**Eidgenössische Technische Hochschule Zürich**

*Ecole polytechnique fédérale de Zurich*
*Politecnico federale di Zurigo*
*Swiss Federal Institute of Technology Zurich*

# Rate of Convergence of Regularization Procedures and Finite Element Approximations for the Total Variation Flow

X. Feng*, M. von Oehsen and A. Prohl

Research Report No. 2003-12
August 2003

Seminar für Angewandte Mathematik
Eidgenössische Technische Hochschule
CH-8092 Zürich
Switzerland

*Department of Mathematics, University of Tennessee, Knoxville, TN 37996, U.S.A., email:xfeng@math.utk.edu

# Rate of Convergence of Regularization Procedures and Finite Element Approximations for the Total Variation Flow

X. Feng*, M. von Oehsen and A. Prohl

Seminar für Angewandte Mathematik
Eidgenössische Technische Hochschule
CH-8092 Zürich
Switzerland

Research Report No. 2003-12 August 2003

**Abstract**

Following the work of [12], in this paper we continue to carry out mathematical and numerical analysis of the gradient flow for the total variation functional, which has applications to image processing, geometric analysis and materials sciences. The main objectives of the paper are to study the long time behavior of the total variation flow, to analyze rate of convergence for regularization procedures and finite element approximations for the total variation gradient flow with $L^2$ initial data, and to derive explicit scaling laws which relate mesh parameters to the regularization parameter. We also provide numerical experiments which complement our theoretical results. In addition, we present an a priori (model) error estimate for the total variation image denoising model of Rudin-Osher-Fatemi [17] and for other related image denoising models. This (model) error estimate provides a theoretical explanation for the good performance of the total variation model in image denoising.

**Keywords:** Bounded variation, gradient flow, variational inequality, finite elements, a priori error analysis

**Subject Classification:** 35B25, 35K57, 35Q99, 65M60, 65M12, 65N30

*Department of Mathematics, University of Tennessee, Knoxville, TN 37996, U.S.A., email:xfeng@math.utk.edu

## 1. Introduction

This paper is the second in a series (cf. [12]) which devotes to mathematical and numerical analysis of the gradient flow for the total variation functional

$$J_\lambda(u) = |Du|(\Omega) + \frac{\lambda}{2}\int_\Omega (u-g)^2\,dx \tag{1}$$

for a given bounded domain $\Omega \subset \mathbf{R}^d\,(d \geq 1)$, a function $g$ and a nonnegative number $\lambda$. Where $|Du|(\Omega)$ denotes the total variation of the function $u$ defined by (cf. [2])

$$|Du|(\Omega) := \sup\Big\{ \int_\Omega -u \operatorname{div} \mathbf{v}\,dx\,;\ \mathbf{v} \in [C_0^1(\Omega)]^d,\ \|\,\mathbf{v}\,\|_{L^\infty} \leq 1 \Big\},$$

and $BV(\Omega)$ will be used to denote the space of functions of bounded total variation.

Formally, total variation (TV) flow is described by the following initial boundary value problem

$$\frac{\partial u}{\partial t} = \operatorname{div}\Big(\frac{Du}{|\,Du\,|}\Big) - \lambda\big(u-g\big) \qquad \text{in } \Omega_T \equiv \Omega\times(0,T),\quad T>0, \tag{2}$$

$$\frac{\partial u}{\partial n} = 0 \qquad \text{on } \partial\Omega_T \equiv \partial\Omega\times(0,T), \tag{3}$$

$$u(\cdot,0) = u_0(\cdot) \qquad \text{in } \Omega\,, \tag{4}$$

where $Du$, a vector-valued Radon measure, denotes the distributional gradient of $u$.

The best known application of the above gradient flow arises from image processing. The well-known noise removal and image restoration model, which was proposed by Rudin, Osher and Fatemi [17], and analyzed by Acar and Vogel [1], and Chambolle and Lions [10], seeks the minimizer of the functional $J_\lambda$ as the "best" restored image for a given noisy image $g$. Solving the minimization problem by using the popular steepest descent method then motivates to consider the above gradient flow. In such an application, the constant $\lambda$ is known as a Lagrange multiplier which is determined by the original constrained minimization problem (see [10] for a detailed exposition). In addition to applications in image processing, the TV flow also appears in geometric measure theory for studying the evolution of a set with finite perimeter without distortion of the boundary [6] and in materials science for studying the crystalline flow [15].

Although the above TV flow has been addressed and approximated extensively in the literature (see [10, 11, 9] and references therein), its rigorous mathematical analysis (for rough initial data) only has appeared in the literature very recently. The first such work was done by Hardt and Zhou in [14], which studied the gradient flow for a class of linear growth functionals with $L^\infty$ initial values. A comprehensive study for the TV flow for $\lambda = 0$ with $L^1$ initial data was carried out lately by Andreu-Ballester-Caselles-Mazón in [3, 4, 5]. Existence of weak solutions was proved by using Crandall-Liggett's semigroup generation theory [10], on the other hand, uniqueness and stability of entropy solutions were established using Kruzkov's doubling variable technique originally proposed for studying hyperbolic conservation law problems.

Very recently, we developed an $L^2$-variational theory for the TV flow in [12], where a simpler notion of weak solution was introduced, and well-posedness of the problem and regularities of the weak solution were established using *the energy method*. The approach of [12] is based on carefully analyzing the following regularized problem

$$\frac{\partial u^\varepsilon}{\partial t} = \operatorname{div}\Big(\frac{f_\varepsilon'(|\,Du^\varepsilon\,|)Du^\varepsilon}{|\,Du^\varepsilon\,|}\Big) - \lambda(u^\varepsilon - g) \qquad \text{in } \Omega_T\,, \tag{5}$$

$$\frac{\partial u^\varepsilon}{\partial n} = 0 \qquad \text{on } \partial\Omega_T\,, \tag{6}$$

$$u^\varepsilon(\cdot,0) = u_0^\varepsilon(\cdot) \qquad \text{in } \Omega\,, \tag{7}$$

where $f_\varepsilon(z) = \sqrt{z^2+\varepsilon^2}$, and on utilizing well-known results for the prescribed mean curvature flow [13, 16]. This approach not only provides a characterization of the weak solution to (2)–(4), depending on regularity properties of $(u_0, g)$, but also allows to establish convergence of its finite element approximations (cf. [12]).

Recall that (5)–(7) also can be interpreted as the gradient flow for the following regularized functional

$$J_{\lambda,\varepsilon}(u) = \int_\Omega f_\varepsilon(|\,Du\,|)\,dx + \frac{\lambda}{2}\int_\Omega (u-g)^2\,dx\,, \tag{8}$$

where for any $u \in BV(\Omega)$, the first term on the right hand side of (8) is defined as (cf. [2])

$$\int_\Omega \sqrt{|\,Du\,|^2+\varepsilon^2}\,dx := \sup\Big\{ \int_\Omega \big[\, -u\,\mathrm{div}\,\mathbf{v} + \varepsilon\sqrt{1-|\mathbf{v}(x)|^2}\,\big]\,dx\,;\ \mathbf{v} \in [C_0^1(\Omega)]^d,\ \|\,\mathbf{v}\,\|_{L^\infty} \le 1 \Big\}.$$

Based on the work of [12], this paper addresses a number of issues which were not covered in [12]. First, in Section 2 we study the long time behavior of the weak solution to (2)–(4), and prove that as $t \to \infty$ its weak solution indeed converges to the unique minimizer of the total variation functional $J_\lambda$. Corresponding results have been obtained earlier by Hardt and Zhou [14] and Vese [19] for initial data $u_0 \in \mathrm{Dom}(\partial J_{\lambda,\varepsilon})$, $\varepsilon \ge 0$, by nonlinear semigroup theory [7] in combination with an abstract result of Bruck [8]; in contrast, our approach uses a simple energy argument to study asymptotics as $t \to \infty$ for $u_0 \in L^2(\Omega)$. Second, in Section 3 we establish rates of convergence for various regularization procedures, corresponding to various choices of the density function $f_\varepsilon$, for approximating the TV flow. Possible regularizations include those given in (8) and [10]. Recall that in the case $f_\varepsilon(z) = \sqrt{z^2+\varepsilon^2}$, it was proved in [12] that $u^\varepsilon$ converges to $u$ in $L^p$ for $1 \le p < \frac{d}{d-1}$, but no rate of convergence was given. Third, in Section 4 we revisit the finite element method proposed in [12] for approximating (5)–(7) and (2)–(4), and establish optimal rate of convergence for the finite element scheme for approximating the TV flow (2)–(4). Again, we recall from [12] that only convergence in $L^p$ for $1 \le p < \frac{d}{d-1}$ was proved for the finite element scheme for approximating the TV flow, although optimal rate of convergence was proved for the finite element scheme for approximating the regularized problem (5)–(7). Fourth, in Section 5 we provide some numerical experiments, which are especially designed to exemplify our theoretical results for general regularization strategies, and to find explicit scaling laws which relate regularization and discretization parameters in order to obtain optimal rate of convergence. Finally, in Section 6 we propose an a priori (model) error analysis for the TV and other related image denoising models. Specifically, for a given (pure) image function $u$, a blur operator $A$ and a noise $n$ let $g = Au + n$ be the "noisy image" and $\widetilde{u}$ denote a restored image using an image denoising model such as the TV model, we are interested in estimating the error $u - \widetilde{u}$ in terms of quantities $u, \widetilde{u}$, $n$ and parameters used in the image denoising model such as $\lambda$ in the case of TV model, and are able to obtain such an error bound, which depends on $u$ and $\widetilde{u}$ through the energy residual (see Section 6 for details), for the TV model and other related image denoising models. This (model) error analysis provides a theoretical justification for a well observed fact that the TV model is superior to other image denoising models.

