(BV, L^2) Multiscale Hierarchical Decomposition: Modes and Rates of Convergence

Ming-Jun Lai

Department of Mathematics

The University of Georgia

Athens, GA 30602

Leopold Matamba Messi

Mathematical Biosciences Institute

The Ohio State University

Columbus, OH 43210

Abstract

Tadmor, Nezzar and Vese [Eitan Tadmor, Suzanne Nezzar, and Luminita Vese. A multiscale image representation using hierarchical (BV, L^2) decompositions. *Multiscale Model. Simul.*, 2(4):554–579, 2004.] developed a total variation based multiscale method for decomposing a function $f \in BV$ into a countable set of features $\{u_k : k = 0, 1, 2 \ldots\}$ associated to a sequence of dyadic scales $\{\lambda_k = \lambda_0 2^{-k} : k = 0, 1, 2, \ldots\}$ such that for each k, $[u_{k+1}, v_{k+1}] = \arg\min\{\lambda_k |Du| + \|v\|_{L^2} : u + v = f\}$. They showed that $f = \sum_{k=0}^{\infty} u_k$ in $L^2(\Omega)$ and strongly in $W^{-1,\infty}(\Omega)$. In this paper, we study the convergence of the series $\sum_{k=0}^{\infty} u_k$ in the weak*, strict and normed topologies of the space of functions with bounded variation. We show that in general, the convergence of the series $f = \sum_{k=0}^{\infty} u_k$ in any of the three topologies of BV is conditioned by its rate of convergence in L^2, and prove that the convergence in L^2 is geometric.

1 Introduction

The pixel values of a digital image are samples of an intensity field $F(x)$ defined on a planar domain Ω, however, pixel values systematically confound the various scales represented in the original intensity field. Nonetheless, the more noticeable features of the intensity field are preserved to some extent, and readily identified by human subjects. For example, edges and textures are well preserved by most sampling processes and humans use that information to classify and identify objects in a digital image.

1

Intensity fields could be realized as square integrable functions. A central question in digital image processing is then to identify appropriate subsets of the class of L^2 intensity fields that encode the most noticeable features of a digital image. A typical digital representation of a natural scene contains flat regions of varying intensities; the interface between these regions is made edges, thus the intensity field from which the image was sampled is presumably with bounded variation on Ω. Indeed, many digital image enhancement methods modeled on BV have been proposed since the seminal work of Rudin, Osher and Fatemi [5]. The explosion of total variation based techniques in the mathematical image analysis community has led Gousseau and Morel [3] to question the universality of the total variation model for natural images; thus concluding that natural images are overwhelming not of bounded variation and providing a plausible argument for why these methods are unable to perform adequately on natural images that are rich on textures.

The Rudin-Osher-Fatemi (ROF) total variation approach to image denoising consist in extracting the critical features of an image f at a scale $\lambda > 0$ by decomposing $f = u_\lambda + v_\lambda$ such that

$$[u_\lambda, v_\lambda] = \arg \inf_{f=u+v} \left\{ \lambda |Du| + \|v\|_2^2 \colon u \in BV(\Omega), \ v \in L^2(\Omega) \right\}.$$

For a good choice of scale λ, the features function u_λ represents the cleaned image and v_λ the noise. The ROF model seeks to achieve reasonable fidelity to the image f while capturing enough of the features to the extent that they are measured by a total variation proportionally to $1/\lambda$. We observe that if λ is too small, then $\|v_\lambda\|_{L^2}$ is small and u_λ captures most of the features in f including those encoding noise. However, if λ is chosen too big, then the model leads to a decomposition in which the total variation of u_λ is too small and only a cartoon version of f is recovered in u_λ with most of edges blurred away. So for adequate performance, one should choose a consensus value of λ, hence the practical challenge in using the ROF model in practice.

To lessen the impact of choosing the consensus λ, Tadmor, Nezzar and Vese [6] proposed a hierarchical multiscale decomposition procedure based on the ROF decomposition. The starting point is Meyer's interpretation [4] of the ROF model as a decomposition of f into edges u_λ and textures v_λ. These scholars intuit that the concepts of edge and texture are scale dependent. For instance at scale λ_0, the ROF model yields the decomposition $f = u_0 + v_0$ of f into edges u_0 and textures v_0; the textures v_0 at the scale λ_0 consists of edges and textures at a finer scale, say $\lambda_0/2$. One can then repeat the ROF decomposition

on v_0 at scale $\lambda_0/2$ to get

$$[u_1, v_1] = \arg \inf_{v_0=u+v} \left\{ \frac{\lambda_0}{2}|Du| + \|v\|_2^2 : \ u \in BV(\Omega), \ v \in L^2(\Omega) \right\},$$

and we now have an improved two-scale reconstruction of f defined by $f = u_0 + u_1 + v_1$, with textures below scale $\lambda_0/2$ now captured in v_1. By iterating the dyadic refinement above, one generates a (BV, L^2) multiscale approximation of f over a dyadic cascade of scales

$$f \approx u_0 + u_1 + \ldots + u_k + \ldots,$$

where for any nonnegative integer $i \geq 0$

$$[u_{i+1}, v_{i+1}] = \arg \inf_{v_i=u+v} \left\{ \frac{\lambda_0}{2^i}|Du| + \|v\|_2^2 : \ u \in BV(\Omega), \ v \in L^2(\Omega) \right\}. \tag{1}$$

