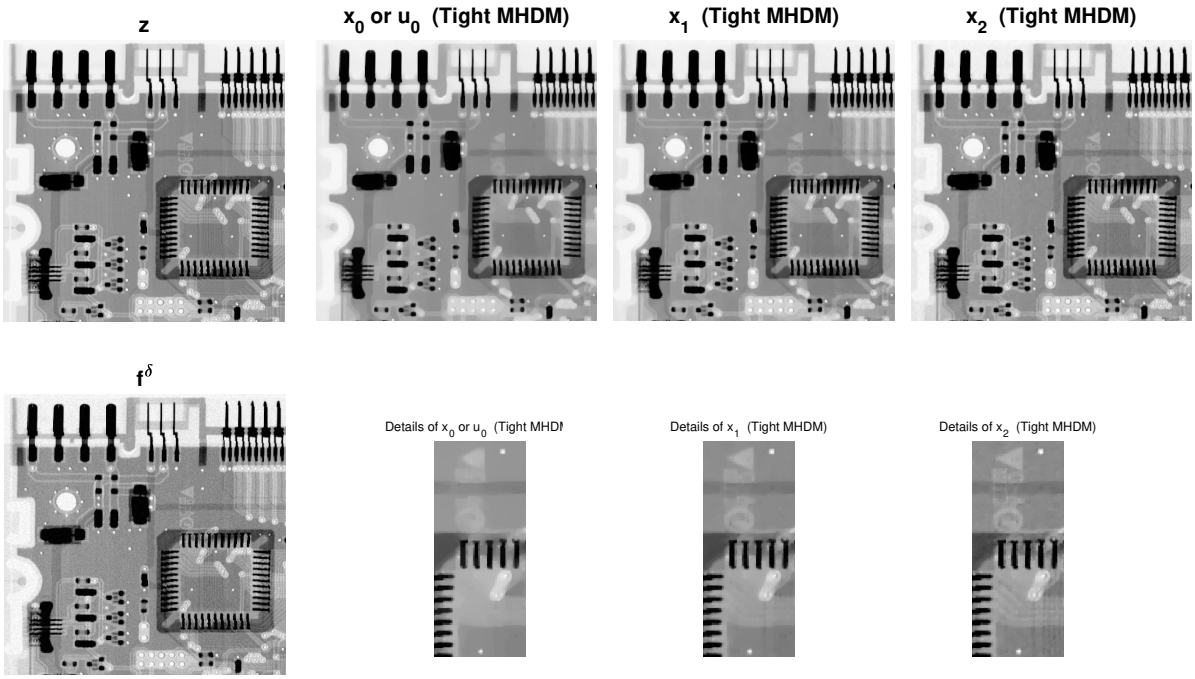

Figure 8: Denoising results of the circuit image with Gaussian noise (variance = 10^{-3}) using tighter MHDM. Here $\delta = 3559$, $N = 448 \times 464$. (Original image courtesy of Mr. Joseph E. Pascente, Lixi, Inc.)

In order to further validate the effectiveness of the tighter MHDM and its early stopping criterion for image denoising, we use it to restore another noisy image f^δ in Fig. 8. The noisy image f^δ is degraded by adding Gaussian noise with variance = 10^{-3} to the original circuit board image z. In this example, $k^* = 1$ and $k_{min} = 2$. When $k = 0, 1, 2$, $RMSE = 0.0197, 0.0136, 0.0106$, $d(x_k, z) = 2.33 \times 10^6, 1.70 \times 10^6, 1.19 \times 10^6$ respectively. From Fig. 8, we can see that the details of the circuit board in x_1 are much clearer than in x_0. It can be observed that x_1 retains most of the edge information. Comparing x_1 with x_2 in Fig. 8, we can see that the loss of edge details in x_1 is acceptable. So the early stopping rule is very reasonable.

7.3 MHDM and tighter MHDM for deblurring-denoising

In this subsection, both methods are applied and compared when restoring images degraded by both noise and blur. First, we still use the same original Cameraman image z from Fig. 2 and Fig. 4, where 5×5 window Gaussian blur (of variance 2) and zero-mean Gaussian noise of different variance levels are added. All the parameters have the same values as in subsection 7.1.

The $RMSE$ and $d(x_k, z) = ||x_k - z||_1 + |J(x_k) - J(z)|$ versus k are shown in Fig. 9. Noise variance values 10^{-3}, 10^{-4}, and 10^{-5} correspond to $\delta = ||f - f^\delta|| = 2048, 657, 219$ respectively. The phenomenon

of error accumulation is more obvious when δ is larger. Overall, the tighter MHDM gives better performance than MHDM. As shown in Fig. 10, both methods have similar $RMSE$ convergence rates, and the strict convergence behavior of tighter MHDM is better than MHDM.

When $\delta = 2048$, corresponding to the first two plots in Fig. 9, as well as the last square point in Fig. 10, there doesn't exist a $k \in \mathbb{N}$ satisfying equation (3.5) using MHDM. Therefore, we let $k^* = 0$ for MHDM in this case. This is also the reason why the curve of MHDM in Fig. 10 does not increase monotonically.

According to the curves of $RMSE$ and $d(x_k, z)$ in Fig. 9, the k^* for MHDM seems not so well predicted as for tight MHDM. However, from the 1-norm curves as shown in Fig. 11, we can see that k^* is close to the k that minimizes $||x_k - z||_1$. It is shown in Fig. 12 that $||x_{k^*(\delta)} - z||_1$ decreases as δ decreases.

