



On the Taut String Interpretation and Other Properties of the Rudin–Osher–Fatemi Model in One Dimension

Niels Chr Overgaard¹

Received: 14 December 2018 / Accepted: 9 September 2019 / Published online: 20 September 2019
© The Author(s) 2019

Abstract

We study the one-dimensional version of the Rudin–Osher–Fatemi (ROF) denoising model and some related TV-minimization problems. A new proof of the equivalence between the ROF model and the so-called taut string algorithm is presented, and a fundamental estimate on the denoised signal in terms of the corrupted signal is derived. Based on duality and the projection theorem in Hilbert space, the proof of the taut string interpretation is strictly elementary with the existence and uniqueness of solutions (in the continuous setting) to both models following as by-products. The standard convergence properties of the denoised signal, as the regularizing parameter tends to zero, are recalled and efficient proofs provided. The taut string interpretation plays an essential role in the proof of the fundamental estimate. This estimate implies, among other things, the strong convergence (in the space of functions of bounded variation) of the denoised signal to the corrupted signal as the regularization parameter vanishes. It can also be used to prove semi-group properties of the denoising model. Finally, it is indicated how the methods developed can be applied to related problems such as the fused lasso model, isotonic regression and signal restoration with higher-order total variation regularization.

Keywords Total variation minimization · Taut string · Regression splines · Lewy–Stampacchia inequality · Denoising semi-group · Fused Lasso · Isotonic regression · Higher-order total variation

1 Introduction

More than 25 years have passed since Leonid Rudin, Stanley Osher and Emad Fatemi proposed their classical model for edge-preserving denoising of images [37]. Given an image corrupted with zero-mean Gaussian noise. If this image is modelled as a function $f : \Omega \rightarrow \mathbf{R}$, where Ω is a bounded, piecewise smooth domain in the plane, then the original Rudin–Osher–Fatemi (ROF) model proposes to recover the clean image as the function $u_0 : \Omega \rightarrow \mathbf{R}$ which solves the following constrained minimization problem:

$$\min \int_{\Omega} |\nabla u(x)| \, dx \quad \text{subject to} \\ \int_{\Omega} u \, dx = \int_{\Omega} f \, dx \quad \text{and} \quad \int_{\Omega} (u - f)^2 \, dx = \sigma^2.$$

Here, the expression $\int_{\Omega} |\nabla u(x)| \, dx$ represents the total variation in the function u . (The set of functions for which the total variation is finite is denoted $BV(\Omega)$.) The linear constraint models that the noise has zero mean, and the quadratic constraint, that it has variance σ^2 . In practice, one studies the Lagrange formulation of this problem and minimizes the functional

$$\lambda \int_{\Omega} |\nabla u| \, dx + \frac{1}{2} \int_{\Omega} (u - f)^2 \, dx$$

over functions $u \in BV(\Omega)$. The linear constraint is then automatically satisfied (see Sect. 4), and it was shown early on, by Chambolle and Lions [19], that there exists a positive value for the Lagrange multiplier λ such that the quadratic constraint is satisfied as well. (It is convenient to place the Lagrange multiplier in front of the objective function instead of in front of the constraint.)

The present paper will carry out a thorough analysis of the one-dimensional ROF model: we take Ω to be a bounded interval $I = (a, b)$ and let $f \in L^2(I)$ denote a given (noisy) signal. To this signal, we associate the (ROF) functional

✉ Niels Chr Overgaard
nco@maths.lth.se

¹ Centre for Mathematical Sciences, Lund University, Lund, Sweden

$$E_\lambda(u) = \lambda \int_a^b |u'(x)| dx + \frac{1}{2} \int_a^b (f(x) - u(x))^2 dx ,$$

where $\lambda > 0$ is a parameter, and define the denoised signal as the function $u_\lambda \in \text{BV}(I)$ which minimizes this energy, i.e.

$$u_\lambda := \arg \min_{u \in \text{BV}(I)} E_\lambda(u) . \tag{1}$$

Precise definitions of the total variation and the space $\text{BV}(I)$ will be given in Sects. 2 and 3.

We are going to compare the one-dimensional ROF model to the *taut string algorithm*, which is an alternative method for denoising of signals with applications in statistics, non-parametric estimation, real-time communication systems and stochastic analysis. In the continuous setting, for analogue signals, the taut string algorithm can be stated in the following manner (cf. Fig. 1):

The Taut String Algorithm

INPUT: A bounded interval $I = (a, b)$, a (noisy) signal $f \in L^2(I)$ and a parameter $\lambda > 0$.

OUTPUT: The denoised signal $f_\lambda \in L^2(I)$.