In this paper, we revisit the question of convergence of (BV, L^2) hierarchical multiscale decompositions in the space of functions with bounded variation. In section 2, we review properties of the ROF model that are relevant to this work and recall the definition of the three topologies on the space of functions with bounded variations that we seek convergence in. Section 3 is devoted to an alternate proof of the convergence of the (BV, L^2) hierarchical approximation. We establish a new result on the convergence rate of the hierarchical decomposition in $L^2(\Omega)$ and give a necessary condition for the convergence in $BV(\Omega)$.

2 Preliminaries

In this section, we review three topologies on the space $BV(\Omega)$ of functions with bounded variation and a characterization of the ROF model of image decomposition. In the sequel Ω is either a convex polygonal domain of $\mathbb{R}^2$ or $\Omega = \mathbb{R}^2$, unless otherwise specified.

2.1 Functions with bounded variation

The total variation of a function u defined on Ω is defined by

$$|Du| := \sup \left\{ -\int_\Omega u \operatorname{div}(\varphi)dx : \varphi \in C_c^1(\Omega, \mathbb{R}^2), |\varphi(x)| \leq 1, \ \forall x \in \Omega \right\}. \tag{2}$$

A function is said to be with bounded variation on Ω if $|Du|$ is finite. The space of functions with bounded variation functions, $BV(\Omega)$, is the subspace of integrable functions on Ω with finite total variation:

$$BV(\Omega) := \left\{ u \in L^1(\Omega) : |Du| < \infty \right\}. \tag{3}$$

The space $BV(\Omega)$ is a Banach space for the norm

$$\|u\|_{BV} = \int_\Omega |u| dx + |Du|; \tag{4}$$

we will refer to the corresponding topology on $BV(\Omega)$ as the *strong topology*.

Definition 2.1 (Strong Convergence). A sequence $\{u_n\}$ in $BV(\Omega)$ converges strongly to $u \in BV(\Omega)$ if

$$\lim_{n \to \infty} \|u_n - u\|_{BV} = 0.$$

Remark 2.2. The total variation defines a semi-norm on $BV(\Omega)$. In particular by the Poincaré inequality, the total variation $|Du|$ is a norm on

$$BV_0(\Omega) := \{u \in BV(\Omega) : \int_\Omega u dx = 0\}$$

that is equivalent to the norm $\|u\|_{BV}$.

Alternatively, one could define the space $BV(\Omega)$ as the subspace to functions $u \in L^1(\Omega)$ such that the weak gradient $Du = \left(\dfrac{\partial u}{\partial x_1}, \dfrac{\partial u}{\partial x_2} \right)$ is a vector valued Radon measure on Ω. In this interpretation, $BV(\Omega)$ can be endowed with a topology that combines the strong topology on $L^1(\Omega)$ and the weak* topology on the space of vector valued Radon measures, leading to the notion of *weak* topology*.

Definition 2.3 (weak* convergence). A sequence $\{u_n\}$ in $BV(\Omega)$ weakly* converges to $u \in BV(\Omega)$ if u_n converges strongly to u in $L^1(\Omega)$ and Du_n converges weakly* to Du, i.e

$$\lim_{n \to \infty} \int_\Omega \varphi \, dDu_n = \int_\Omega \varphi \, dDu, \quad \forall \varphi \in C_c^0(\Omega, \mathbb{R}^2).$$

A simple criterion for proving weakly* convergence is as follows

4

Proposition 2.4 ([1, Proposition 3.13]). *A sequence $\{u_n\} \subset BV(\Omega)$ weakly* converges to $u \in BV(\Omega)$ if and only if $\{|Du_n|\}$ is bounded as a sequence of numbers and $\int_\Omega |u_n - u|dx \to 0$ as $n \to \infty$.*

Finally, the space $BV(\Omega)$ can be endowed with a metric space structure where the distance is defined by

$$d(u, v) = \int_\Omega |u - v|dx + \Big||Du| - |Dv|\Big| \tag{5}$$

The metric space topology induced by the distance d above is called the *strict topology.*

Definition 2.5. A sequence $\{u_n\}$ in $BV(\Omega)$ strictly converges to $u \in BV(\Omega)$ if

$$\lim_{n\to\infty} \int_\Omega |u_n - u|dx = 0 \text{ and } \lim_{n\to\infty} |Du_n| = |Du|. \tag{6}$$

Remark 2.6. The three modes on convergence defined above are related as follows:

$$\text{Strong Convergence} \implies \text{Strict Convergence} \implies \text{Weak* Convergence;} \tag{7}$$

all converses being false.