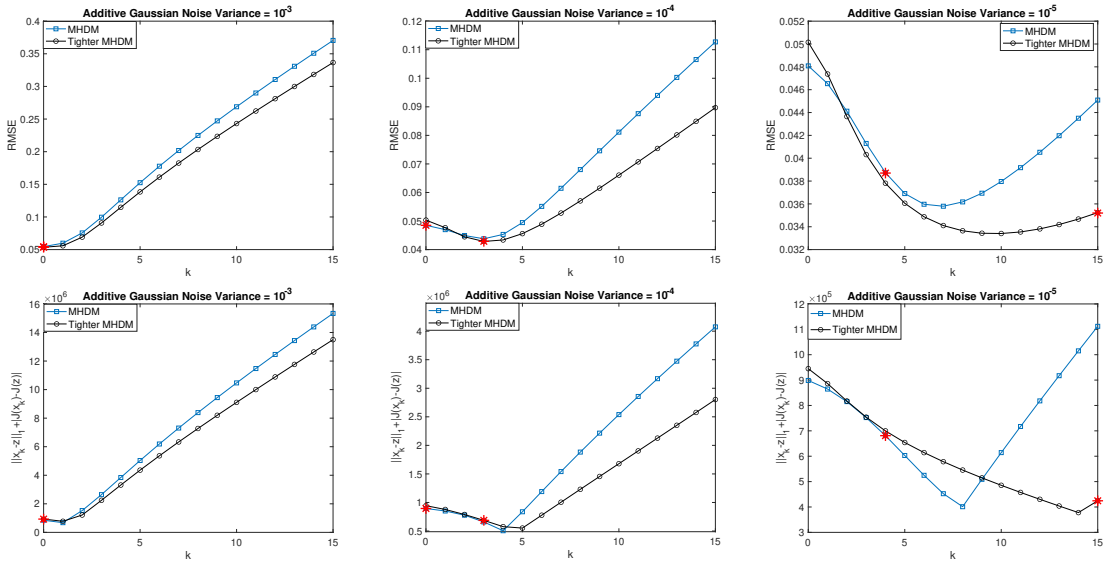

Figure 9: Errors of debluring-denoising results of image with Gaussian blur (variance=2) and zero-mean Gaussian noise with different variances. k^* is marked by red asterisk.

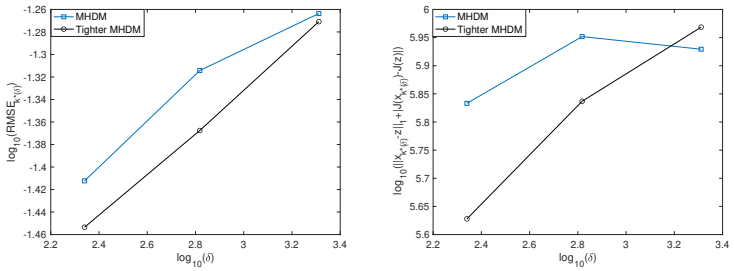

Figure 10: Comparison of convergence rate and strict convergence rate of MHDM and tighter MHDM for deblurring-denoising based on the index stopping rule.

Fig. 13 shows that the restored image x_{k^*} obtained by using MHDM is more blurred than the one using tighter MHDM, because k^* determined by equation (3.5) is less than that determined by equation (4.16). The detailed difference of v_{k^*} using both methods can also be noticed in Fig. 13. The restored images x_{15} ($k^* < 15$) using both methods obviously have more noise than x_{k^*}. In addition, the image x_{15} obtained by MHDM is noisier than the one obtained by tighter MHDM. Also, it can be seen in Fig. 9 that errors of x_k using MHDM are greater than those from tighter MHDM when k is relatively large. This illustrates that the tighter version can better suppress error propagation in this example.

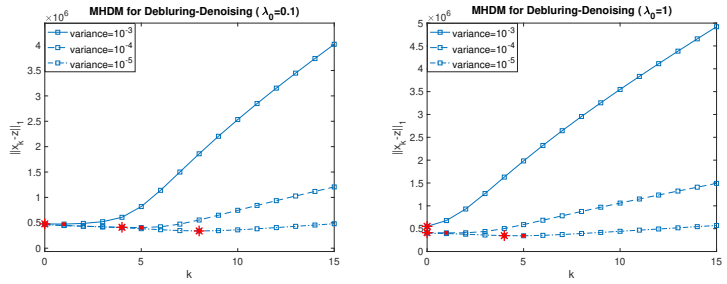

Figure 11: 1-norm errors of MHDM for deblurring-denoising the image with Gaussian blur (variance=2) and zero-mean Gaussian noise (variance=$10^{-3}, 10^{-4}, 10^{-5}$); k that has minimum $||x_k - z||_1$ is marked by red square, and k^* is marked by red asterisk.

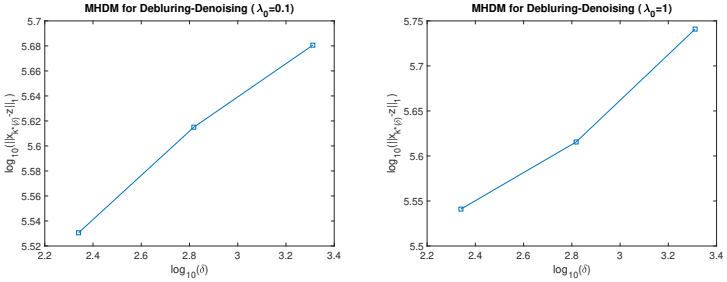

Figure 12: Convergence rate (1-norm) of MHDM for deblurring-denoising based on the index stopping rule.