## 2. Convergence of the total variation flow as $t \to \infty$

In this section we show that as $t \to \infty$ the solution to the TV flow (2)–(4) with any $L^2(\Omega)$ starting value $u_0$ indeed converges to the minimizer of the total variation functional $J_\lambda$ defined in (1). This result immediately implies the convergence of the gradient descent method for approximating the minimizer of $J_\lambda$. Hence, we provide a theoretical justification for using the TV flow as a means to study and approximate the total variation denoising model.

First, we recall the definition of weak solutions to the TV flow and quote some results of [12], which will serve as the starting point of our analysis in this paper.

**Definition 2.1.** *Let $\Omega \subset \mathbf{R}^d\,(d \geq 2)$ be a bounded open domain with Lipschitz boundary $\partial\Omega$ and $u_0,\ g \in L^2(\Omega)$, a function $u$ is said to be a weak solution to the initial boundary value problem (2)–(4) if $u \in L^1\big((0,T);BV(\Omega)\big) \cap C^0\big([0,T];L^2(\Omega)\big) \cap H^1((0,T);H^{-1}(\Omega))$, $u(x,0) = u_0(x)$ for a.e. $x \in \Omega$, and satisfies for any $s \in [0,T]$*

$$\int_0^s \int_\Omega v_t(v-u)\,dxdt + \int_0^s \big[J_\lambda(v) - J_\lambda(u)\big]\,dt \geq \frac{1}{2}\big[\,\|\,v(s)-u(s)\,\|_{L^2}^2 - \|\,v(0)-u_0\,\|_{L^2}^2\,\big] \tag{9}$$
$$\forall\, v \in L^1\big((0,T);BV(\Omega)\big) \cap L^2(\Omega_T) \text{ such that } v_t \in L^2(\Omega_T).$$

**Theorem 2.1. (cf. Theorems 1.1–1.4 of [12])** *Let $\Omega \subset \mathbf{R}^d$ be a bounded Lipschitz domain, and $u_0,\ g \in L^2(\Omega)$.*

*(i) There exists a unique weak solution to initial boundary value problem (2)–(4). Moreover, suppose $u_i\,(i=1,2)$ are the weak solutions for the given data $u_i(0)$, $g_i\,(i=1,2)$, respectively, then, there holds the following stability estimate*

$$\|\,u_1(s)-u_2(s)\,\|_{L^2} \leq \|\,u_1(0)-u_2(0)\,\|_{L^2} + \sqrt{\lambda}\|\,g_1-g_2\,\|_{L^2} \qquad \forall s \in [0,T]. \tag{10}$$

*(ii) If $u_0 \in BV(\Omega)\cap L^2(\Omega)$, then the weak solution satisfies $u \in L^\infty\big((0,T);BV(\Omega)\big) \cap H^1\big((0,T);L^2(\Omega)\big)$ and the inequality for any $s \in [0,T]$*

$$\int_0^s \int_\Omega u_t(v-u)\,dxdt + \int_0^s \big[J_\lambda(v) - J_\lambda(u)\big]\,dt \geq 0 \quad \forall v \in L^1\big((0,T);BV(\Omega)\big) \cap L^2(\Omega_T). \tag{11}$$

*(iii) If $u_0 \in W^{1,1}(\Omega) \cap H^1_{loc}(\Omega)$, $g \in L^2(\Omega) \cap H^1_{loc}(\Omega)$, $\partial\Omega \in C^2$, then $u \in L^\infty\big((0,T);W^{1,1}(\Omega)\big)$ $\cap L^\infty\big((0,T);H^1_{loc}(\Omega)\big)$.*

*(iv) Suppose that $u_0^\varepsilon \in L^2(\Omega)$ and $g \in L^2(\Omega)$, then for each $\varepsilon > 0$ the regularized problem (5)–(7) has a unique weak solution $u^\varepsilon \in L^1\big((0,T);BV(\Omega)\big) \cap H^1\big((0,T);H^{-1}(\Omega)\big)$ which satisfies that $u^\varepsilon(\cdot,0) = u_0^\varepsilon(\cdot)$, and for any $s \in [0,T]$*

$$\int_0^s \int_\Omega v_t(v-u^\varepsilon)\,dxdt + \int_0^s \big[J_{\lambda,\varepsilon}(v) - J_{\lambda,\varepsilon}(v)\big]\,dt \geq \frac{1}{2}\big[\,\|\,v(s)-u^\varepsilon(s)\,\|_{L^2}^2 - \|\,v(0)-u^\varepsilon(0)\,\|_{L^2}^2\,\big] \tag{12}$$
$$\forall v \in L^1\big((0,T);BV(\Omega)\big) \cap L^2(\Omega_T) \text{ such that } v_t \in L^2(\Omega_T).$$

*In addition, the corresponding stability estimate to (10) holds for the weak solution $u^\varepsilon$. Furthermore, suppose that $u_0^\varepsilon = u^0$, then there also holds*

$$\lim_{\varepsilon\to 0} \|\,u-u^\varepsilon\,\|_{L^1((0,T),L^p(\Omega))} = 0 \qquad \forall p \in \big[1, \frac{d}{d-1}\big)\,. \tag{13}$$

To analyze the long time behavior of the weak solution $u$, a simple but important observation is the following alternative formulation for (9).

**Lemma 2.1.** *Let $u_0 \in L^2(\Omega)$ and $g \in L^2(\Omega)$, then for every $s_0 > 0$ the weak solution $u$ to (2)-(4) satisfies $u \in L^\infty((s_0,T),BV(\Omega)) \cap H^1((s_0,T),L^2(\Omega))$ and*

$$\int_{s_0}^s \int_\Omega u_t(w-u)\,dxdt + \int_{s_0}^s \big[J_\lambda(w) - J_\lambda(u)\big]\,dt \geq 0 \quad \forall w \in L^1\big((0,T);BV(\Omega)\big) \cap L^2(\Omega_T) \tag{14}$$

*for all $s \in [s_0,T]$. Hence, the weak solution $u$ also satisfies*

$$\int_\Omega u_t(t)(w-u(t))\,dx + J_\lambda(w) - J_\lambda(u(t)) \geq 0 \quad \forall w \in BV(\Omega)\cap L^2(\Omega) \text{ and a.e. } t \in (s_0,T). \tag{15}$$

*Proof.* For almost every $s_0 > 0$, notice that $u(s_0) \in BV(\Omega)$. Consider the TV flow which starts at $t = s_0$ with the starting value $u(s_0)$, let $\widehat{u}$ denote the solution. From (ii) of Theorem 2.1 we know that $\widehat{u} \in L^\infty((s_0, T); BV(\Omega)) \cap H^1((s_0, T); L^2(\Omega))$. Define the function

$$\widehat{\widehat{u}}(\cdot, t) = \begin{cases} u(\cdot, t) & \text{for } 0 \le t < s_0, \\ \\ \widehat{u}(\cdot, t) & \text{for } s_0 \le t \le T. \end{cases}$$

It is easy to check that $\widehat{\widehat{u}}$ is a weak solution of (2)–(4) with initial datum $u_0$. By uniqueness of weak solutions (cf. (i) of Theorem 2.1), we have $\widehat{\widehat{u}} = u$ on $[0, T]$. Hence, $u \in L^\infty((s_0, T); BV(\Omega)) \cap H^1((s_0, T); L^2(\Omega))$.