2.2 Relevant properties of the ROF decomposition model

The ROF image decomposition model consists in extracting the critical features of an image f at a scale $\lambda > 0$ by decomposing $f = u_\lambda + v_\lambda$ such that

$$[u_\lambda, v_\lambda] = \arg\inf_{f=u+v} \Big\{\lambda|Du| + \|v\|_2^2 : u \in BV(\Omega), \ v \in L^2(\Omega)\Big\}. \tag{ROF}$$

The existence and uniqueness of the pair $[u_\lambda, v_\lambda]$ is an interesting exercise in convex analysis, see Chambolle et al. [2] and the references therein. In particular, the optimal pair $[u_\lambda, v_\lambda]$ is characterized as follows.

Theorem 2.7 ([2]). *Let $\lambda > 0$ and $f \in L^2(\Omega)$ be fixed. Then, the pair of functions $f = u_\lambda + v_\lambda$ is a solution of the convex optimization problem* (ROF) *if and only if there exists*

$z \in L^\infty(\Omega, \mathbb{R}^2)$ *such that*

$$\begin{cases} -\dfrac{\lambda}{2} \operatorname{div}(z(x)) + u_\lambda(x) = f(x) & \text{a.e. } x \in \Omega, \\[2mm] |z(x)| \leq 1 & \text{a.e. } x \in \Omega, \\[2mm] z \cdot \nu = 0 & \text{weakly on } \partial\Omega, \\[2mm] \displaystyle\int_\Omega -\operatorname{div}(z) u_\lambda \, dx = |Du_\lambda|. \end{cases} \tag{8}$$

Let $E_\lambda(u)$ denote the objective functional of the ROF model, i.e.

$$E_\lambda(u) = \lambda |Du| + \int_\Omega |u - f|^2 dx. \tag{9}$$

We infer from the characterization of the optimal pair above the following extremal value identity.

Corollary 2.8. *Let* $f = u_\lambda + v_\lambda$ *be the optimal decomposition of* f *per the ROF model. Then, we have*

$$E_\lambda(u_\lambda) = \|f\|_2^2 - \|u_\lambda\|_2^2. \tag{10}$$

Proof. By Theorem 2.7 above, we have

$$|Du_\lambda| = \frac{2}{\lambda} \int_\Omega v_\lambda u_\lambda dx$$

and it follows that

$$E_\lambda(u_\lambda) = 2 \int_\Omega v_\lambda u_\lambda dx + \int_\Omega v_\lambda^2 dx = \int_\Omega (u_\lambda + v_\lambda)^2 - u_\lambda^2 \, dx$$
$$= \|f\|_2^2 - \|u_\lambda\|_2^2 \quad \text{since } f = u_\lambda + v_\lambda,$$

which is the desired result. $\qquad\qquad\square$

We conclude the section by showing that the ROF model (for Ω bounded) preserves the average value of functions with zero mean value.

Proposition 2.9. *Suppose* Ω *is a bounded domain. Let* $f \in L^2(\Omega)$ *and* $\lambda > 0$ *be fixed. If* $f = u_\lambda + v_\lambda$ *is the ROF decomposition of* f *at the scale* λ, *then* $\displaystyle\int_\Omega u_\lambda \, dx = \int_\Omega f \, dx$.

Proof. Since u_λ minimizes E_λ over $BV(\Omega)$, it follows that for any constant $c \in \mathbb{R}$, we have

$$\lambda|Du_\lambda| + \int_\Omega |f - u_\lambda|^2 \le \lambda|D(u_\lambda + c)|\,dx + \int_\Omega |f - (u_\lambda + c)|^2\,dx$$

$$\le \lambda|D(u_\lambda)| + \int_\Omega |f - (u_\lambda + c)|^2\,dx,$$

that is,

$$0 \le 2c \int_\Omega (f - u_\lambda)dx + |\Omega|\,c^2 \quad \forall\, c \in \mathbb{R}$$

where $|\Omega|$ is the size of Ω. Consequently, $\int_\Omega (f - u_\lambda)\,dx = 0$ and the proof is complete. $\quad\square$

3 (BV, L²) hierarchical decomposition

Let $\lambda_0 > 0$ and $f \in L^2(\Omega)$ be fixed. The Tadmor-Nezzar-Vese (TNV) dyadic sequence $\{u_n : n = 0, 1, 2, \ldots\}$ for f with base scale λ_0 is given by the recurrence relation

$$u_0 = \arg\inf_{u \in BV(\Omega)} \left\{E_0(u) := \lambda_0|Du| + \|f - u\|_2^2\right\}, \tag{11}$$

and for any $n \ge 0$

$$u_{n+1} = \arg\inf_{u \in BV(\Omega)} \left\{E_{n+1}(u) := \frac{\lambda_0}{2^{n+1}}|Du| + \left\|f - \sum_{i=0}^{n} u_i\right\|_2^2\right\}. \tag{12}$$