STEP 1. Compute the cumulative signal,

$$F(x) = \int_a^x f(t) dt , \quad x \in \bar{I} = [a, b].$$

STEP 2. Set

$$T_\lambda = \left\{ W \in H^1(I) : W(a) = F(a), W(b) = F(b), \right. \\ \left. \text{and } F - \lambda \leq W \leq F + \lambda \right\} .$$

(The set of L^2 -functions with weak derivatives in L^2 and graphs lying within a tube around F of width λ .)

STEP 3. Compute the unique minimizer $W_\lambda \in T_\lambda$ of the shortest path problem

$$\min_{W \in T_\lambda} L(W) := \int_a^b \sqrt{1 + W'(x)^2} dx \tag{2}$$

(W_λ is the ‘Taut string’ in the name of the algorithm.)

STEP 4. Set $f_\lambda = W'_\lambda$ (distributional derivative.)

END.

The taut string algorithm has been extensively studied in the discrete setting by Mammen and de Geer as well as Davies and Kovacs [20,29] and Dumgen and Kovacs [21]. Recently, using the highly developed methods of real interpolation

theory (Peetre’s K -functional and the notion of invariant K -minimal sets, etc.), Niyobuhingiro [31] has investigated the ROF model in the discrete case and Setterqvist [40] has probed the limits to which taut string methods may be extended.

The taut string algorithm instructs us to minimize the curve length functional $L(W)$ in (2) among all functions W whose graphs are curves through the points $(a, F(a))$ and $(b, F(b))$ and which lies within the tube T_λ . The name of the algorithm derives from this shortest path problem. It turns out that one may just as well minimize the energy associated with an elastic rubber band satisfying the same boundary conditions and the same constraints:

$$\min_{W \in T_\lambda} E(W) := \frac{1}{2} \int_a^b W'(x)^2 dx , \tag{3}$$

That the new problem (3) has the same solution (W_λ) as (2) is the content of the following interesting lemma:

Lemma 1 *Let $H : \mathbf{R} \rightarrow \mathbf{R}$ be a convex C^1 -function and set*

$$L_H(W) = \int_I H(W'(x)) dx .$$

If $W_ := \arg \min_{W \in T_\lambda} E(W)$, then W_* is also a solution to the minimization problem $\min_{W \in T_\lambda} L_H(W)$. Moreover, if H is strictly convex then W_* is the unique minimizer in T_λ of L_H .*

If we take $H(s) = (1 + s^2)^{1/2}$, it follows that (2) and (3) have precisely the same minimizer in T_λ , namely W_λ . While this statement seems intuitively clear from our everyday experience with rubber bands and strings, the mathematical assertion is not equally self-evident. A proof is consequently offered in ‘‘Appendix A’’.

The paper, which is a considerably modified and enlarged version of the author’s preprint [32] and subsequent conference paper [33], has two main purposes. The first is to present a new, elementary proof of the following remarkable result:

Theorem 1 *The taut string algorithm and the ROF model yield the same solution; $f_\lambda = u_\lambda$.*

This is not new—a discrete version of this theorem was proved in [20,29]. In the continuous setting, the equivalence result was explicitly stated and proved by Grassmair [24]. There is also an extensive treatment in [39, Ch. 4]. Indeed, a few years earlier, Hintermüller and Kunisch [26, p.7], in a brief (but inconclusive) remark, refer to the close relation between the one-dimensional ROF model and the taut string algorithm. The new proof is given in Sect. 5 after the basic existence theory for the ROF model has been developed.

The second main purpose of the paper is to prove the following ‘‘fundamental’’ estimate on the denoised signal (see Sect. 8):