Now, inequality (14) immediately follows from setting

$$v(\cdot, t) = \begin{cases} u(\cdot, t), & \text{for } 0 \le t \le s_0, \\ \\ w(\cdot, t), & \text{for } s_0 < t \le T \end{cases}$$

in (9), for any $w \in L^1\big((0, T); BV(\Omega)\big) \cap L^2(\Omega_T)$. Inequality (15) follows from (14) and the Lebesgue Differentiation Theorem (cf. [2]). The proof is complete. □

*Remark* 2.1. The above lemma can be regarded as a smoothing property of the TV flow, it says that starting with any $L^2$ initial datum at $t = 0$, the solution of the TV flow is a BV function at every later time $t > 0$.

We are now ready to state the first main result of this paper.

**Theorem 2.2.** *Suppose $u_0 \in L^2(\Omega)$ and $g \in L^2(\Omega)$, let $u(x, t)$ denote the weak solution to* (2)-(4) *and $\overline{u} \in BV(\Omega)$ denote the unique minimizer of $J_\lambda$. Then,*

$$\lim_{t \to \infty} \| u(t) - \overline{u} \|_{L^p(\Omega)} = 0 \qquad \forall p \in \big[1, \frac{d}{d-1}\big). \tag{16}$$

*Proof.* Existence and uniqueness of $\overline{u}$ was proved in [10]. Let $s_0 > 0$ be any time incidence such that $u(s_0) \in BV(\Omega)$. Taking $w(t) = u(t - \tau)$ for $\tau > 0$ in (14) with $s = T$, dividing the resulted inequality by $-\tau$ and then setting $\tau \to 0$ yields

$$\int_{s_0}^{T} \| u_t(t) \|_{L^2}^2 \, dt + J_\lambda(u(T)) \le J_\lambda(u(s_0)) < \infty \qquad \text{for any } s_0 \le T < \infty.$$

Hence, there exists a sequence $\{t_j\}$ with $t_j \to \infty$ as $j \to \infty$ such that

$$\lim_{j \to \infty} \| u_t(t_j) \|_{L^2} = 0, \tag{17}$$

$$\| u(t_j) \|_{BV(\Omega) \cap L^2(\Omega)} < C \quad \text{uniformly for } j \ge 1. \tag{18}$$

By compactness of $BV(\Omega)$ (cf. [2]), there exists a subsequence of $\{u(t_j)\}$ (still denoted by the same notation) and $\hat{u} \in BV(\Omega) \cap L^2(\Omega)$ such that $u(t_j)$ converges to $\hat{u}$ weak$*$ in $BV(\Omega)$, strongly in $L^p(\Omega)$ for $1 \le p < \frac{d}{d-1}$, and weakly in $L^2(\Omega)$ as $j \to \infty$.

Finally, setting $j \to \infty$ in (15) after choosing $t = t_j$ and using the fact that $J_\lambda$ is lower semi-continuous with respect to weak$*$ topology in $BV(\Omega)$ (cf. [2]) we get

$$J_\lambda(w) \ge J_\lambda(\hat{u}) \qquad \forall w \in BV(\Omega) \cap L^2(\Omega),$$

which implies that $\hat{u}$ is a minimizer of $J_\lambda$. Then, the uniqueness of the minimizer implies $\hat{u} = \overline{u}$ and that the whole sequence $\{u(\cdot, t)\}$ converges to $\overline{u}$ as $t \to \infty$. The proof is complete. □

*Remark* 2.2. (a) The proof clearly relies on the smoothing property of the TV flow.

(b) It is easy to see that the conclusion of Theorem 2.2 also holds for the solution of the regularized flow (5)–(7) for each $\varepsilon > 0$, that is, $u^\varepsilon(\cdot,t)$ converges to the unique minimizer $\overline{u}^\varepsilon$ of $J_{\lambda,\varepsilon}$ strongly in $L^p(\Omega)$ for $1 \le p < \frac{d}{d-1}$ as $t \to \infty$.

## 3. Rate of convergence of the regularized flow as $\varepsilon \to 0$

From (iv) of Theorem 2.1 we know that the solution of the regularized flow (5)–(7) converges to the solution of the TV flow (2)–(4) strongly in $L^p(\Omega)$ for $1 \le p < \frac{d}{d-1}$ as $\varepsilon \to 0$. However, it does not tell how fast it converges. We will address the issue in this section by establishing a rate of convergence (in powers of $\varepsilon$). Moreover, we will consider more general regularization procedures by stating some structural assumptions for $J_{\lambda,\varepsilon}$, which cover commonly used regularization procedures; in particular, they include $J_{\lambda,\varepsilon}$ defined in (8) and a modified regularization procedure introduced by Chambolle and Lions in [10].

**Theorem 3.1.** *Suppose that $u_0, u_0^\varepsilon \in L^2(\Omega)$ and $g \in L^2(\Omega)$. Let $u, u^\varepsilon$ be the weak solutions of (2)–(4) and (5)–(7), respectively. Assume there exist two positive constants $\alpha, C_0(T)$ such that $J_{\lambda,\varepsilon}$ satisfies*

$$\int_0^T |\, J_{\lambda,\varepsilon}(v) - J_\lambda(v)\,|\, dt \le C_0(T)\varepsilon^\alpha \qquad \forall\, v \in L^1((0,T);BV(\Omega)) \cap L^2(\Omega_T), \tag{19}$$

*then, there holds*

$$\operatorname*{ess\,sup}_{t\in[0,T]} \|\, u(t) - u^\varepsilon(t)\,\|_{L^2(\Omega)} \le \|\, u_0 - u_0^\varepsilon\,\|_{L^2(\Omega)} + 2\sqrt{C_0(T)}\,\varepsilon^{\frac{\alpha}{2}}\,. \tag{20}$$

*Proof.* For any $f \in L^2(\Omega)$ and $\rho \ll 1$, let $(f)_\rho := \mathcal{M}_\rho * f \in C^\infty(\Omega)$ denote its mollification. Here $\mathcal{M}_\rho$ can be chosen as any well-known mollifier.

Let $u_\rho$ and $u_\rho^\varepsilon$ denote the weak solutions of (2)–(4) with initial datum $(u_0)_\rho$ and (5)–(7) with initial datum $(u_0^\varepsilon)_\rho$, respectively, then from Theorem 2.1 we know that $u_\rho$ and $u_\rho^\varepsilon$ satisfies

$$\int_0^s \int_\Omega u_{\rho t}(v - u_\rho)\,dxdt + \int_0^s \big[J_\lambda(v) - J_\lambda(u_\rho)\big]\,dt \ge 0 \quad \forall v \in L^1\big((0,T);BV(\Omega)\big) \cap L^2(\Omega_T), \tag{21}$$

$$\int_0^s \int_\Omega u^\varepsilon_{\rho t}(v - u^\varepsilon_\rho)\,dxdt + \int_0^s \big[J_{\lambda,\varepsilon}(v) - J_{\lambda,\varepsilon}(u^\varepsilon_\rho)\big]\,dt \ge 0 \quad \forall v \in L^1\big((0,T);BV(\Omega)\big), \tag{22}$$

and

$$\|\, u(s) - u_\rho(s)\,\|_{L^2} \le \|\, u_0 - (u_0)_\rho\,\|_{L^2} \qquad \forall s \in [0,T], \tag{23}$$

$$\|\, u^\varepsilon(s) - u^\varepsilon_\rho(s)\,\|_{L^2} \le \|\, u^\varepsilon_0 - (u^\varepsilon_0)_\rho\,\|_{L^2} \qquad \forall s \in [0,T]. \tag{24}$$

Choosing $v = u_\rho^\varepsilon$ in (21), $v = u_\rho$ in (22), and adding the resulting inequalities yield

$$-\int_0^s \int_\Omega (u_\rho - u^\varepsilon_\rho)_t (u_\rho - u^\varepsilon_\rho)\,dx\,dt + \int_0^s \Big\{ \big[J_\lambda(u^\varepsilon_\rho) - J_{\lambda,\varepsilon}(u^\varepsilon_\rho)\big] + \big[J_{\lambda,\varepsilon}(u_\rho) - J_\lambda(u_\rho)\big] \Big\}\,dt \ge 0,$$

which and (19) imply that

$$\begin{aligned} \|\, u_\rho(s) - u^\varepsilon_\rho(s)\,\|^2_{L^2} &\le \|\, u_\rho(0) - u^\varepsilon_\rho(0)\,\|^2_{L^2} + 2\int_0^s \Big\{ \big[J_\lambda(u^\varepsilon_\rho) - J_{\lambda,\varepsilon}(u^\varepsilon_\rho)\big] + \big[J_{\lambda,\varepsilon}(u_\rho) - J_\lambda(u_\rho)\big] \Big\}\,dt \\ &\le \|\, (u_0)_\rho - (u^\varepsilon_0)_\rho\,\|^2_{L^2} + 4C_0(T)\,\varepsilon^\alpha \qquad \forall s \in [0,T]. \end{aligned} \tag{25}$$

Finally, it follows from (23)-(25) and the triangle inequality that

$$\begin{aligned} \|\, u(s) - u^\varepsilon(s)\,\|_{L^2} &\le \|\, u(s) - u_\rho(s)\,\|_{L^2} + \|\, u_\rho(s) - u^\varepsilon_\rho(s)\,\|_{L^2} + \|\, u^\varepsilon_\rho(s) - u^\varepsilon(s)\,\|_{L^2} \\ &\le \|\, u_0 - (u_0)_\rho\,\|_{L^2} + \|\, (u_0)_\rho - (u^\varepsilon_0)_\rho\,\|_{L^2} + 2\sqrt{C_0(T)}\,\varepsilon^{\frac{\alpha}{2}} + \|\, u^\varepsilon_0 - (u^\varepsilon_0)_\rho\,\|_{L^2}. \end{aligned}$$

The desired estimate (20) follows from setting $\rho \to 0$ in the above inequality. The proof is complete. □

For readers' convenience, we now verify the assumption (19) for some commonly used regularization procedures.