Let us quickly review some facts about the (BV, L^2) hierarchical decomposition that are readily obtained from the properties of the ROF model highlighted in the previous section. Firstly, in view of Theorem 2.7, for all $n = 0, 1, 2, \ldots$, $f - \sum_{i=0}^{n} u_i$ defines a bounded linear functional on the Sobolev spaces $W_0^{1,1}(\Omega)$, as such $f - \sum_{i=0}^{n} u_i \in W^{-1,\infty}(\Omega)$. It is readily deduced from (8) and the definition of u_n that

$$\left\|f - \sum_{i=0}^{n} u_i\right\|_{W^{-1,\infty}} \le \lambda_0 2^{-n-1}, \qquad \forall n = 0, 1, 2 \ldots. \tag{13}$$

Thus, for any $f \in L^2(\Omega)$ and $\lambda_0 > 0$, the (BV, L^2) dyadic multiscale series $\sum_{i=0}^{\infty} u_i$ converges to f in $W^{-1,\infty}(\Omega)$ with a rate of $O(2^{-n})$.

Secondly, thanks to the minimum value identity (10) applied to u_{k+1}, we have

$$\left\| f - \sum_{i=0}^{k+1} u_i \right\|_2^2 \le E_{k+1}(u_{k+1}) \le \left\| f - \sum_{i=0}^{k} u_i \right\|_2^2. \tag{14}$$

Thus, the sequence $\{\| f - \sum_{i=0}^{k} u_i \|_2 : i = 0, 1, 2, \ldots\}$ is monotonically nonincreasing. This suggests that the hierarchical decomposition may converge to f in $L^2(\Omega)$ in some cases. Indeed, Tadmor, Nezzar and Vese proved the following.

Theorem 3.1 ([6, Theorem 2.2]). *Suppose that $f \in BV(\Omega)$. Then, the series $\sum_{k=0}^{\infty} u_k$ converges strongly to f in $L^2(\Omega)$ and the energy of f in $L^2(\Omega)$ satisfies*

$$\|f\|_2^2 = \sum_{k=0}^{\infty} \|u_k\|_2^2 + \lambda_0 \sum_{k=0}^{\infty} 2^{-k} |Du_k|. \tag{15}$$

Proof. When studying the convergence rate of a TNV sequence below, we will give a new proof of this theorem. □

The following result is an observation about the average values of the terms of TNV sequences. Namely, we assert that when Ω is bounded all but the first term of a TNV sequence must have zero as mean value.

Proposition 3.2. *Suppose that Ω is bounded. Let $f \in L^2(\Omega)$ and $\lambda_0 > 0$ be fixed. If $\{u_n\}_n$ is the TNV dyadic sequence for f with base scale λ_0, then*

$$\int_{\Omega} u_0 dx = \int_{\Omega} f dx \text{ and } \int_{\Omega} u_n dx = 0 \quad \forall n \ge 1. \tag{16}$$

Proof. Indeed by Proposition 2.9, we have

$$\int_{\Omega} u_0 dx = \int_{\Omega} f dx \text{ and } \int_{\Omega} u_n dx = \int_{\Omega} f - \sum_{i=0}^{n-1} u_i \, dx, \quad n = 1, 2, \ldots.$$

By induction on $n = 1, 2, \ldots$, we get

$$\int_{\Omega} u_n dx = \int_{\Omega} f - u_0 \, dx = 0 \quad \forall n \in \mathbb{N},$$

and the proof is complete. □

3.1 Convergence rate in the norm of L^2

In this section, we establish a convergence rate to complement Theorem 3.1. Before proving our first contribution of this paper, let us first establish a lemma that is the cornerstone of our result.

Lemma 3.3. *Let a sequence $\{a_k\}_{k \geq 0}$ of nonnegative real numbers be fixed. Then, for any $\delta \in (0, 1)$ and for all $0 \leq \rho \leq 1/\delta - 1$, there holds*

$$\sum_{k=0}^{n} \delta^k a_k \geq \rho \sum_{k=0}^{n} \delta^k \sum_{j=0}^{k-1} a_j, \quad \forall\, n \geq 1. \tag{17}$$

Proof. Let $n \geq 1$ and $\delta \in [0, 1)$ be fixed. Then,

$$\sum_{k=0}^{n} \delta^k \sum_{j=0}^{k-1} a_j = \sum_{k=0}^{n-1} a_k \sum_{j=k+1}^{n} \delta^j$$

$$= \sum_{k=0}^{n-1} a_k \delta^k \sum_{j=1}^{n-k} \delta^j = \sum_{k=0}^{n-1} a_k \delta^k \frac{\delta(1 - \delta^{n-k})}{1 - \delta}.$$