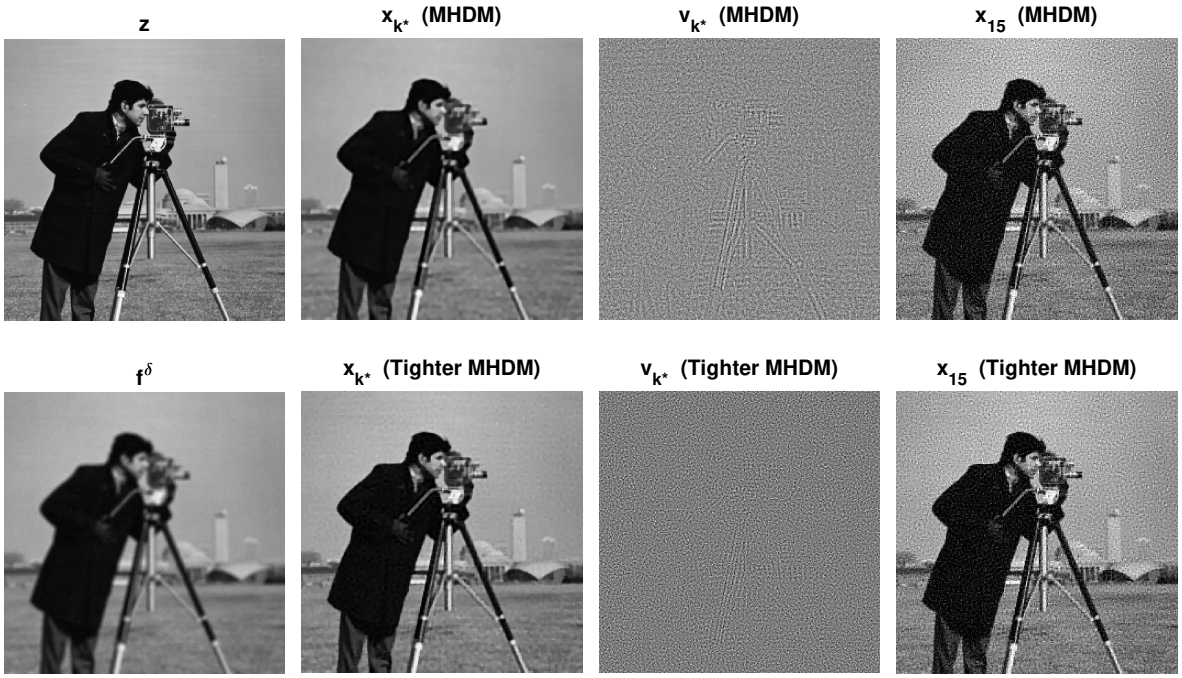

Figure 13: Deblurring-denoising results and v_{k^*} of both methods for the image with Gaussian blur (variance=2) and Gaussian noise (variance=10^{-4}, $\delta = 657$).

We restore another degraded image f^δ in Fig. 14 using both algorithms, where f^δ is obtained by applying Gaussian blur and additive normally distributed random noise to the original image z. One observes that x_{k^*} using tighter MHDM retains more edge information than using MHDM, which can be more clearly observed in Fig. 15. Here $RMSE_{k^*} = 0.030$, $d(x_{k^*}, z) = 5.93 \times 10^6$ for MHDM and $RMSE_{k^*} = 0.029$, $d(x_{k^*}, z) = 5.63 \times 10^6$ for tighter MHDM. It can also be seen in Fig. 15 that x_{10}

Figure 14: Deblurring-denoising results of the image with Gaussian blur (variance=2) and normally distributed random noise ($\delta = 1128$). Here $N = 512^2$, $\lambda_0 = 0.1$, $\Delta t = 0.01$, $k^* = 2$ for MHDM and $k^* = 3$ for tighter MHDM.

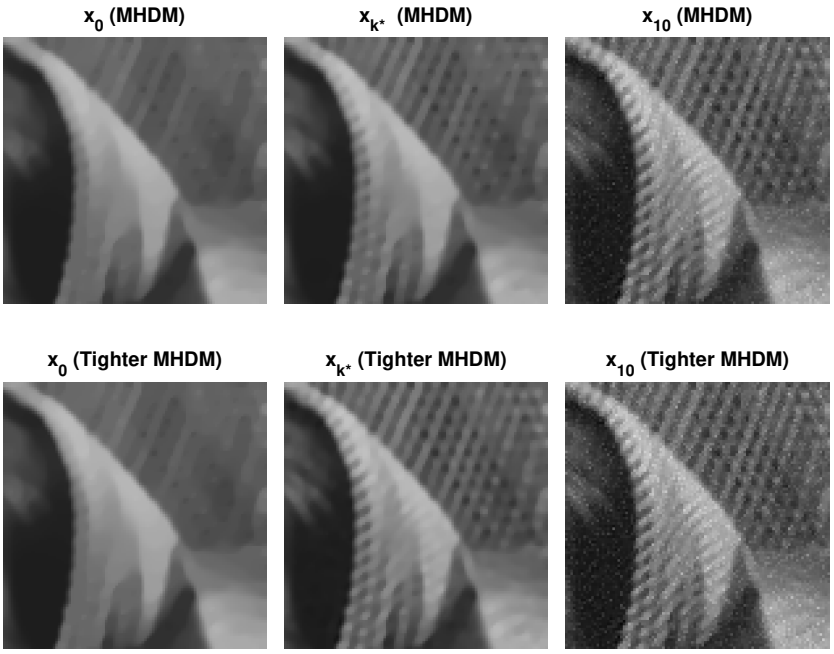

Figure 15: Right upper corner of the restoration of f^δ in Fig.14; $k^* = 2$ for MHDM and $k^* = 3$ for tighter MHDM.