**Example 3.1.** *For any $1 < q < \infty$, define*

$$J_{\lambda,\varepsilon,q}(u) := \int_\Omega f_{\varepsilon,q}(|Du|)\,dx + \frac{\lambda}{2}\int_\Omega (u-g)^2\,dx, \tag{26}$$

*where $f_{\varepsilon,q}$ is given by*

$$f'_{\varepsilon,q}(z) = \frac{z}{\sqrt[q]{z^q+\varepsilon^q}}.$$

*Note that $f_{\varepsilon,2}(z) = f_\varepsilon(z) = \sqrt{z^2+\varepsilon^2}$, this case was studied in great details in* [12].

*For any $u \in C^1(\Omega_T)$, a direct calculation yields*

$$\big|J_{\lambda,\varepsilon,2}(u) - J_\lambda(u)\big| = \int_\Omega \Big[\sqrt{|Du|^2+\varepsilon^2} - |Du|\Big]\,dx = \int_\Omega \frac{\varepsilon^2}{\sqrt{|Du|^2+\varepsilon^2}+|Du|} \le |\Omega|\,\varepsilon. \tag{27}$$

*Since $C^\infty(\Omega_T)$ is dense in $L^1((0,T);BV(\Omega))\cap L^2(\Omega_T)$, it follows from above estimate and a standard density argumentation that* (19) *holds with $C_0(T) = |\Omega|T$, and $\alpha = 1$.*

*Remark* 3.1. The above estimate gives the worst scenario. In the case that $|Du|$ has a positive low bound almost everywhere, that is, $\{|Du| \ge c_0\} \approx \Omega$, we have $\alpha = 2$. Hence, we get linear rate of convergence in $\varepsilon$.

**Example 3.2.** *The following regularization procedure is a modification of the one proposed and analyzed by Chambolle and Lions in* [10]

$$J_{\lambda,\varepsilon,CL}(u) := \int_\Omega \phi_\varepsilon(|Du|)\,dx + \frac{\lambda}{2}\int_\Omega (u-g)^2\,dx, \tag{28}$$

*where $\phi_\varepsilon$ is given by*

$$\phi_\varepsilon(z) := \begin{cases} \dfrac{1}{2\varepsilon}z^2 & \text{if } 0 \le z \le \varepsilon, \\ z - \dfrac{\varepsilon}{2} & \text{if } z \ge \varepsilon. \end{cases}$$

*For any $u \in C^1(\Omega_T)$, a direct calculation gives*

$$\begin{aligned} \big|J_{\lambda,\varepsilon,CL}(u) - J_\lambda(u)\big| &= \left|\int_{\{|Du|\le\varepsilon\}} |Du|\Big[\frac{1}{2\varepsilon}|Du| - 1\Big]\,dx - \int_{\{|Du|\ge\varepsilon\}} \frac{\varepsilon}{2}\,dx\right| \\ &\le |\Omega|\,\varepsilon + \frac{|\Omega|}{2}\,\varepsilon \le \frac{3}{2}|\Omega|\,\varepsilon, \end{aligned}$$

*which and a standard density argumentation imply that* (19) *holds with $C_0(T) = \frac{3}{2}|\Omega|T$, and $\alpha = 1$.*

## 4. Rate of convergence of finite element approximations

A fully discrete finite element method for the regularized flow (5)–(7) was proposed in [12], and optimal order error estimates were established. In addition, it was shown that the finite element solution converges to the solution of the TV flow (2)–(4) as the mesh sizes and the parameter $\varepsilon$ all tend to zero. On the other hand, no rate of convergence was given there.

In this section, we first propose and analyze a semi-discrete (in time) scheme to approximate the weak solution of the TV flow (2)–(4). In particular, we verify error estimates for the semi-discrete scheme. Then, we revisit the fully discrete finite element method developed in [12], and establish a rate of convergence for using the method to approximate the TV flow.

4.1. **An implicit time discretization for the TV flow.** Let $\{t_m\}_{m=0}^M$ be an equidistant partition of $[0,T]$ of mesh size $k \in (0,1)$, and $d_t u^m := \frac{1}{k}(u^m - u^{m-1})$. Our semi-discrete in time scheme for approximating the TV flow (2)–(4) is defined as follows: Given $u^0 \in BV(\Omega) \cap L^2(\Omega)$, find $\{u^m\}_{m=1}^M \in BV(\Omega) \cap L^2(\Omega)$ such that

$$\int_\Omega d_t u^m (w - u^m)\, dx + J_\lambda(w) - J_\lambda(u^m) \geq 0 \qquad \forall\, w \in BV(\Omega) \cap L^2(\Omega), \quad m = 1, 2 \cdots, M. \tag{29}$$

Summing over $m$ from 1 to $\ell\,(1 \leq \ell \leq M)$ after taking $w = u^{m-1}$ in (29) leads to the following a priori estimate for $\{u^m\}_{m=1}^M$

$$k \sum_{m=1}^{\ell} \| d_t u^m \|_{L^2}^2 + J_\lambda(u^\ell) \leq J_\lambda(u^0) \qquad \forall \ell \leq M. \tag{30}$$

Well-posedness of problem (29) can be shown by following the proof of (11) in [12]. Convergence behavior depends on regularity of initial data: if $u_0 \in \mathrm{Dom}(\partial J_\lambda)$, optimal order rate of convergence follows from a general result of Rulla [18] for the backward Euler time-discretization of differential inclusions $u_t + \mathcal{A}(u) \ni 0$ with a maximal monotone operator $\mathcal{A} = \partial J_\lambda$. In case of more general initial data $u_0 \in \overline{\mathrm{Dom}(\partial J_\lambda)} \equiv BV(\Omega) \cap L^2(\Omega)$, we prove suboptimal order rate of convergence for the semi-discrete scheme.

**Theorem 4.1.** *(i) Suppose that $g \in L^2(\Omega)$ and $u_0 \in BV(\Omega) \cap L^2(\Omega)$. Let $u$, $\{u^m\}_{m=0}^M$ be the solutions of* (11) *and* (29)*, respectively. Define*

$$\overline{\overline{u}}^k(\cdot, t) := \frac{t - t_{m-1}}{k} u^m(\cdot) + \frac{t_m - t}{k} u^{m-1}(\cdot) \qquad \forall t \in [t_{m-1}, t_m].$$

*Then, there holds*

$$\operatorname*{ess\,sup}_{t \in [0,T]} \| u(t) - \overline{\overline{u}}^k(t) \|_{L^2(\Omega)} \leq \| u_0 - u^0 \|_{L^2(\Omega)} + C\sqrt{k}\, \sqrt{J_\lambda(u_0) + J_\lambda(u^0)}. \tag{31}$$

*(ii) If $u_0 \in Dom(\partial J_\lambda)$, there holds*

$$\operatorname*{ess\,sup}_{t \in [0,T]} \| u(t) - \overline{\overline{u}}^k(t) \|_{L^2(\Omega)} \leq \| u_0 - u^0 \|_{L^2(\Omega)} + Ck \| (\partial J_\lambda)^0(u_0) \|_{L^2}\,, \tag{32}$$

*where $(\partial J_\lambda)^0(u_0)$ is the unique element of minimal norm from the closed and convex set $\partial J_\lambda(u_0)$.*