Consequently, for any $\rho \geq 0$ we have

$$\sum_{k=0}^{n} \delta^k a_k - \rho \sum_{k=0}^{n} \delta^k \sum_{j=0}^{k-1} a_j = \sum_{k=0}^{n} \delta^k a_k - \rho \sum_{k=0}^{n-1} a_k \delta^k \frac{\delta(1 - \delta^{n-k})}{1 - \delta}$$

$$= \sum_{k=0}^{n} \delta^k a_k - \frac{\rho \delta}{1 - \delta} \sum_{k=0}^{n-1} a_k \delta^k + \frac{\rho \delta^n}{1 - \delta} \sum_{k=0}^{n-1} a_k$$

$$\geq \delta^n a_n + (1 - \frac{\rho \delta}{1 - \delta}) \sum_{k=0}^{n-1} a_k \delta^k \quad \text{(dropping the rightmost term).}$$

In particular, if $0 \leq \rho \leq 1/\delta - 1$, then $1 - \frac{\rho \delta}{1-\delta} \geq 0$ and it follows that

$$\sum_{k=0}^{n} \delta^k a_k - \rho \sum_{k=0}^{n} \delta^k \sum_{j=0}^{k-1} a_j \geq \delta^n a_n \geq 0.$$

Since n was arbitrarily chosen, we get inequality (17) and the lemma is proved. $\qquad \square$

We are now ready to state and prove our first result of this paper complementing Theorem 3.1 with a geometric convergence rate.

Theorem 3.4. *Suppose that $f \in BV(\Omega)$ and let $\lambda_0 > 0$ be fixed. Let $\{u_n\}$ be the sequence of functions defined by the recurrence relation* (11)–(12). *Then, for any $\eta \in \left(\frac{9-\sqrt{17}}{8}, 1 \right]$, the series $\sum_{k=0}^{\infty} \eta^{-k} \left\| f - \sum_{i=0}^{k} u_i \right\|_2^2$ converges. Furthermore, there exists a positive constant C such that*

$$\left\| f - \sum_{i=0}^{k} u_i \right\|_2^2 \le C \left(\frac{9-\sqrt{17}}{8} \right)^{k+1} \qquad \forall\, k = 0, 1, 2, \ldots. \tag{18}$$

Proof. Let $\rho \ge 0$ and $\eta > 0$ be fixed. By definition of u_k and using the assumption that $f \in BV(\Omega)$, we have

$$\lambda_0 2^{-k} |Du_k| + \left\| f - \sum_{i=0}^{k} u_i \right\|_2^2 \le \lambda_0 2^{-k} \rho \left| D\!\left(f - \sum_{i=0}^{k-1} u_i \right) \right| + (1-\rho)^2 \left\| f - \sum_{i=0}^{k-1} u_i \right\|_2^2,$$

where we have used the convention that $\sum_{i=0}^{-1} u_i = 0$. Dividing the latter inequality by η^k and using the fact that $\left| D\!\left(f - \sum_{i=0}^{k-1} u_i \right) \right| \le |Df| + \sum_{i=0}^{k-1} |Du_i|$, we obtain

$$\lambda_0 (2\eta)^{-k} |Du_k| + \eta^{-k} \left\| f - \sum_{i=0}^{k} u_i \right\|_2^2 \le \lambda_0 (2\eta)^{-k} \rho |Df|$$

$$+ \underbrace{\lambda_0 (2\eta)^{-k} \rho \sum_{i=0}^{k-1} |Du_i|}_{(A)} + (1-\rho)^2\, \eta^{-k} \left\| f - \sum_{i=0}^{k-1} u_i \right\|_2^2. \tag{19}$$

Summing the latter inequality over k ranging from 0 to n, and moving the term (A) to the left yields

$$\lambda_0 \underbrace{\left(\sum_{k=0}^{n} (2\eta)^{-k} |Du_k| - \rho \sum_{k=0}^{n} (2\eta)^{-k} \sum_{i=0}^{k-1} |Du_i| \right)}_{(B)} + \sum_{k=0}^{n} \eta^{-k} \left\| f - \sum_{i=0}^{k} u_i \right\|_2^2 \le$$

$$\lambda_0 \rho |Df| \sum_{k=0}^{n} (2\eta)^{-k} + (1-\rho)^2 \underbrace{\sum_{k=0}^{n} \eta^{-k} \left\| f - \sum_{i=0}^{k-1} u_i \right\|_2^2}_{(C)}. \tag{20}$$