shows more edge details, but is noisier than x_{k^*}. When $k = 10$, $RMSE = 0.032$, $d(x_{10}, z) = 3.85 \times 10^6$ for MHDM and $RMSE = 0.034$, $d(x_{10}, z) = 3.42 \times 10^6$ for tighter MHDM.

7.4 A refinement of the tight MHDM for deblurring-denoising

The effects of a refined version of the tight MHDM for deblurring-denoising are tested in this subsection by replacing each $R_k(u)$ in (5.1) with the weaker texture norm $||u||_*$, while for J the total variation seminorm is considered. Note that (5.3) is fulfilled in this case cf. [26, page 73].

To compare with MHDM and the original tight MHDM, the refined tight MHDM is first applied to deblurring-denoising the degraded image f^δ (blurred image with random noise) in Fig. 14. All the parameters, including iteration number, used in this example are the same as in the original tight MHDM. Numerical results show that the restored image x_k obtained by the refined version can preserve better texture features than the other two methods for each k. By looking at the details of x_1 shown in Fig. 16 for example, we can see that the refined version restores texture of the image better than the other two methods. In addition, the refinement version also produces smaller $RMSE$ and metric $d(x_1, z)$.

We further compare these three versions of MHDM by deblurring and denoising Gaussian blurred (variance=2) and Gaussian noisy (variance=10^{-4}) images as shown in Fig. 17 and Fig. 18. Here, $\lambda_0 = 0.05$, $\Delta t = 0.025$ for restoring both images.

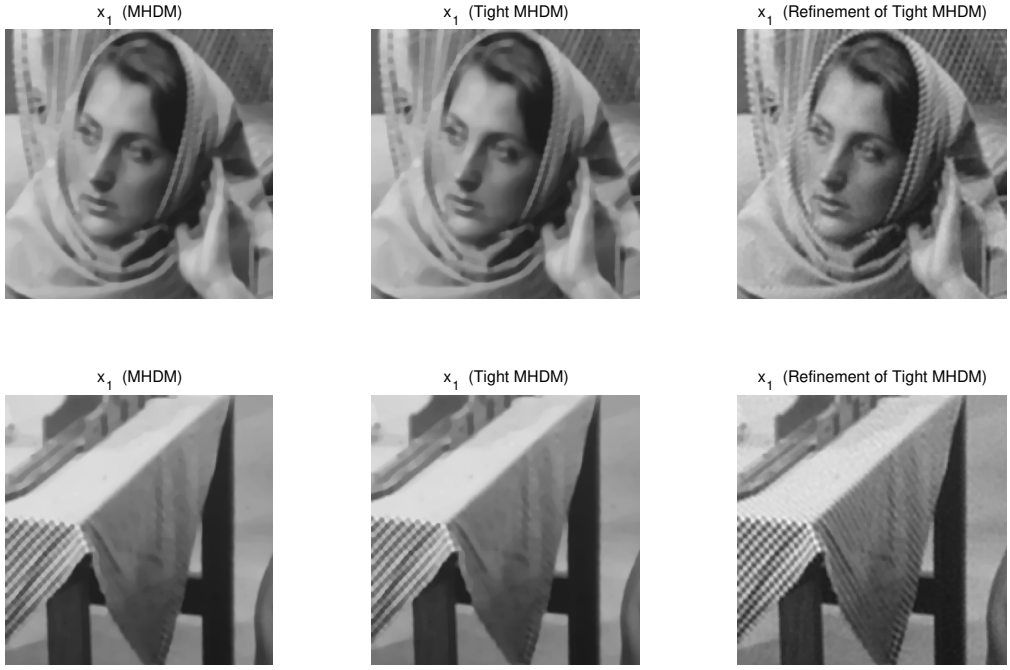

Figure 16: Details of the restoration images x_1 of f^δ in Fig. 14 using MHDM ($RMSE = 0.0304, d(x_1, z) = 6.06 \times 10^6$), tight MHDM ($RMSE = 0.0298, d(x_1, z) = 6.03 \times 10^6$), and the refinement of tight MHDM ($RMSE = 0.0293, d(x_1, z) = 5.53 \times 10^6$).

Fig. 17 shows the restored images x_{k^*} obtained by these methods following the stopping rules. The refined version still keeps better texture than the others, which can been seen from both x_{k^*} and the residual images v_{k^*}. It also has smaller restoration errors as explained in the title of Fig. 17. It cannot be ignored that the better texture and smaller errors of the refined version are also based on fewer iterations k than in the other two methods.

It is worth mentioning that the weaker norm $||u||_*$ can also preserve the noise, being of oscillatory nature. Therefore, to prevent error accumulation caused by noise, the refined version usually stops earlier than the other two versions, which means k^* of the refined version is smaller. This observation can be verified by both deblurring-denoising examples. In the tungsten filament image restoration example, $k^* = 4, 3, 2$ for MHDM, the tight MHDM, and the refined tight version respectively. In the pollen image

example, $k^* = 2$ for the refined version, which is also smaller than MHDM ($k^* = 5$) and the tight MHDM ($k^* = 4$).