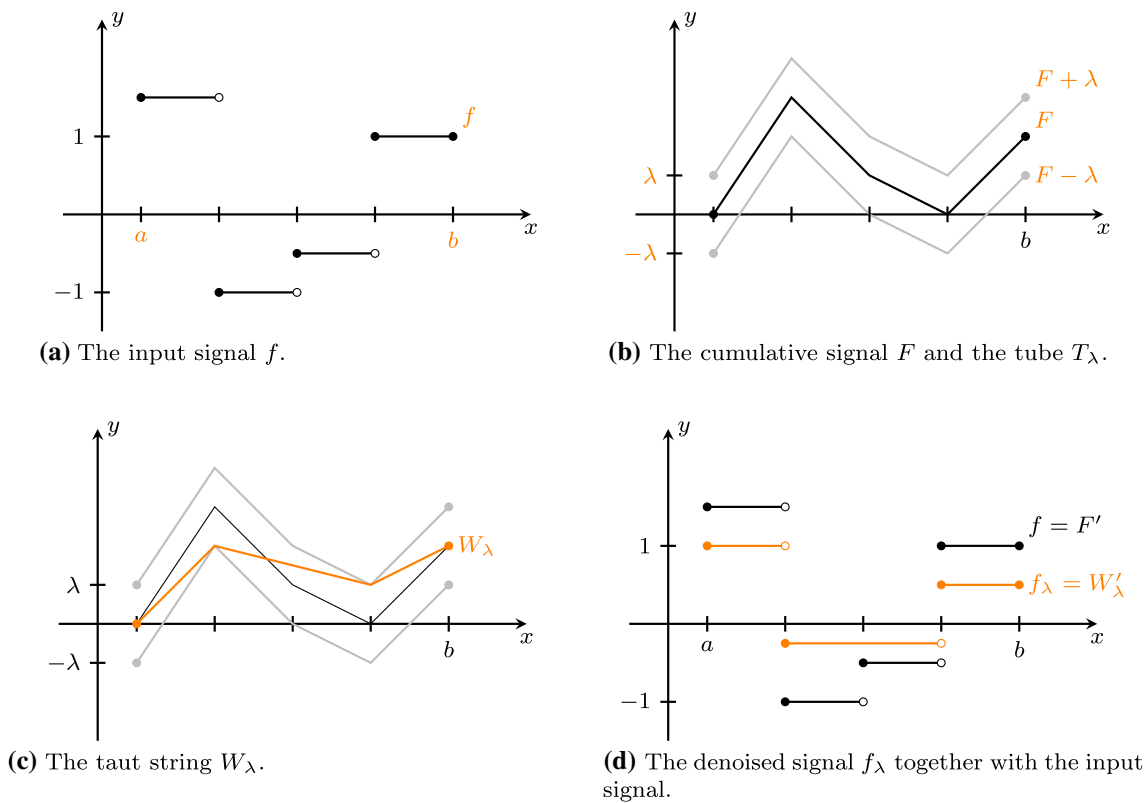


Fig. 1 Graphical illustrations of the steps in the taut string algorithm applied to a piecewise constant signal

Theorem 2 *If the signal f belongs to $BV(I)$, then for any $\lambda > 0$, the denoised signal u_λ satisfies the inequality*

$$-(f')^- \leq u'_\lambda \leq (f')^+, \tag{4}$$

where $(f')^+$ and $(f')^-$ denote the positive and the negative variations, respectively, of f' (distributional derivative).

Here, f' , as well as the derivative u'_λ , is computed in the distributional sense and is, in general, a signed measure. Recall that $(f')^+$ and $(f')^-$ are finite positive measures satisfying $f' = (f')^+ - (f')^-$, see, e.g. [38, Sect. 6.6]. As an example, the reader may compute the derivatives of f and $u_\lambda = f_\lambda$ shown in Fig. 1d. Theorem 2 immediately implies the “edge-preserving” property of the one-dimensional ROF model: the solution u_λ can only have a “jump” at points where the data f has a jump, a qualitative result which holds even in the multidimensional case, as proved by Caselles et al. [16]. In the one-dimensional case, we obtain, in addition, a quantification of these jumps; they have the same sign as and are dominated by the jumps in the data. This assertion is embodied in the estimate (4) which does not carry over to higher dimensions. The proof of Theorem 2 is based on (an extension to bilateral obstacle problems of) the classic Lewy–Stampacchia inequality [28] and uses the taut string

interpretation (Theorem 1) in an essential way. The proof of Theorem 2 and its implications are given in Sect. 8.

Our Theorem 2 turns out to be a special case of an estimate proved in Briani et al. [14, Lemmas 3, 4] and is related to a result (Lemmas 2, 3) in Bonfore and Figalli [11, p. 4459]. Both papers study the gradient flows associated with certain one-homogeneous functionals. In [14], it is functionals of the form $\int_\Omega |\operatorname{div} \mathbf{w}| dx$ for vector fields \mathbf{w} defined on Ω , and the paper was not directly concerned with the ROF model. The relevance of these papers was first pointed out the author after the publication of [33]. Moreover, our method of proof differs from those in [11, 14] and we use Theorem 2 to derive the number of fundamental properties of the one-dimensional ROF model.

Perhaps, the most significant consequence of Theorem 2 is that for any in-signal f belonging to $BV(I)$, we get $u_\lambda \rightarrow f$ strongly in $BV(I)$ as $\lambda \rightarrow 0+$, in particular we show that $\int |f' - u'_\lambda| dx \rightarrow 0$ as $\lambda \rightarrow 0+$, see Proposition 8. The usual Moreau–Yosida approximation result, see, for example, [2, Ch. 17], only contains the weaker assertion that $u_\lambda \rightarrow f$ in $L^2(I)$ and $\int_I |u'_\lambda| dx \rightarrow \int_I |f'| dx$ as λ tends to zero.

The literature on the ROF model is extensive, and many of the results about the one-dimensional case are often scattered throughout research articles and monographs as examples illustrating the more general multidimensional theory that is

their real focus. This sometimes makes these results hard to find and, once found, these examples may be hard to follow as they rest on the general theoretical framework already developed up to that point in the text. Newcomers to the field as well as more application-oriented researchers would probably welcome an introduction to the topic which could give an overview of the theory of the one-dimensional ROF model, on the one hand, and introduce him to some of the ideas which is used in the analysis of the general case.