*Proof.* Since (ii) follows from Theorem 5 of [18], it suffices to prove (i). Notice that (29) can be rewritten as

$$\int_\Omega \overline{\overline{u}}_t^k (w - \overline{u}^k)\, dx + J_\lambda(w) - J_\lambda(\overline{u}^k) \geq 0 \qquad \forall w \in BV(\Omega) \cap L^2(\Omega), \quad t \in (0,T), \tag{33}$$

where $\overline{u}^k(\cdot, t) := u^m(\cdot)$ for $t \in (t_{m-1}, t_m]$. Now, choosing $w = \overline{u}^k$ in (15) with $s_0 = 0$ and $w = u$ in (33) we get

$$\int_\Omega u_t(\overline{u}^k - u)\, dx + J_\lambda(\overline{u}^k) - J_\lambda(u) \geq 0\,,$$
$$\int_\Omega \overline{\overline{u}}_t^k (u - \overline{u}^k)\, dx + J_\lambda(u) - J_\lambda(\overline{u}^k) \geq 0\,.$$

Adding the above two inequalities and using the notation

$$\overline{e}(\cdot, t) := u(\cdot, t) - \overline{u}^k(\cdot, t), \qquad \overline{\overline{e}}(\cdot, t) := u(\cdot, t) - \overline{\overline{u}}^k(\cdot, t) \qquad \forall\, t \in (t_{m-1}, t_m]\,,$$

we get

$$\frac{1}{2} \frac{d}{dt} \| \overline{\overline{e}} \|_{L^2}^2 \leq - \int_\Omega \overline{\overline{e}}_t (\overline{\overline{u}}^k - \overline{u}^k)\, dx\,. \tag{34}$$

It follows from a direct calculation and (30) that for any $1 \le \ell \le M$

$$\int_0^{t_\ell} \| \overline{u}^k - \overline{\overline{u}}^k \|_{L^2}^2 \, dt = \sum_{m=1}^{\ell} \| d_t u^m \|_{L^2}^2 \int_{t_{m-1}}^{t_m} (t - t_{m-1})^2 \, dt = \frac{k^3}{3} \sum_{m=1}^{\ell} \| d_t u^m \|_{L^2}^2 \le C\, k^2,$$

which, together with (30), then leads to

$$\begin{aligned} \Big| \int_0^{t_\ell} \int_\Omega \overline{\overline{e}}_t (\overline{\overline{u}}^k - \overline{u}^k) \, dx dt \Big| &\le \Big\{ \int_0^{t_\ell} \| u_t \|_{L^2}^2 \, dt + k \sum_{m=1}^{\ell} \| d_t u^m \|_{L^2}^2 \Big\}^{\frac{1}{2}} \Big\{ \int_0^{t_\ell} \| \overline{u}^k - \overline{\overline{u}}^k \|_{L^2}^2 \, dt \Big\}^{\frac{1}{2}} \\ &\le C\, k. \end{aligned} \tag{35}$$

Finally, the desired estimate (31) follows from integrating (34) from 0 to $t_\ell$ and appealing to (19) and (35). The proof is complete. □

Efficient control of spatial discretization effects requires additional regularity properties of solutions which is another motivation for us to come back to a fully discrete version of the discretization (5)–(7) in the next subsection.

4.2. **A fully discrete finite element method for the TV flow.** Let $\mathcal{T}_h$ be a quasiuniform triangulation of $\Omega$ with mesh size $h \in (0,1)$, and $V^h$ denote the continuous, piecewise linear finite element space associated with $\mathcal{T}_h$, that is,

$$V^h := \big\{ v_h \in C^0(\overline{\Omega});\ v_h\big|_K \in P_1(K),\ \forall\, K \in \mathcal{T}_h \big\}.$$

We recall that the fully discrete finite element method of [12] for the gradient flow (5)–(7) is defined as follows: Given $U^0 \in V^h$, find $\{U^m\}_{m=1}^M \in V^h$ such that

$$\big(d_t U^m, v_h\big) + \Big( \frac{f_\varepsilon'(|\nabla U^m|)}{|\nabla U^m|} \nabla U^m, \nabla v_h \Big) + \lambda \big( U^m - g, v_h \big) = 0 \qquad \forall\, v_h \in V^h, \tag{36}$$

where $(\cdot,\cdot)$ stands for the standard $L^2(\Omega)$ inner product, and $f_\varepsilon(z) = \sqrt{z^2 + \varepsilon^2}$.

Define the linear interpolation of $\{U^m\}_{m=0}^M$ in time as

$$\overline{\overline{U}}^{\varepsilon,h,k}(\cdot,t) := \frac{t - t_{m-1}}{k} U^m(\cdot) + \frac{t_m - t}{k} U^{m-1}(\cdot) \qquad \forall\, t \in [t_{m-1}, t_m]. \tag{37}$$

If $u_0 \in W^{1,1}(\Omega) \cap H^1_{loc}(\Omega)$ and $g \in L^2(\Omega) \cap H^1_{loc}(\Omega)$, it was shown in Theorem 1.6 of [12] that

$$\lim_{\varepsilon \to 0} \lim_{h,k \to 0} \| u - \overline{\overline{U}}^{\varepsilon,h,k} \|_{L^\infty(0,T);L^p(\Omega))} = 0, \qquad 1 \le p < \frac{d}{d-1}, \tag{38}$$

provided that $\lim_{h \to 0} \| u_0 - U^0 \|_{L^2} = 0$.

We now state a theorem which provides a rate of convergence for the error $u - \overline{\overline{U}}^{\varepsilon,h,k}$.

**Theorem 4.2.** *In addition to the assumptions of Theorem 4.1, suppose that $u_0 \in C^2(\overline{\Omega})$, $g \in L^\infty\big((0,T);\, W^{1,\infty}(\Omega)\big)$, $\partial\Omega \in C^3$, and $f_\varepsilon(z) = \sqrt{z^2 + \varepsilon^2}$. Then, under the following starting value and mesh constraints*

$$\| u_0^\varepsilon - U^0 \|_{L^2} \le Ch^2 \qquad \textit{and} \qquad k = \mathcal{O}(h^2),$$

*there holds the error estimate*

$$\operatorname*{ess\,sup}_{t \in [0,T]} \| u(t) - \overline{\overline{U}}^{\varepsilon,h,k}(t) \|_{L^2(\Omega)} \le \| u_0 - u_0^\varepsilon \|_{L^2} + 2\sqrt{|\Omega| T}\, \sqrt{\varepsilon} + C_1(\varepsilon) k + C_2(\varepsilon) h^2, \tag{39}$$

*where $C_i(\varepsilon)$ for $i = 1, 2$ are positive constants which depends on $\varepsilon^{-1}$ in some low polynomial order.*

*Proof.* (39) follows immediately from the triangle inequality

$$\| u - \overline{\overline{U}}^{\varepsilon,h,k} \|_{L^2} \le \| u - u^\varepsilon \|_{L^2} + \| u^\varepsilon - \overline{\overline{U}}^{\varepsilon,h,k} \|_{L^2},$$

and appealing to (20) and Theorem 1.7 of [12]. The proof is complete. □

*Remark* 4.1. (a). Assumptions in Theorem 4.2 require data $(u_0, g)$ to be regular, and (39) suggests that any (smooth) approximation $(\tilde{u}_0, \tilde{g})$ of $(u_0, g)$ satisfying

$$\| u_0 - \tilde{u}_0 \|_{L^2(\Omega)} + \| g - \tilde{g} \|_{L^2(\Omega)} = \mathcal{O}(\sqrt{\varepsilon})$$

should be sufficient.

(b). The coefficients $C_1(\varepsilon)$ and $C_2(\varepsilon)$ in (39) depend on $\varepsilon^{-1}$ in some low polynomial orders, and no sharp bounds are available from the analysis. In Section 5, we will provide computational evidences for appropriate scalings for $k$ and $h$ with respect to $\varepsilon$ in order to obtain optimal order convergence.

## 5. Numerical experiments

We present some numerical experiments for the regularization procedures (26) and (28), and numerically identify scaling laws between mesh sizes $h, k$ and regularization parameter $\varepsilon$. In all our numerical tests, we solve the nonlinear (algebraic) equations at each time step using a fixed-point iteration, where the diffusivity is evaluated at the previous iterate. This algorithm performs remarkably well, for all regularizations we obtained good approximations after only few iterations. Furthermore, the number of iterations required to trigger the stopping criteria is essentially independent of the regularization procedures. In that sense, we found that all regularization strategies lead to efficient numerical schemes of the same quality.