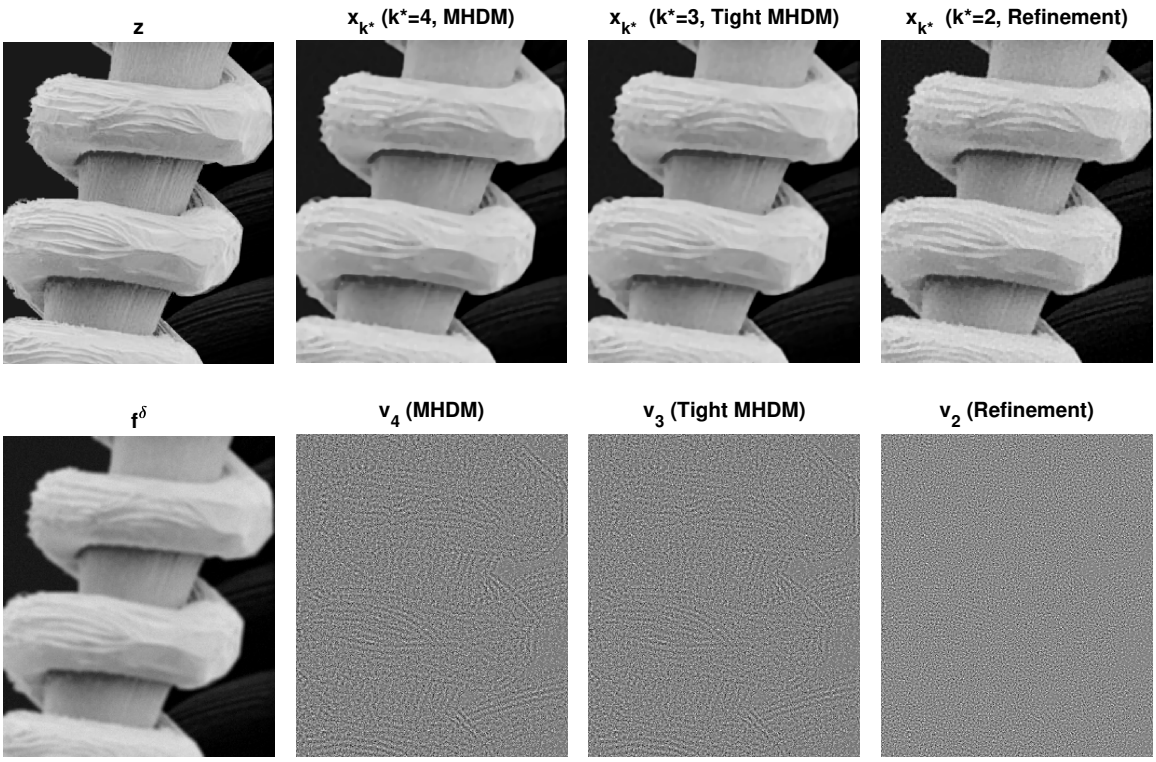

Figure 17: Degraded image of magnified tungsten filament and support with Gaussian blur and Gaussian noise. The restoration images x_{k^*} and residuals v_{k^*} of f^δ using MHDM ($RMSE_{x_4} = 0.0225, d(x_4, z) = 5.087 \times 10^5$), tight MHDM ($RMSE_{x_3} = 0.0220, d(x_3, z) = 4.937 \times 10^5$), and the refinement of tight MHDM ($RMSE_{x_2} = 0.0218, d(x_2, z) = 3.779 \times 10^5$). (Original image courtesy of Mr. Michael Shaffer, Department of Geological Sciences, University of Oregon, Eugene.)

The restoration errors of the pollen image in Fig. 18 can be seen in Fig. 19. These error plots also reveal the reason why the refined version stops earlier than the other two versions. According to the numerical results, using the refined version, the texture can be well preserved with small errors even in the very first several iterations of x_k. From Fig. 18 we can see that the restored image x_0 of the refined tight version is much better than the original tight version, and it is even better than x_2 of the tight MHDM. This can also be observed in the error plots in Fig. 19.

As seen in 6.1 and 6.2, calculation of the refined tight MHDM is more complicated than that of the tight MHDM. The execution CPU time of each iteration of the refined tight MHDM is about 45% longer than in the case of the tight MHDM. However, as mentioned before, the refined tight version stops earlier, so in fact the overall running time of the compact version is shorter than for the tight MHDM. Take the restoration in Fig. 18 for example, where the CPU running time of three iterations ($k^* = 2$) of refined tight MHDM is about 14% less than five iterations ($k^* = 4$) of tight MHDM.

We also apply the refined tight MHDM to restoring the degraded image f^δ with disk-shaped blur kernel [21] of radius $r = 3$ and Gaussian noise of standard deviation $\sigma = 0.01$, which is shown in Fig. 20. We compare the refined tight MHDM to the ROF total variation (TV) deconvolution with the split Bregman algorithm [21]. In [21], a method for estimating the optimal fidelity weight λ is proposed, and the optimal λ is determined by r and σ. However, the optimal value of λ is big when σ is small. As shown in Fig. 20, oscillation artifacts appear in the restoration image using TV deconvolution with the optimal $\lambda = 16443.9$, and $RMSE = 0.0726$. We tested the TV deconvolution with different values of λ. A better image restoration can be obtained with $\lambda = 991.65$, and $RMSE = 0.0243$. In contrast, the refined tight MHDM is insensitive to the value of λ_0. No matter if $\lambda_0 = 16443.9$ (which is the same as the optimal fidelity weight value of TV deconvolution) or $\lambda_0 = 0.05$, we can get a better image $x_{k^*}(k^* = 3)$, as shown

in the last column of Fig. 20. Both values of λ_0 yield $RMSE_{x_0} = 0.0243$ and $RMSE_{x_3} = 0.0210$, the latter being smaller than for the TV deconvolution. Thus, the refined tight MHDM is very robust and effective.