The present paper may be seen as such an overview. Here, the theory is developed from scratch and known properties of the ROF model are collected in one place and given efficient proofs within a unified framework. The style is expository and is considered to be accessible to anyone who wants to learn about total variation techniques—a little bit of measure theory, knowledge of basic functional analysis and of Sobolev spaces in one dimension are the only prerequisites needed to follow the text. In fact, once the total variation of a function has been properly defined, it turns out that the theory of the ROF model in one dimension hinges on little more than the projection theorem (onto closed convex sets) and completion by squares. As we shall see, this elementary setting allows us to introduce and highlight, in a concrete setting, some of the interesting phenomena which occur in the analysis of more general convex variational problems.

Some of the known results, apart from Theorems 1 and 2, for which new proofs have been supplied are: (i) Propositions 4 where some basic properties of the ROF model are re-derived, and (ii) Propositions 6 and 5, where some precise results on the rate of convergence $u_\lambda \rightarrow f$, and of the value function $E_\lambda(u_\lambda)$, are given as λ tends to zero—collecting all such result in one place! (iii) Moreover, a new and slick proof of the fact that u_λ is a semi-group with respect to λ is given (Proposition 10), and we also derive the infinitesimal generator of this semi-group. (iv) Finally, we indicated how our method can be modified prove the “lower convex envelope” interpretation of the solution to the *isotonic regression* problem, Sect. 11.

Some entirely new results have also emerged: Proposition 8, which asserts that $u_\lambda \rightarrow f$ in $BV(I)$ as $\lambda \rightarrow 0+$ whenever $f \in BV(I)$ is new. The author also believes that the statement in part 2 of Proposition 9 is new and likewise is its consequence in Corollary 2. The explicit solution to the ROF model given in Example 5 also seems to appear here for the first time. And, at least in the context of the ROF model, the improved convergence rate proved in Proposition 6 seems to have gone unnoticed until now. Moreover, Proposition 3 gives an improvement to a known “gap” estimate found in [44]. Finally, our treatment of the *fused lasso model* in the continuous setting in Sect. 10 seems to be the first of its kind.

2 Our Analysis Toolbox

Throughout this paper, I denotes an open, bounded interval (a, b) , where $a < b$ are real numbers and $\bar{I} = [a, b]$ is the corresponding closed interval.

$C_0^1(I)$ denotes the space of continuously differentiable (test) functions $\xi : I \rightarrow \mathbf{R}$ with compact support in I , and $C(\bar{I})$ is the space of continuous functions on the closure of I .

For $1 \leq p \leq \infty$, $L^p(I)$ denotes the Lebesgue space of measurable functions $f : I \rightarrow \mathbf{R}$ with finite p -norm; $\|f\|_p := (\int_a^b |f(x)|^p dx)^{1/p} < \infty$, when p is finite, and $\|f\|_\infty = \text{ess sup}_{x \in I} |f(x)| < \infty$ when $p = \infty$. The space $L^2(I)$ is a Hilbert space with the inner product $\langle f, g \rangle = \langle f, g \rangle_{L^2(I)} := \int_a^b f(x)g(x) dx$ and the corresponding norm $\|f\| := (\langle f, f \rangle_{L^2(I)})^{1/2} = \|f\|_2$.

We are going to need the Sobolev spaces over L^2 :

$$H^1(I) = \{u \in L^2(I) : u' \in L^2(I)\},$$

where u' denotes the distributional derivative of u . This is a Hilbert space when equipped with the inner product $\langle u, v \rangle_{H^1} := \langle u, v \rangle + \langle u', v' \rangle$ and the corresponding norm $\|u\|_{H^1} = (\|u'\|_2^2 + \|u\|_2^2)^{1/2}$. Any $u \in H^1(I)$ can, after correction on a set of measure zero, be identified with a unique function in $C(\bar{I})$. In particular, a unique value $u(x)$ can be assigned to u for every $x \in \bar{I}$.

The following subspace of $H^1(I)$ plays an important role in our analysis:

$$H_0^1(I) = \{u \in H^1(I) : u(a) = 0 \text{ and } u(b) = 0\}.$$

Here, $\langle u, v \rangle_{H_0^1(I)} := \int_a^b u'(x)v'(x) dx$ defines an inner product on $H_0^1(I)$ whose induced norm

$$\|u\|_{H_0^1(I)} = \|u'\|_2$$

is equivalent to the norm inherited from $H^1(I)$ (by the Poincaré inequality).