Our first set of numerical experiments studies the temporal discretization effect and its relationship with the regularization effect. Figure 1 shows the given image $g$, and two smoothened samples using the TV flow with $\lambda = 0$. Figure 2 gives $L^2$-errors at $t = 5 \cdot 10^{-3}$ for different scalings $\varepsilon = \mathcal{O}(k^r)$ for $r = 0, 1/2, 1, 3/2$, which validates a proper scaling law $\varepsilon = \mathcal{O}(k)$, for $f_\varepsilon = f_{\varepsilon,2}$; this scaling law is better compared to the theoretically predicted one in Theorem 4.2, which may be explained by the 'pessimistic' crucial last bound in (27) in situations where $\{|\, Du \,| > c\}$ almost covers $\Omega$. Figure 3 displays corresponding results for the regularizations $f_{\varepsilon,q}$, for $q = 1.5$ and 5; Analogous experiments for the case $f_{\varepsilon,\mathrm{CL}}$-regularization are shown in Figure 5, which again evidence an optimal scaling $\varepsilon = \mathcal{O}(k)$.

Spatial discretization effect in the case $f_{\varepsilon,2}$-regularization is reported in Figure 5 using initial data of different regularities (see Figure 4). We observe a decrease in rate of convergence when using rough initial data (cf. Theorem 4.2).

Finally, we examine the regularization procedures using $f_{\varepsilon,\mathrm{CL}}$ and $f_{\varepsilon,q}$ for $q = 1, 2, 100$. Figure 6 plots the diffusivity function $f'_{\varepsilon,*}(s)/s$ of each regularization. We remark that the regularization procedure of Chambolle and Lions [10] amounts to applying the heat equation (constant diffusivity) in regions of small gradients and using the TV flow at places where the gradient is big. For large $q$, the $f_{\varepsilon,q}$-regularization has the similar properties. However, for small $q$ the diffusivity of $f_{\varepsilon,q}$ is large for small gradients and small for large gradients where the edges reside. For two-dimensional image denoising applications, our numerical tests indicate that all these regularization strategies essentially perform equally well. Figures 7 and 8 depict three samples of the regularized flow with $f_{\varepsilon,2}$. Corresponding tests with the Chambolle-Lions regularization $f_{\varepsilon,\mathrm{CL}}$ and the regularizations $f_{\varepsilon,q}$ for $q \neq 2$ produce results which do not differ qualitatively from those of Figures 7 and 8.

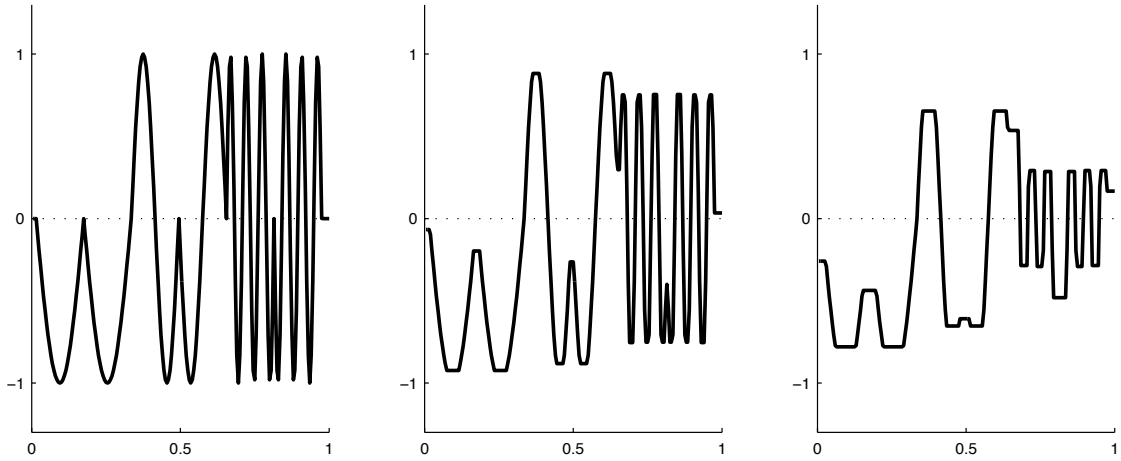


FIGURE 1. Computed solutions of the regularized flow at $t = 0, 10^{-3}, 5 \cdot 10^{-3}$ using $f_{\varepsilon,2}$-regularization with $(\varepsilon, k, h) = (10^{-10}, 5 \cdot 10^{-8}, 10^{-3})$ and $\lambda = 0$.

## 6. Error estimate for the TV and other image denoising models

In this final section, we derive an a priori model error estimate for the total variation image denoising model of Rudin-Osher-Fatemi [17] and for other related image denoising models. Specifically, given a (pure) image $u : \Omega \to \mathbf{R}$, a (linear and injective) blur operator $A$, and some noise $n : \Omega \to \mathbf{R}$. Let $g = Au + n$ be the given "noisy image", and $\widetilde{u} : \Omega \to \mathbf{R}$ denotes a recovered image using an image denoising model such as the TV model. We now ask for an estimate of the model error $u - \widetilde{u}$ in terms of the known quantities $u, \widetilde{u}$, $n$ and parameters used in the image denoising model such as $\lambda$ in the case of TV model. We will show that the error depends on $u$ and $\widetilde{u}$ through the energy residual for each image denoising models. This model error estimate provides a theoretical explanation for the good performance of the TV model in image denoising.

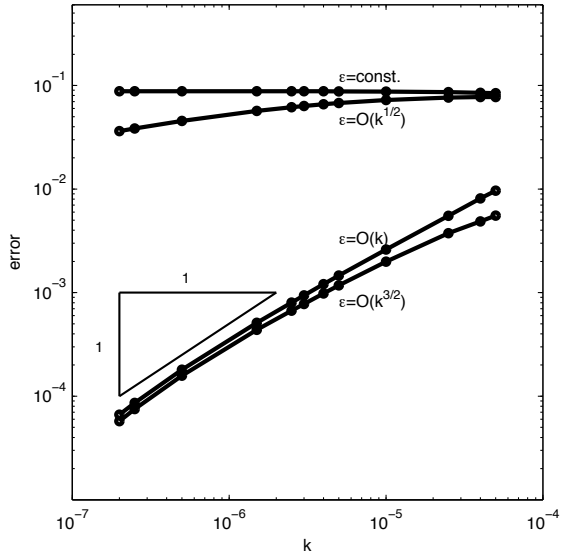


FIGURE 2. $L^2$-error at $t = 5 \cdot 10^{-3}$ using $f_{\varepsilon,2}$-regularization with different scaling laws $\varepsilon = \mathcal{O}(k^r)$.

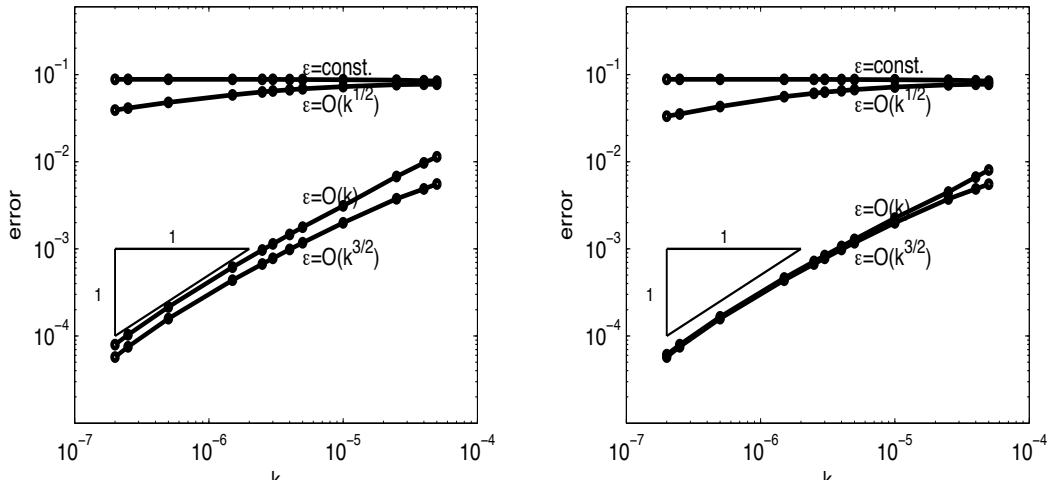


FIGURE 3. $L^2$-error at $t = 5 \cdot 10^{-3}$ using $f_{\varepsilon,q}$-regularization with different scaling laws $\varepsilon = \mathcal{O}(k^r)$. $q = 1.5$ (left), $q = 5$ (right).