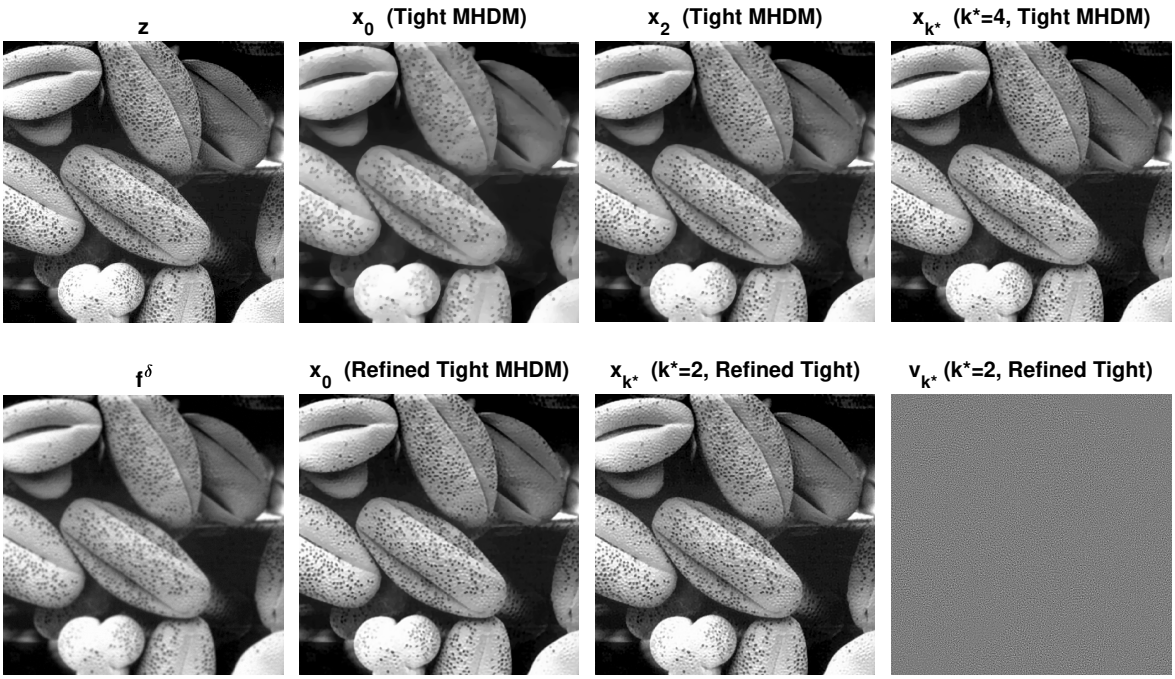

Figure 18: Degraded image of magnified pollen with Gaussian blur and Gaussian noise. Comparison of the restoration images of f^δ using the tight MHDM and the refined tight MHDM. (Original image courtesy of Dr. Roger Heady, Research School of Biological Sciences, Australian National University, Canberra, Australia.)

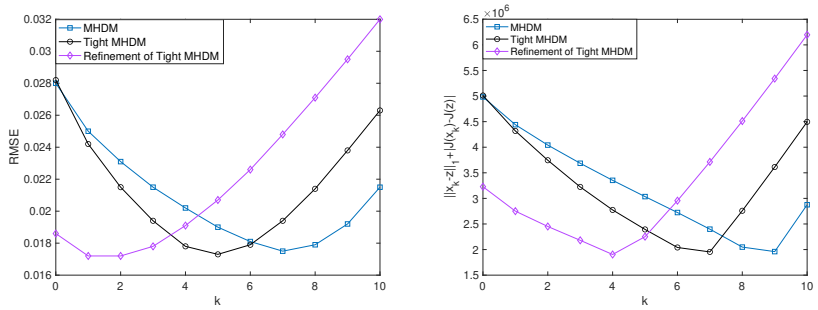

Figure 19: Errors of deblurring-denoising results of the degraded pollen image.

8 Conclusion

The theoretical findings in this work complete existing results about the multiscale hierarchical decomposition method previously proposed and studied in [37], [38], and [27]. More precisely, we derive error estimates and early stopping criteria in the case of perturbed or noisy data. These are supported by our numerical experiments, which also compare the original MHDM and the tighter version, and show that the tighter version performs better in practice in terms of convergence rate and restoration quality. Moreover, a refinement of the tight MHDM is proposed, which preserves texture better during the restoration process.

As future plans, we would like to extend this analysis to the case of nonquadratic data-fitting terms and to nonlinear problems. We plan also to develop and test more broadly the refined tight MHDM.