Finally, let H be a (general) real Hilbert space with inner product between $u, v \in H$ denoted by $\langle u, v \rangle$ and the corresponding norm $\|u\| = \sqrt{\langle u, u \rangle}$. The following result is standard [13, Théorème V.2]:

Proposition 1 (Projection Theorem) *Let $K \subset H$ be a non-empty closed convex set. Then, for every $z \in H$ there exists a unique point $x_* \in K$ such that*

$$\|z - x_*\| = \min_{x \in K} \|z - x\|.$$

Moreover, the minimizer x_ is characterized by the following property:*

$$x_* \in K \text{ and } \langle z - x_*, x - x_* \rangle \leq 0, \text{ for all } x \in K. \tag{5}$$

The point x_* is called the projection of z onto K and is denoted $x_* = Pr_K(z)$.

Recall that the projection onto a set K is a non-expansive mapping:

$$\|Pr_K(y) - Pr_K(x)\| \leq \|y - x\| \tag{6}$$

for all $x, y \in H$. We also need the nonlinear mapping $Sr_K : H \rightarrow H$ defined as the residual after projection onto K ;

$$Sr_K(z) := z - Pr_K(z), \quad (z \in H). \tag{7}$$

This is called the shrinkage mapping associated with Pr_K , and it is also non-expansive,

$$\|Sr_K(y) - Sr_K(x)\| \leq \|y - x\| \tag{8}$$

for all $x, y \in H$.

The two estimates (6) and (8) follow immediately from the inequality,

$$\begin{aligned} &\|Pr_K(y) - Pr_K(x)\|^2 \\ &+ \|Sr_K(y) - Sr_K(x)\|^2 \leq \|y - x\|^2 \end{aligned} \tag{9}$$

which is valid for all $x, y \in H$. To derive this inequality, take $u = Pr_K(y) - Pr_K(x)$ and $v = Sr_K(y) - Sr_K(x)$ in the parallelogram identity $2(\|u\|^2 + \|v\|^2) = \|u - v\|^2 + \|u + v\|^2$. After the second term on the right-hand side has been expanded and the resulting identity simplified, we obtain:

$$\begin{aligned} &\|Pr_K(y) - Pr_K(x)\|^2 + \|Sr_K(y) - Sr_K(x)\|^2 \\ &= \|y - x\|^2 - 2\langle Pr_K(y) - Pr_K(x), Sr_K(y) - Sr_K(x) \rangle. \end{aligned}$$

Now, (9) follows if it can be proved that

$$-\langle Pr_K(y) - Pr_K(x), Sr_K(y) - Sr_K(x) \rangle \geq 0.$$

This follows from the characterization of the projection map (Theorem 1) by first applying (5) to the projection of y taking $v = Pr_K(x)$ and then applying (5) to the projection of x with $v = Pr_K(y)$. If the resulting inequalities are added, the above inequality follows. This proves (9).

Example 1 As an illustration, consider the existence and uniqueness of a taut string as defined by (3). Assume, for simplicity, that the cumulative signal F satisfies $F(a) = F(b) = 0$. Then, T_λ is a closed convex subset of $H_0^1(I)$ —non-empty because $F \in T_\lambda$. Since $\|u\|_{H_0^1(I)} = \|u'\|_{L^2(I)}$, we immediately see that (3) is equivalent to

$$\min_{W \in T_\lambda} \|z - W\|_{H_0^1(I)},$$

with $z = 0$ (the origin). It follows from the projection theorem that there exists a unique solution $W_\lambda = Pr_{T_\lambda}(0)$ of this minimization problem.

Example 2 An even simpler example is the following. The set $B := \{u \in L^2(I) : \|u\|_\infty \leq 1\}$ is a closed convex subset of $L^2(I)$. For $\lambda \geq 0$, let $\lambda B := \{\lambda u : u \in B\}$. The projection of $\varphi \in L^2(I)$ onto λB is given by truncation:

$$Pr_{\lambda B}(\varphi) = Pr_{[-\lambda, \lambda]} \circ \varphi, \tag{10}$$

where $Pr_{[-\lambda, \lambda]} : \mathbf{R} \rightarrow \mathbf{R}$ is the function given by

$$Pr_{[-\lambda, \lambda]}(t) = \begin{cases} t & \text{if } |t| \leq \lambda, \\ \lambda \operatorname{sign}(t) & \text{otherwise.} \end{cases} \tag{11}$$