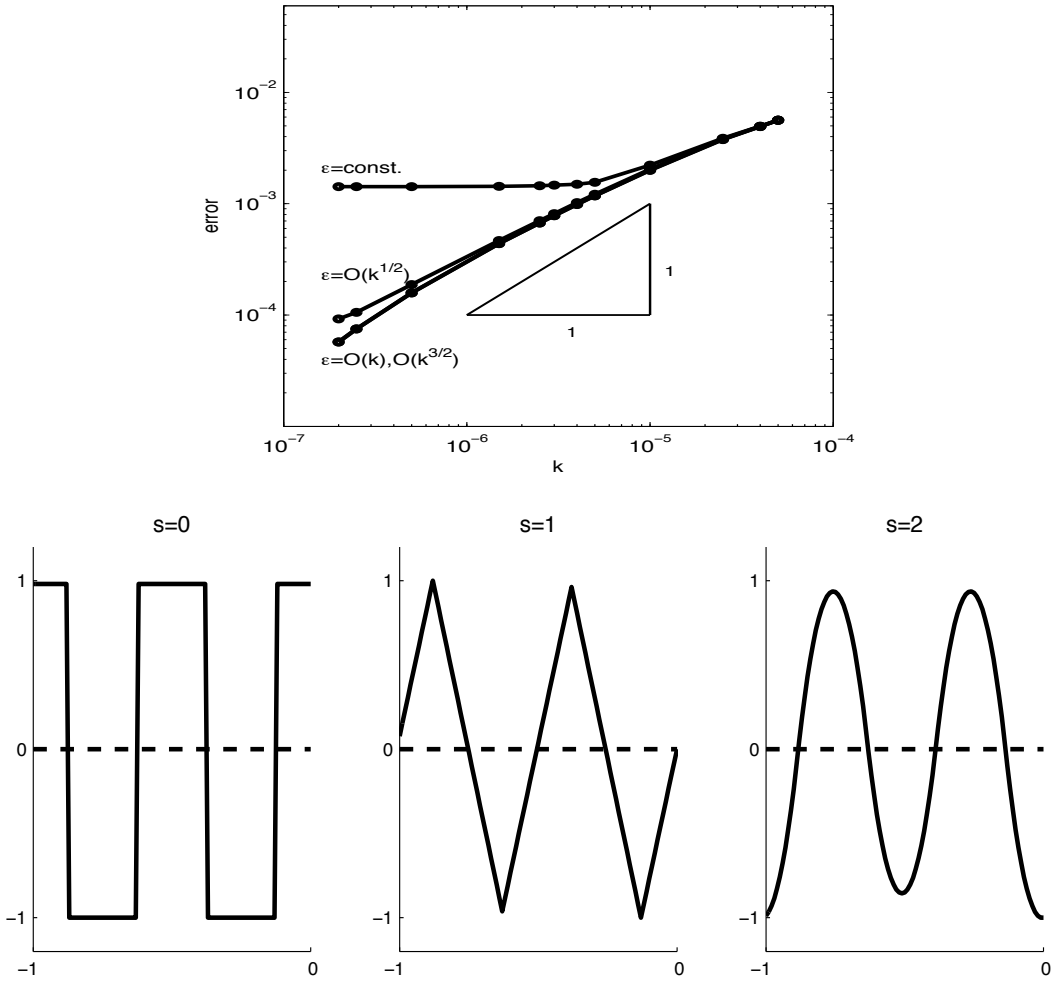


FIGURE 4. Piecewise polynomial initial data with discontinuities in the $s$-th derivative for $s = 0, 1, 2$.

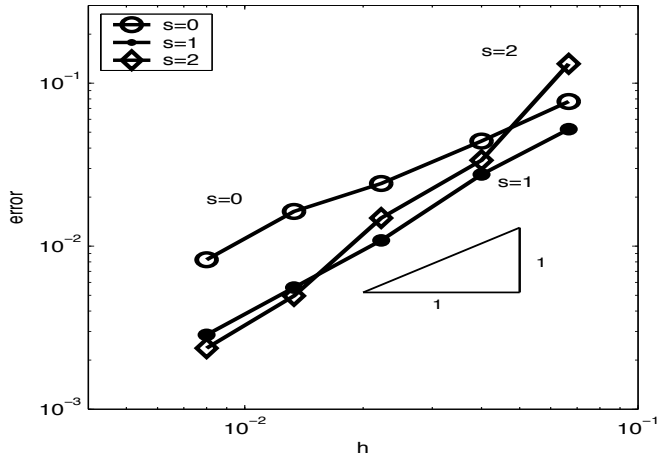


FIGURE 5. $L^2$-error at $t = 10^{-3}$ using $f_{\varepsilon,2}$-regularization. $\varepsilon = 10^{-3}$, $k = 10^{-6}$, and the initial data of Figure 4 are used in the tests.

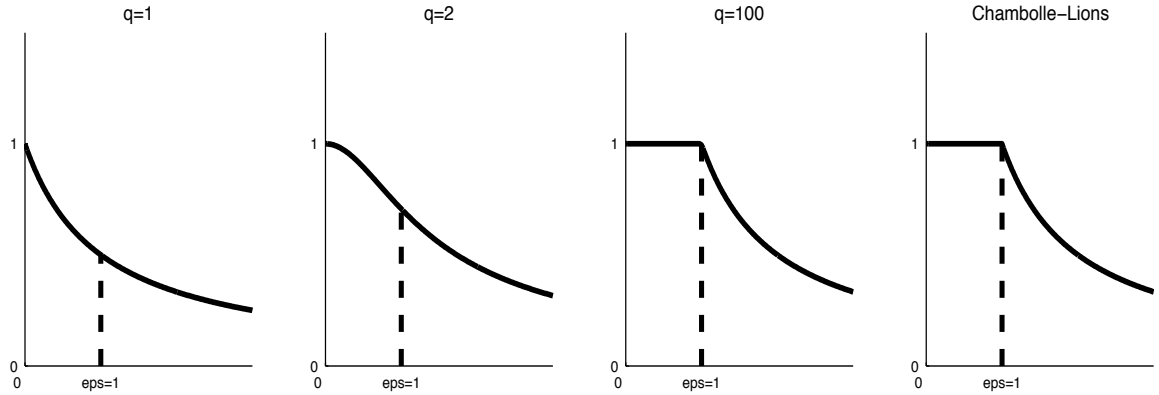


FIGURE 6. The diffusivity for $f_{\varepsilon,q}$, $q = 1, 2, 100$ and for $f_{\varepsilon,\mathrm{CL}}$.

FIGURE 7. The initial noisy image and two smoothed samples at times $t = 5 \cdot 10^{-4}, 10^{-3}$ using $f_{\varepsilon,2}$-regularization. $\varepsilon = 10^{-5}, k = 10^{-5}, h = 2^{-8}$ and $\lambda = 0$ are used in the test.

FIGURE 8. The initial noisy image and two smoothed samples at times $t = 5 \cdot 10^{-4}, 10^{-3}$ using $f_{\varepsilon,2}$-regularization. $\varepsilon = 10^{-5}, k = 10^{-5}, h = 2^{-8}$ and $\lambda = 0$ are used in the test.

Let us consider a class of image selection functionals of the form

$$E_\lambda(v) := E(v) + \frac{\lambda}{2} \| Av - g \|_{L^2}^2 \qquad \forall v \in V \equiv BV(\Omega) \cap L^2(\Omega), \tag{40}$$

$$E(v) := \int_\Omega \varphi(|Dv|)\, dx, \tag{41}$$

where $\lambda \geq 0$ is a given constant, and $\varphi(z)$ is a Lipschitz continuous function of $z \in \mathbf{R}$. Then the recovered image $\widetilde{u}$ is defined by

$$\widetilde{u} := \operatorname*{argmin}_{v \in V} E_\lambda(v). \tag{42}$$

*Remark* 6.1. Recall that the total variation image denoising model of Rudin-Osher-Fatemi [17] uses $\varphi(z) = z$ and the harmonic image denoising model has $\varphi(z) = z^2$.

We are now ready to state the main result of this section.