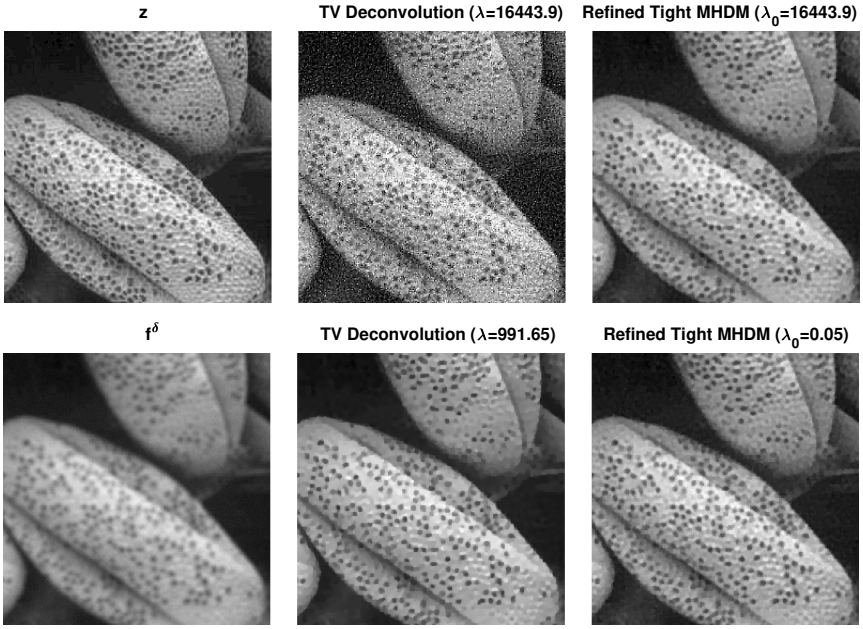

Figure 20: The first column: original image z and degraded image with disk-shape blur (radius=3) and Gaussian noise ($\sigma = 0.01$). The second column: restoration images using TV deconvolution with the split Bregman algorithm with optimal $\lambda = 16443.9$ and $\lambda = 991.65$. The third column: restoration images $x_3(k^* = 3)$ using MHDM with $\lambda_0 = 16443.9$ and $\lambda_0 = 0.05$.

9 Acknowledgments

E.R. is grateful for the opportunity to work on this topic during her sabbatical spent at the UCLA Mathematics Department, hosted by Luminita Vese. Moreover, the proof reading done by Tram Thi Ngoc Nguyen (Karl Franzens University of Graz) is warmly acknowledged. L.V. has been supported by NSF-DMS award 2012868 while working on this project. The authors acknowledge the facilities and the scientific and technical assistance of Microscopy Australia at the Centre for Advanced Microscopy, Australian National University, a facility that is funded by the University and the Federal Government. Last but not least, the authors would like to thank the reviewers for the insightful comments and suggestions that have greatly increased the overall quality of the manuscript.

References

[1] R. Acar and C. R. Vogel, *Analysis of bounded variation penalty methods for ill-posed problems*, Inverse Problems, 10 (1994), p. 1217.

[2] L. Ambrosio, N. Fusco, and D. Pallara, *Functions of bounded variation and free discontinuity problems*, Courier Corporation, 2000.

[3] F. Andreu-Vaillo, V. Caselles, J. M. Mazón, and J. M. Mazón, *Parabolic quasilinear equations minimizing linear growth functionals*, vol. 223, Springer Science & Business Media, 2004.

[4] F. J. A. Artacho, R. Campoy, and M. K. Tam, *The Douglas-Rachford algorithm for convex and nonconvex feasibility problems*, Mathematical Methods of Operations Research, 91 (2020), pp. 201–240.

[5] P. Athavale, R. Xu, P. Radau, A. Nachman, and G. A. Wright, *Multiscale properties of weighted total variation flow with applications to denoising and registration*, Medical Image Analysis, 23 (2015), pp. 28–42.

[6] G. Aubert and J.-F. Aujol, *Modeling very oscillating signals. Application to image processing*, Applied Mathematics and Optimization, 51 (2005), pp. 163–182.

[7] G. Aubert and L. Vese, *A variational method in image recovery*, SIAM Journal on Numerical Analysis, 34 (1997), pp. 1948–1979.

[8] J.-F. Aujol, G. Aubert, L. Blanc-Féraud, and A. Chambolle, *Image decomposition application to SAR images*, in International Conference on Scale-Space Theories in Computer Vision, Springer, 2003, pp. 297–312.

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

[10] J.-F. Aujol and A. Chambolle, *Dual norms and image decomposition models*, International Journal of Computer Vision, 63 (2005), pp. 85–104.

[11] M. Burger, E. Resmerita, and L. He, *Error estimation for Bregman iterations and inverse scale space methods in image restoration*, Computing, 81 (2007), pp. 109–135.

[12] R. H. Chan, M. Tao, and X. Yuan, *Constrained total variation deblurring models and fast algorithms based on alternating direction method of multipliers*, SIAM Journal on Imaging Sciences, 6 (2013), pp. 680–697.

[13] G. Chung, T. M. Le, L. H. Lieu, N. M. Tanushev, and L. Vese, *Computational methods for image restoration, image segmentation, and texture modeling*, in Computational Imaging IV, vol. 6065, International Society for Optics and Photonics, 2006, p. 60650J.

[14] L. Condat, *A generic proximal algorithm for convex optimization-application to total variation minimization*, IEEE Signal Processing Letters, 21 (2014), pp. 985–989.

[15] I. Daubechies and G. Teschke, *Wavelet-based image decomposition by variational functionals*, in Wavelet Applications in Industrial Processing, vol. 5266, International Society for Optics and Photonics, 2004, pp. 94–105.