This is, in fact, the projection in one dimension of $t \in \mathbf{R}$ onto the closed interval $[-\lambda, \lambda]$. Since $Pr_{[-\lambda, \lambda]} \circ \varphi$ clearly belongs to λB , we only need to verify the condition $\langle \varphi - Pr_{[-\lambda, \lambda]} \circ \varphi, v - Pr_{[-\lambda, \lambda]} \circ \varphi \rangle \leq 0$ for all $v \in \lambda B$. Now, this inequality follows from a simple calculation:

$$\begin{aligned} &\langle \varphi - Pr_{[-\lambda, \lambda]} \circ \varphi, v - Pr_{[-\lambda, \lambda]} \circ \varphi \rangle \\ &= \int_{\{\varphi \geq \lambda\}} (\varphi - \lambda)(v - \lambda) + \int_{\{\varphi \leq -\lambda\}} (\varphi + \lambda)(v + \lambda) \leq 0 \end{aligned}$$

which holds because $v \in \lambda B$ implies $-\lambda \leq v(x) \leq \lambda$ for almost all $x \in I$. The “residual” $\varphi - Pr_{\lambda \hat{C}}(\varphi)$ of the projection is given by the formula $Sr_{[-\lambda, \lambda]} \circ \varphi$, where $Sr_{[-\lambda, \lambda]} : \mathbf{R} \rightarrow \mathbf{R}$ is the so-called *soft threshold map* or *shrinkage map* defined by $Sr_{[-\lambda, \lambda]}(t) = t - Pr_{[-\lambda, \lambda]}(t)$. We shall meet both functions again in the sequel.

3 Precise Definition of the ROF Model

The expression $\int_I |u'| dx$ for the total variation makes sense for $u \in H^1(I)$ but is otherwise merely a convenient symbol. A more general and precise definition is needed: one which works when u' does not exist in the classical sense. The standard way to define the total variation is via duality: For $u \in L^1(I)$ set,

$$J(u) = \sup \left\{ \int_a^b u(x) \xi'(x) dx : \xi \in C_0^1(I), \|\xi\|_\infty \leq 1 \right\}.$$

If $J(u) < \infty$, u is said to be a *function of bounded variation* on I , and $J(u)$ is called the *total variation* of u (using the same notation as in [17,19]). The set of all integrable functions on I of bounded variation is denoted $BV(I)$, that is, $BV(I) = \{u \in L^1(I) : J(u) < \infty\}$. This becomes a Banach space when equipped with the norm

$\|u\|_{BV} := J(u) + \|u\|_{L^1}$. Notice that, as already mentioned, if $u \in H^1(I)$ $J(u) = \int_I |u'| dx < \infty$, so $u \in BV(I)$.

Let us illustrate how the definition works for a function with a jump discontinuity:

Example 3 Let $u(x) = \text{sign}(x)$ for $x \in I = (-1, 1)$. For any $\xi \in C_0^1(I)$, satisfying $|\xi(x)| \leq 1$ for all $x \in I$, we have

$$\begin{aligned} & \int_{-1}^1 u(x)\xi'(x) dx \\ &= \int_0^1 \xi'(x) dx - \int_{-1}^0 \xi'(x) dx = -2\xi(0) \leq 2, \end{aligned}$$

where equality holds for any admissible ξ which satisfies $\xi(0) = -1$. So $J(u) = 2$ and $u \in BV(I)$, as predicted by intuition. Here, the supremum is attained for many choices of ξ . This is not always the case; if $u(x) = x$ on $I = (0, 1)$, then $J(u) = 1$, but the supremum is not attained by any admissible test function.

The following lemma shows that the definition of the total variation J and the space $BV(I)$ can be moved to a Hilbert space setting involving L^2 and H_0^1 .

Lemma 2 Every $u \in BV(I)$ belongs to $L^2(I)$ and

$$J(u) = \sup_{\xi \in K} \langle u, \xi' \rangle_{L^2(I)}, \tag{12}$$

where $K = \{ \xi \in H_0^1(I) : \|\xi\|_\infty \leq 1 \}$, which is a closed and convex set in $H_0^1(I)$.

Proof If $u \in BV(I)$, then Sobolev’s lemma for functions of bounded variation, see [2, p. 152], ensures that $u \in L^\infty(I)$. This in turn implies $u \in L^2(I)$ because I is bounded. The (ordinary) Sobolev’s lemma asserts that $H_0^1(I)$ is continuously embedded in $L^\infty(I)$. Since K is the inverse image under the embedding map of the unit ball in $L^\infty(I)$, which is both closed and convex, we draw the conclusion that K is closed and convex in H_0^1 .

It only remains to prove (12). Clearly, $J(u)$ cannot exceed the right-hand side because the set $\{ \xi \in C_0^1(I) : \|\xi\|_\infty \leq 1 \}$ is contained in K . To verify that equality holds, it is enough to prove the inequality

$$\langle u, \xi' \rangle_{L^2(I)} \leq J(u)\|\xi\|_\infty, \quad \text{for all } \xi \in H_0^1(I), \tag{13}$$

as it implies that the right-hand side of (12) cannot exceed $J(u)$. To do this, we first notice that the inequality holds for all $\zeta \in C_0^1(I)$. This follows by applying homogeneity to the definition of $J(u)$. Second, if $\xi \in H_0^1(I)$ we can use that $C_0^1(I)$ is dense in $H_0^1(I)$ and find functions $\zeta_n \in C_0^1(I)$ such that $\zeta_n \rightarrow \xi$ in $H_0^1(I)$ (and in $L^\infty(I)$ by the continuous embedding). It follows that