10

Next, by Lemma 3.3 with $\delta = \dfrac{1}{2\eta}$ and $0 \le \rho \le \dfrac{1}{\delta} - 1 = 2\eta - 1$, the term labeled (B) above is nonnegative provided that $\eta \ge 1/2$. Thus, dropping (B) and moving (C) to the left, inequality (20) reduces to

$$\eta^{-n}\Big\|f - \sum_{i=0}^{n} u_i\Big\|_2^2 - \left(\frac{(1-\rho)^2}{\eta} - 1\right)\sum_{k=0}^{n-1}\eta^{-k}\Big\|f - \sum_{i=0}^{k} u_i\Big\|_2^2 \le \frac{2\lambda_0\eta}{2\eta - 1}|Df| + \frac{(1-\rho)^2}{\eta}\|f\|_2^2,$$

where we used the convention that $\sum_{i=0}^{-1} u_i = 0$. In particular for $\rho = 2\eta - 1$, we get

$$\eta^{-n}\Big\|f - \sum_{i=0}^{n} u_i\Big\|_2^2 - \frac{4 - 9\eta + 4\eta^2}{\eta}\sum_{k=0}^{n-1}\eta^{-k}\Big\|f - \sum_{i=0}^{k} u_i\Big\|_2^2 \le \frac{2\lambda_0\eta}{2\eta - 1}|Df| + \frac{4(1-\eta)^2}{\eta}\|f\|_2^2.$$

$$(21)$$

Consequently, for any $\eta \ge 1/2$ such that $4 - 9\eta + 4\eta^2 \le 0$ and in particular for $\dfrac{9 - \sqrt{17}}{8} \le \eta \le 1$, we have

$$\Big\|f - \sum_{i=0}^{n} u_i\Big\|_2^2 \le C_0(\lambda_0|Df| + \|f\|_2^2)\,\eta^{n+1},$$

where C is a positive universal constant.

Taking the infimum of the latter inequality with respect to η over the interval $\left[\frac{9-\sqrt{17}}{8},\, 1\right]$ yields the geometric convergence (18) with $C = C_0(\lambda_0|Df| + \|f\|_2^2)$. Furthermore, for any $\eta \in \left(\frac{9-\sqrt{17}}{8},\, 1\right]$, it follows from (21) that

$$\sum_{k=0}^{n}\eta^{-k}\Big\|f - \sum_{i=0}^{k} u_i\Big\|_2^2 \le \max(2\lambda_0|Df|, 4\|f\|_2^2))\Psi(\eta),$$

where $\Psi(\eta)$ is the rational function defined by

$$\Psi(\eta) = \frac{\eta^2 + (2\eta - 1)(1 - \eta)^2}{(2\eta - 1)(-4\eta^2 + 9\eta - 4)}.$$

Consequently, the series $\sum\limits_{k=0}^{\infty}\eta^{-k}\|f - \sum_{i=0}^{k} u_i\|_2^2$ converges and the proof is complete. $\qquad\square$

Remark 3.5. Note the gap between the convergence rate in $W^{-1,\infty}(\Omega)$ and the convergence rate in $L^2(\Omega)$. Our argument above points to the fact that one cannot bridge this gap by merely changing the scale refinement strategy. For example, if one adopts the following sequence of scales, $\lambda_k = \lambda_0 r^k$ where $0 < r < 1$, then the convergence rate in $W^{-1,\infty}(\Omega)$ is

$O(r^k)$ while per our proof above the L^2 convergence rate is $O(\eta^k)$ with $\eta > r$. *It would be interesting to characterize the functions $f \in BV(\Omega)$ for which the two convergence rates are identical.*

Example 1: *The set of functions for which the convergence rate is λ_k is nonempty.* Let $\Omega = \mathbb{R}^2$ and for $R > 0$ fixed define the function f by $f(x) = \begin{cases} 1 & |x| \le R, \\ 0 & |x| > R \end{cases}$. We would like to compute TNV dyadic decompositions of f.

Firstly, by a result of Meyer [4, page 36], for any $\lambda > 0$ the optimal ROF decomposition of $f = u_\lambda + v_\lambda$ yields

$$u_\lambda = \max(1 - \lambda/R, 0)f \quad \text{and} \quad v_\lambda = \min(1, \lambda/R)f. \tag{22}$$

Therefore, to obtain a decomposition of f, we have to repeatedly find the ROF decomposition of functions of the form $g = \beta f$ where β is a constant factor. Secondly, for any $\beta > 0$ the ROF decomposition βf at the scale λ is

$$\beta f = \beta u_{\lambda/\beta} + \beta v_{\lambda/\beta} \tag{23}$$

where $f = u_{\lambda/\beta} + v_{\lambda/\beta}$ is ROF decomposition of f at the scale λ/β. Consequently, the TNV sequence for f is completely determine by the ROF decomposition of f at the scales $\{\lambda_k = \lambda_0 2^{-k}\}_{k=0}^\infty$ with the appropriate corrections as indicated by equation (22).