**Theorem 6.1.** *For a given image function $u$, let $\widetilde{u}$ be defined by* (42). *Then, there holds the following error estimate*

$$\| u - \widetilde{u} \|_A^2 \leq \| g - A\widetilde{u} \|_{L^2}^2 + \frac{2}{\lambda} \big[ E_\lambda(u) - E_\lambda(\widetilde{u}) \big] - \frac{2}{\lambda} Q(u, \widetilde{u}), \tag{43}$$

*where $\| v \|_A := \| Av \|_{L^2}$, and $Q(u, \widetilde{u})$ is defined by*

$$Q(u, \widetilde{u}) := E(u) - E(\widetilde{u}) - E'(\widetilde{u})(u - \widetilde{u}), \tag{44}$$

*where $E'(\widetilde{u})$ denotes the Fréchet derivative of $E$ at $\widetilde{u}$.*

*Proof.* It is easy to check that the Euler-Lagrange equation for $\widetilde{u}$ is given by

$$-\operatorname{div}\left( \frac{\varphi'(|D\widetilde{u}|)}{|D\widetilde{u}|}\, D\widetilde{u} \right) + \lambda A^*(A\widetilde{u} - g) = 0,$$

where $A^*$ stands for the adjoint operator of $A$ with respect to $L^2(\Omega)$ inner product. Rewrite the above equation as

$$A^*A(\widetilde{u} - u) = A^*(g - Au) + \frac{1}{\lambda} \operatorname{div}\left( \frac{\varphi'(|D\widetilde{u}|)}{|D\widetilde{u}|}\, D\widetilde{u} \right),$$

then test is against $\widetilde{u} - u$ and use the relation (44) to get

$$\begin{aligned} \| \widetilde{u} - u \|_A^2 &= \big(g - Au, A(\widetilde{u} - u)\big) + \frac{1}{\lambda} \left( \frac{\varphi'(|D\widetilde{u}|)}{|D\widetilde{u}|} D\widetilde{u}\,, D(u - \widetilde{u}) \right) \\ &\le \frac{1}{2} \| \widetilde{u} - u \|_A^2 + \frac{1}{2} \| g - Au \|_{L^2}^2 + \frac{1}{\lambda} \big[ E(u) - E(\widetilde{u}) - Q(u, \widetilde{u}) \big] \\ &= \frac{1}{2} \| \widetilde{u} - u \|_A^2 + \frac{1}{\lambda} \big[ E_\lambda(u) - E(\widetilde{u}) - Q(u, \widetilde{u}) \big]. \end{aligned}$$

Hence,

$$\| \widetilde{u} - u \|_A^2 \le \| g - A\widetilde{u} \|_{L^2}^2 + \frac{2}{\lambda} \big[ E_\lambda(u) - E_\lambda(\widetilde{u}) \big] - \frac{2}{\lambda} Q(u, \widetilde{u}). \tag{45}$$

The proof is complete. □

*Remark* 6.2. (a). The three terms on the right hand side of (45) clearly show sources of the total error in the recovered image $\widetilde{u}$. The first term can be regarded as the fidelity error for $\widetilde{u}$. The second term, which is positive, measures how far the energy $E_\lambda(u)$ from the minimum energy $E_\lambda(\widetilde{u})$, hence, it can be regarded as energy residual error. The third term, which does not have a fixed sign in general, adds a contribution to the global error if $Q(u, \widetilde{u})$ is negative (which is the case for nonconvex density function $\varphi$), but reduces the global error $Q(u, \widetilde{u})$ if it is nonnegative (which occurs when $\varphi$ is convex).

(b). Although both TV model and harmonic model have convex density function $\varphi$, which results in nonnegative $Q(u, \widetilde{u})$, harmonic model gives much larger energy residual error in general.

**Acknowledgment:** The first author would like to thank Professor Hongkai Zhao of UC Irvine for raising the question of model error analysis addressed in Section 6 and also for his stimulating discussion. X.F. and A.P. would like to thank the Mathematisches Forschungsinstitut Oberwolfach for the kind hospitality and opportunity of its "Research in Pairs" program.

## References

[1] R. Acar, C.R. Vogel, *Analysis of bounded penalty methods for ill-posed problems*, Inv. Problems **10**, pp. 1217–1229 (1994).

[2] L. Ambrosio, N. Fusco, D. Pallara, *Functions of bounded variation and free discontinuity problems*, Oxford University Press, New York (2000).

[3] F. Andreu, C. Ballester, V. Caselles, J.M. Mazon, *The Dirichlet problem for the total variation flow*, J. Funct. Anal. **180**, pp. 347–403 (2001).

[4] F. Andreu, C. Ballester, V. Caselles, J.M. Mazon, *Minimizing total variation flow*, Diff. Int. Equations **14**, pp. 321–360 (2001).

[5] F. Andreu, C. Ballester, V. Caselles, J.M. Mazon, *Some qualitative properties for the total variation flow*, J. Funct. Anal. **188**, pp. 516–547 (2002).

[6] G. Bellettini, V. Caselles, *The total variation flow in* $\mathbf{R}^N$, J. Diff. Equations (accepted) (2001).

[7] H. Brezis, *Operateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*, North-Holland (1973)

[8] R.E. Bruck, *Asymptotic convergence of nonlinear contraction semi-groups in Hilbert spaces*, J. Funct. Anal. **18**, pp. 15-26 (1975).

[9] T. Chan and J. Shen, *On the role of the BV image model in image restoration*, Tech. Report CAM 02-14, Department of Mathematics, UCLA (2002).

[10] A. Chambolle, P.L. Lions, *Image recovery via total variation minimization and related problems*, Numer. Math. **76**, pp. 167–188 (1997).

[11] D. C. Dobson and C. R. Vogel, *Convergence of an iterative method for total variation denoising*, SIAM J. Numer. Anal. **34**, pp. 1779–1791 (1997).

[12] X. Feng, A. Prohl, *Analysis of total variation flow and its finite element approximations*, Math. Mod. Num. Analysis **37**, pp. 533–556 (2003).

[13] C. Gerhardt, *Evolutionary surfaces of prescribed mean curvature*, J. Diff. Equations **36**, pp. 139–172 (1980).

[14] R. Hardt, X. Zhou, *An evolution problem for linear growth functionals*, Comm. PDEs **19**, pp. 1879–1907 (1994).
[15] R. Kobayashi, J. A. Warren and W. C. Carter, *A simple continuum model of grain boundaries*, Physica D, **140**, pp. 141–150 (2000).
[16] A. Lichnewsky, R. Temam, *Pseudosolutions of the time-dependent minimal surface problem*, J. Diff. Equations **30**, pp. 340–364 (1978).
[17] L. Rudin, S. Osher, E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Physica D **60**, pp. 259–268 (1992).
[18] J. Rulla, *Error analysis for implicit approximations to solutions to Cauchy problems*, SIAM J. Num. Analysis **33**, pp. 68–87 (1996).
[19] L. Vese, *A study in the BV Space of a denoising-deblurring variational problem*, Appl. Math. & Opt. **44**, pp. 131–161 (2001).

# Research Reports

| No. | Authors | Title |
|---|---|---|
| 03-12 | X. Feng, M. von Oehsen, A. Prohl | Rate of Convergence of Regularization Procedures and Finite Element Approximations for the Total Variation Flow |
| 03-11 | P.-A. Nitsche | Best $N$ Term Approximation Spaces for Sparse Grids |
| 03-10 | J.T. Becerra Sagredo | Z-splines: Moment Conserving Cardinal Spline Interpolation of Compact Support for Arbitrarily Spaced Data |
| 03-09 | P. Houston, D. Schötzau, Th. Wihler | Energy norm a-posteriori error estimation for mixed discontinuous Galerkin approximations of the Stokes problem |
| 03-08 | R. Hiptmair, A. Buffa | Coercive combined field integral equations |
| 03-07 | R. Hiptmair, O. Sterz | Current and Voltage Excitations for the Eddy Current Model |
| 03-06 | A.-M. Matache, P.-A. Nitsche, C. Schwab | Wavelet Galerkin Pricing of American Options on Lévy Driven Assets |
| 03-05 | M. Becheanu, R.A. Todor | On the Set of Diameters of Finite Point-Sets in the Plane |
| 03-04 | C. Schwab, R.A. Todor | Sparse finite elements for stochastic elliptic problems - higher order moments |
| 03-03 | R. Sperb | Bounds for the first eigenvalue of the elastically supported membrane on convex domains |
| 03-02 | F.M. Buchmann | Computing exit times with the Euler scheme |
| 03-01 | A. Toselli, X. Vasseur | Domain decomposition preconditioners of Neumann-Neumann type for $hp$-approximations on boundary layer meshes in three dimensions |
| 02-26 | M. Savelieva | Theoretical study of axisymmetrical triple flame |
| 02-25 | D. Schötzau, C. Schwab, A. Toselli | Mixed $hp$-DGFEM for incompressible flows III: Pressure stabilization |
| 02-24 | F.M. Buchmann, W.P. Petersen | A stochastically generated preconditioner for stable matrices |
| 02-23 | A.W. Rüegg, A. Schneebeli, R. Lauper | Generalized $hp$-FEM for Lattice Structures |
| 02-22 | L. Filippini, A. Toselli | $hp$ Finite Element Approximations on Non-Matching Grids for the Stokes Problem |
| 02-21 | D. Schötzau, C. Schwab, A. Toselli | Mixed $hp$-DGFEM for incompressible flows II: Geometric edge meshes |