$$\langle u, \xi' \rangle_{L^2(I)} = \lim_{n \rightarrow \infty} \langle u, \zeta_n' \rangle_{L^2(I)}$$

$$\leq J(u) \lim_{n \rightarrow \infty} \|\zeta_n\|_\infty = J(u)\|\xi\|_\infty,$$

which establishes (13) and the proof is complete. \square

The inequality (13) shows that $\xi \rightarrow \langle u, \xi' \rangle$ extends to a continuous linear functional on $C_0(I)$. Riesz’ representation theorem (cf., e.g. [2, Thm. 1.54]) therefore implies that the distributional derivative u' of $u \in BV(I)$ is a signed (Radon) measure μ on I , and that we may write

$$\langle u, \xi' \rangle_{L^2(I)} = \int_I \xi d\mu.$$

This will be useful later on.

We can now give the precise definition of the ROF model: For any $f \in L^2(I)$ and any real number $\lambda > 0$, the ROF functional is the function $E_\lambda : BV(I) \rightarrow \mathbf{R}$ given by

$$E_\lambda(u) = \lambda J(u) + \frac{1}{2} \|f - u\|_{L^2(I)}^2. \tag{14}$$

Denoising according to the ROF model is the mapping $L^2(I) \ni f \mapsto u_\lambda \in BV(I)$ defined by (1). To emphasize the role of the in-signal f , we sometimes write $E_\lambda(f; u)$ instead of $E_\lambda(u)$ and denote the corresponding ROF minimizer u_λ by the more elaborate $u_\lambda(f)$. Well-posedness of the ROF model is demonstrated in Sect. 5 after some simple observations about the symmetry properties of $E_\lambda(f, \cdot)$ have been presented in the next section.

4 Simple Symmetries of the ROF Functional

We begin with some simple observations. For any $u \in BV(I)$, the total variation functional J satisfies $J(u + c) = J(u)$ for all $c \in \mathbf{R}$ and $J(cu) = cJ(u)$ whenever $c > 0$. This implies the following formulas for the ROF functional:

$$\begin{aligned} E_\lambda(f; u) &= E_\lambda(f + c; u + c), \quad \text{for } c \in \mathbf{R}, \text{ and} \\ c^2 E_\lambda(f, u) &= E_{c\lambda}(cf; cu), \quad \text{for } c > 0. \end{aligned}$$

When u_λ denotes the minimizer of $E_\lambda(f; \cdot)$, the first of these identities implies that the minimizer of the functional $E_\lambda(f + c; \cdot)$ is the function $u_\lambda + c$. Likewise, if $c > 0$, the second identity shows that the function $cu_{\lambda/c}$ minimizes $E_\lambda(cf; \cdot)$. These symmetry properties (translation and scaling of the dependent variable) can be expressed as

$$\begin{aligned} u_\lambda(f + c) &= u_\lambda(f) + c \quad \text{and} \quad (c \in \mathbf{R}) \tag{15} \\ u_\lambda(cf) &= cu_{\lambda/c}(f), \quad (c > 0). \tag{16} \end{aligned}$$

We also have $E_\lambda(-f; -u) = E_\lambda(f; u)$, so changing the sign of the signal will just change the sign of the minimizer; $u_\lambda(-f) = -u_\lambda(f)$. For any integrable function f , let $f_I =$

$|I|^{-1} \int_I f \, dx$ denote the mean value of f on the interval I . The formula (15) with $c = -f_I$ becomes

$$u_\lambda(f) = u_\lambda(f - f_I) + f_I \tag{17}$$

which shows that we may restrict ourselves to consider signals with zero mean, $f_I = 0$. In fact, the following easily verified identity

$$E_\lambda(u - u_I) = E_\lambda(u) - \frac{1}{2} \|u_I\|^2$$

implies that the minimizer u_λ of E_λ has zero mean. (For general f , we have $(u_\lambda)_I = f_I$ which is seen by taking mean values in (17).) Consequently, if $f_I = 0$ it is enough to minimize E_λ over functions u with zero mean.

5 Existence Theory for the ROF Model

The following theorem contains the key result for developing the properties of the ROF model. It proves existence and uniqueness of the ROF minimizer u_λ for a general signal f in L^2 and gives a necessary and sufficient characterization of this minimizer in terms of itself and a dual variable (ξ_λ in the theorem). Throughout the analysis, we assume that f has mean value zero in $I = (a, b)$. This assumption, which is strictly speaking not needed in order for the result to be true, implies that the cumulative signal $F(x)$ satisfies $F(a) = F(b) = 0$, and hence, $F \in H_0^1(I)$, which will simplify the exposition.