It follows from above that for $\lambda_0 > 0$ fixed, the TNV sequence of f with base scale λ_0 is given by

$$u_k = \begin{cases} 0 & \text{if } k \le \lfloor \log_2(\lambda_0/R) \rfloor \\ (1 - \dfrac{\lambda_k}{R})f & \text{if } k = \lceil \log_2(\lambda_0/R) \rceil \\ \dfrac{\lambda_k}{R} f & \text{if } k > \lceil \log_2(\lambda_0/R) \rceil \end{cases} \tag{24}$$

Furthermore, for any n

$$\Big\| f - \sum_{k=0}^n u_k \Big\|_2 = \Big\| \frac{\lambda_n}{R} f \Big\|_2 = \sqrt{\pi}\lambda_n.$$

Thus the L^2 convergence rate of any TNV dyadic decomposition of the indicator function of a disc centered at the origin is λ_k which is better than the generic geometric rate established in Theorem 3.4.

So far, we have given a convergence rate of a TNV dyadic decomposition which turned

out to be suboptimal in the class of functions with bounded variation, see Example 1. We now investigate the optimal convergence rate and show that in general the convergence rate in $L^2(\Omega)$ does not exceed λ_k.

Proposition 3.6. *Suppose that Ω is convex. Let $f \in BV(\Omega)$ be fixed and $f = \sum\limits_{k=0}^{\infty} u_k$ a hierarchical decomposition of f according the scales $\{\lambda_k \searrow 0 \colon k = 0, 1, 2, \ldots\}$. Then, there exists $C > 0$ dependent on Ω such that*

$$\left\| f - \sum_{j=0}^{k} u_j \right\|_2 \geq C\lambda_k \quad \forall\, k \geq 1. \tag{25}$$

Proof. Let $k \geq 1$ be fixed. Recall from Theorem 2.7 that

$$\lambda_k |Du_k| = 2 \int_{\Omega} u_k (f - \sum_{i=0}^{k} u_i)dx$$

and if Ω is bounded, then by Proposition 2.9

$$\int_{\Omega} u_k dx = 0 \quad \forall\, k \geq 1.$$

Now, by Cauchy-Schwarz inequality and Poincaré-Wirtinger inequality, we infer from above that

$$\lambda_k |Du_k| \leq 2 \left\| u_k \right\|_2 \left\| f - \sum_{i=0}^{k} u_i \right\|_2 \leq K\, |Du_k| \left\| f - \sum_{i=0}^{k} u_i \right\|_2,$$

where $K > 0$ is constant dependent on Ω if Ω is bounded and a universal constant if $\Omega = \mathbb{R}^2$. It then follows that

$$K^{-1}\lambda_k \leq \left\| f - \sum_{i=0}^{k} u_j \right\|_2 \quad \text{for any } k \text{ for which } |Du_k| \neq 0. \tag{26}$$

Finally, since Ω is either convex or $\Omega = \mathbb{R}^2$, it is easy to see using Poincaré - Wirtinger inequality that for any $k \geq 1$ $|Du_k| = 0$ implies that $u_k = 0$. Therefore since the sequence of scales is decreasing, the inequality (26) actually holds for all $k \geq 1$; hence (25) follows with $C = K^{-1}$ and the proof is complete. $\qquad\square$

Remark 3.7. We infer from Theorem 3.4 and Proposition 3.6 that any TNV dyadic decomposition of a function $f \in BV(\Omega)$ converges geometrically in the norm of $L^2(\Omega)$ with rate

$$r \in \left[\frac{1}{2}, \frac{9 - \sqrt{17}}{8} \right].$$

3.2 Convergence in BV

We now investigate the convergence of a hierarchical decomposition in $BV(\Omega)$. We would like to know under what conditions would the decomposition $f = \sum\limits_{k=0}^{\infty} u_k$ converges strongly, strictly, or weakly* in $BV(\Omega)$.

Recall from remark 2.6 that for $f = \sum\limits_{k=0}^{\infty} u_k$ to holds in either of the three topologies, the sequence $\{\sum_{k=0}^{n} u_k : n = 0, 1, 2, \ldots\}$ must be bounded in $BV(\Omega)$. Also, by definition of the functions $\{u_k\}_k$, we have

$$\left\| f - \sum_{i=0}^{k} u_i \right\|_2^2 \leq \lambda_k \left| D(f - \sum_{i=0}^{k-1} u_i) \right|, \quad \forall k = 0, 1, 2, \ldots . \tag{27}$$

Consequently, the convergence of a hierarchical decomposition of f in $BV(\Omega)$ determine its convergence rate in L^2 as follows.

Proposition 3.8. *Suppose $f \in BV(\Omega)$ and its TNV hierarchical decomposition with base scale λ_0 converges in $BV(\Omega)$. Then*

$$\left\| f - \sum_{j=0}^{k} u_j \right\|_2 \leq C \sqrt{\lambda_k} \quad \forall k = 0, 1, 2, \ldots \tag{28}$$

for some constant $C > 0$ independent of k. Furthermore, if $f = \sum\limits_{j=0}^{\infty} u_j$ strongly in BV, then

$$\left\| f - \sum_{j=0}^{k} u_j \right\|_2 = o(\sqrt{\lambda_k}) \quad as \ k \to \infty. \tag{29}$$

Proof. The two inequalities (28) and (29) are mere restatement of the inequality (27) under the boundedness hypothesis and the strong convergence assumption, respectively. $\square$

Remark 3.9. Proposition 3.8 implies that for a TNV hierarchical decomposition to converge in $BV(\Omega)$, its convergence rate in $L^2(\Omega)$ must be at least $O(\sqrt{\lambda_k})$.