[16] I. Daubechies and G. Teschke, *Variational image restoration by means of wavelets: Simultaneous decomposition, deblurring, and denoising*, Applied and Computational Harmonic Analysis, 19 (2005), pp. 1–16.

[17] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of inverse problems*, vol. 375, Springer Science & Business Media, 1996.

[18] K. Frick, D. A. Lorenz, and E. Resmerita, *Morozov's principle for the augmented Lagrangian method applied to linear inverse problems*, Multiscale Modeling & Simulation, 9 (2011), pp. 1528–1548.

[19] A. Fu, J. Zhang, and S. Boyd, *Anderson accelerated Douglas-Rachford splitting*, SIAM Journal on Scientific Computing, 42 (2020), pp. A3560–A3583.

[20] J. B. Garnett, T. M. Le, Y. Meyer, and L. Vese, *Image decompositions using bounded variation and generalized homogeneous Besov spaces*, Applied and Computational Harmonic Analysis, 23 (2007), pp. 25–56.

[21] P. Getreuer, *Total variation deconvolution using split Bregman*, Image Processing On Line, 2 (2012), pp. 158–174.

[22] P. Giselsson and S. Boyd, *Linear convergence and metric selection for Douglas-Rachford splitting and ADMM*, IEEE Transactions on Automatic Control, 62 (2016), pp. 532–544.

[23] M. Hidane, O. Lézoray, V.-T. Ta, and A. Elmoataz, *Nonlocal multiscale hierarchical decomposition on graphs*, in European Conference on Computer Vision, Springer, 2010, pp. 638–650.

[24] M. Jung, E. Resmerita, and L. Vese, *Dual norm based iterative methods for image restoration*, Journal of Mathematical Imaging and Vision, 44 (2012), pp. 128–149.

[25] L. H. Lieu and L. Vese, *Image restoration and decomposition via bounded total variation and negative Hilbert-Sobolev spaces*, Applied Mathematics and Optimization, 58 (2008), pp. 167–193.

[26] Y. Meyer, *Oscillating patterns in image processing and nonlinear evolution equations: the fifteenth Dean Jacqueline B. Lewis memorial lectures*, vol. 22, American Mathematical Soc., 2001.

[27] K. Modin, A. Nachman, and L. Rondi, *A multiscale theory for image registration and nonlinear inverse problems*, Advances in Mathematics, 346 (2019), pp. 1009–1066.

[28] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin, *An iterative regularization method for total variation-based image restoration*, Multiscale Modeling & Simulation, 4 (2005), pp. 460–489.

[29] S. Osher, A. Solé, and L. Vese, *Image decomposition and restoration using total variation minimization and the* H^{-1} *norm*, Multiscale Modeling & Simulation, 1 (2003), pp. 349–370.

[30] D. C. Paquin, D. Levy, E. Schreibmann, and L. Xing, *Multiscale image registration*, Mathematical Biosciences and Engineering, 3 (2006), p. 389.

[31] L. Rudin and V. Caselles, *Image recovery via multiscale total variation*, in Proceedings of the Second European Conference on Image Processing, Palma, Spain, Citeseer, 1995.

[32] L. I. Rudin and S. Osher, *Total variation based image restoration with free local constraints*, in Proceedings of 1st International Conference on Image Processing, vol. 1, IEEE, 1994, pp. 31–35.

[33] L. I. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Physica D: Nonlinear Phenomena, 60 (1992), pp. 259–268.

[34] M. Schloegl, M. Holler, A. Schwarzl, K. Bredies, and R. Stollberger, *Infimal convolution of total generalized variation functionals for dynamic MRI*, Magnetic Resonance in Medicine, 78 (2017), pp. 142–155.

[35] J.-L. Starck, M. Elad, and D. L. Donoho, *Image decomposition via the combination of sparse representations and a variational approach*, IEEE Transactions on Image Processing, 14 (2005), pp. 1570–1582.

[36] E. Tadmor and P. Athavale, *Multiscale image representation using novel integro-differential equations*, Inverse Problems and Imaging, 35 (2009), p. 693.

[37] E. Tadmor, S. Nezzar, and L. Vese, *A multiscale image representation using hierarchical (BV,* L^2*) decompositions*, Multiscale Modeling & Simulation, 2 (2004), pp. 554–579.

[38] E. Tadmor, S. Nezzar, and L. Vese, *Multiscale hierarchical decomposition of images with applications to deblurring, denoising, and segmentation*, Communications in Mathematical Sciences, 6 (2008), pp. 281–307.

[39] L. Tang and C. He, *Multiscale variational decomposition and its application for image hierarchical restoration*, Computers & Electrical Engineering, 54 (2016), pp. 354–369.

[40] M.-P. Tran and T. N. Nguyen, *Image decomposition using a second-order variational model and wavelet shrinkage*, ELCVIA: Electronic Letters on Computer Vision and Image Analysis, 18 (2019), pp. 0092–107.

[41] L. Vese, *A study in the BV space of a denoising-deblurring variational problem*, Applied Mathematics and Optimization, 44 (2001), pp. 131–161.

[42] L. Vese and S. Osher, *Modeling textures with total variation minimization and oscillating patterns in image processing*, Journal of Scientific Computing, 19 (2003), pp. 553–572.

[43] M. Zhong, *Hierarchical Reconstruction Method for Solving Ill-posed Linear Inverse Problems*, PhD thesis, University of Maryland, 2016.