We conclude the section with a sufficient condition for weakly* convergence of a decomposition in $BV(\Omega)$.

Proposition 3.10. *Suppose that $f \in BV(\Omega)$ is such that the series $\displaystyle\sum_{k=0}^{\infty} \left\| f - \sum_{i=0}^{k-1} u_i \right\|_2^2 / \lambda_k$*

converges. Then, $f = \displaystyle\sum_{k=0}^{\infty} u_k$ weakly* in BV.

Proof. Fix $k \geq 0$. By the triangle inequality we have

$$\left| D(\sum_{i=0}^{k} u_i) \right| \leq \sum_{i=0}^{k} \left| Du_i \right| \leq \sum_{i=0}^{k} \frac{1}{\lambda_i} E_i(0)$$

$$\leq \sum_{i=0}^{k} \frac{1}{\lambda_i} \left\| f - \sum_{k=0}^{i-1} u_k \right\|_2^2 < \sum_{k=0}^{\infty} \frac{1}{\lambda_k} \left\| f - \sum_{i=0}^{k-1} u_i \right\|_2^2.$$

Since the rightmost term in the latter inequality is finite, it follows that the sequence $\left\{ \left| D(\sum_{i=0}^{n} u_i) \right| \right\}_n$ is bounded. Thus, by Proposition 2.4 $f = \displaystyle\sum_{k=0}^{\infty} u_k$ holds weakly* in $BV(\Omega)$. $\square$

Example 2: *The class of functions f for which the decomposition converges in BV is nonempty.* Continuing Example 1 above, we have for every n

$$\left| D(f - \sum_{k=0}^{n} u_k) \right| = \frac{\lambda_n}{R} |Df| = 2\pi\lambda_n \quad \text{and} \quad \left\| f - \sum_{k=0}^{n} u_k \right\|_2^2 / \lambda_n = \pi\lambda_n.$$

Thus, any dyadic decomposition $f = \displaystyle\sum_{k=0}^{\infty} u_k$ of the indicator function of a disk centered at the origin converges strongly in $BV(\mathbb{R}^2)$ with rate λ_n. Moreover, our hypothesis in Proposition 3.10 holds for indicator functions of disks, since the series $\displaystyle\sum_{k=0}^{\infty} \lambda_k$ converges.

Remark 3.11. The geometric L^2 convergence rate of a dyadic hierarchical decomposition necessary for convergence in $BV(\Omega)$ is better than what we established in Theorem 3.4.

4   Conclusion

In this paper we addressed the convergence of a Tadmor-Nezzar-Vese hierarchical decomposition of a function with bounded variation in $BV(\Omega)$ and the convergence rate in $L^2(\Omega)$. We derived a geometric rate of convergence of such a decomposition in $L^2(\Omega)$, and proved that for a TNV dyadic hierarchical decomposition to converge (strong/strict/weakly*) in $BV(\Omega)$ its geometric convergence rate in $L^2(\Omega)$ must be at least $O((\sqrt{2})^{-n})$. We offered a sufficient condition for weakly* convergence in $BV(\Omega)$ and produced a class of functions in $BV(\mathbb{R}^2)$ on which our condition holds.

Acknowledgements

References

[1] L. Ambrosio, N. Fusco, and D. Pallara. *Functions of bounded variation and free discontinuity problems*. Oxford Mathematical Monographs. The Clarendon Press Oxford University Press, New York, 2000.

[2] A. Chambolle, V. Caselles, D. Cremers, M. Novaga, and T. Pock. An introduction to total variation for image analysis. In *Theoretical foundations and numerical methods for sparse recovery*, volume 9 of *Radon Ser. Comput. Appl. Math.*, pages 263–340. Walter de Gruyter, Berlin, 2010.

[3] Y. Gousseau and J.-M. Morel. Are natural images of bounded variation? *SIAM J. Math. Anal.*, 33(3):634–648 (electronic), 2001.

[4] Y. Meyer. *Oscillating patterns in image processing and nonlinear evolution equations*, volume 22 of *University Lecture Series*. American Mathematical Society, Providence, RI, 2001. The fifteenth Dean Jacqueline B. Lewis memorial lectures.

[5] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1-4):259–268, 1992.

[6] E. Tadmor, S. Nezzar, and L. Vese. A multiscale image representation using hierarchical (BV, L^2) decompositions. *Multiscale Model. Simul.*, 2(4):554–579 (electronic), 2004.