Theorem 3 *We have the equality*

$$\min_{u \in \text{BV}(I)} E_\lambda(u) = \max_{\xi \in K} \frac{1}{2} \left\{ \|f\|_{L^2(I)}^2 - \|f - \lambda \xi'\|_{L^2(I)}^2 \right\}, \tag{18}$$

with the minimum achieved by a unique $u_\lambda \in \text{BV}(I)$ and the maximum achieved by a unique $\xi_\lambda \in K$. The two functions are related by the identities

$$u_\lambda = f - \lambda \xi'_\lambda, \tag{19a}$$

and

$$J(u_\lambda) = \langle u_\lambda, \xi'_\lambda \rangle_{L^2(I)}. \tag{19b}$$

Moreover, if $u_\lambda \neq 0$, then $\|\xi_\lambda\|_\infty = 1$. Conversely, the conditions (19a) and (19b) characterize the solution; if a pair of functions $\bar{u} \in \text{BV}(I)$ and $\bar{\xi} \in K$ satisfy $\bar{u} = f - \lambda \bar{\xi}'$ and $J(\bar{u}) = \langle \bar{u}, \bar{\xi}' \rangle_{L^2(I)}$, then $\bar{u} = u_\lambda$ and $\bar{\xi} = \xi_\lambda$.

This result is an instance of the Fenchel–Rockafellar theorem, see, for example, Brezis [13, p. 11]. It is tailored with our specific needs in mind and will be proved with our bare hands using the projection theorem. (The general version was used by Hintermüller and Kunisch [26] in their analysis of the multidimensional ROF model with the “Manhattan metric”.) In one of the first theoretical analyses of the ROF model, Chambolle and Lions [19] proved the existence of a minimizer (for a more general case) using the standard argument where a minimizing sequence is shown to converge weakly to function which can be shown to be the desired solution. The equality (18) has played an important role in the development of numerical algorithms for total variation minimization, either directly, as in Zhu et al. [44] or, more indirectly, as in Chambolle [17].

Before the proof starts, let us remind the reader of the following general fact: If M and N are arbitrary non-empty sets and $\Phi : M \times N \rightarrow \mathbf{R}$ is any real-valued function, then the inequality

$$\inf_{x \in M} \sup_{y \in N} \Phi(x, y) \geq \sup_{y \in N} \inf_{x \in M} \Phi(x, y), \tag{20}$$

is always true, as is easily checked. The use of inf’s and sup’s is crucial, as neither the greatest lower bounds nor the least upper bounds are necessarily attained.

Proof Since $E_\lambda(u) = \sup_{\xi \in K} \lambda \langle u, \xi' \rangle + \frac{1}{2} \|f - u\|^2$, it follows from (20) that

$$\inf_{u \in \text{BV}(I)} E_\lambda(u) \geq \sup_{\xi \in K} \left\{ \inf_{u \in \text{BV}(I)} \lambda \langle u, \xi' \rangle + \frac{1}{2} \|u - f\|^2 \right\}.$$

We first solve, for $\xi \in K$ fixed, the minimization problem on the right-hand side. Expanding $\|f - u\|^2$ and completing squares with respect to u yields:

$$\begin{aligned} \lambda \langle u, \xi' \rangle + \frac{1}{2} \|u - f\|^2 \\ = \frac{1}{2} \left\{ \|u - (f - \lambda \xi')\|^2 - \|f - \lambda \xi'\|^2 + \|f\|^2 \right\} \end{aligned}$$

The right-hand side is clearly minimized by the $L^2(I)$ -function $u = f - \lambda \xi'$ and

$$\inf_{u \in \text{BV}(I)} E_\lambda(u) \geq \sup_{\xi \in K} \frac{1}{2} \left\{ \|f\|^2 - \|f - \lambda \xi'\|^2 \right\} \tag{21}$$

holds. The maximization problem on the right-hand side is equivalent to

$$\begin{aligned} \inf_{\xi \in K} \|f - \lambda \xi'\| &= \inf_{\xi \in K} \|F' - \lambda \xi'\| \\ &= \lambda \inf_{\xi \in K} \|\lambda^{-1} F - \xi\|_{H_0^1(I)}. \end{aligned} \tag{22}$$

By Proposition 1, this problem has the unique solution $\xi_\lambda = P_K(\lambda^{-1}F) \in K$, so the supremum is attained in (21). Now, let the function u_λ be defined by (19a) in the theorem. A priori, u_λ belongs to $L^2(I)$, but we are going to show that $u_\lambda \in \text{BV}(I)$: the characterization of ξ_λ according in the projection theorem states that $\xi_\lambda \in K$ and $\langle f - \lambda\xi'_\lambda, \lambda\xi'_\lambda - \lambda\xi'_\lambda \rangle \leq 0$ for all $\xi \in K$. If we use the definition of u_λ and divide by $\lambda > 0$, this characterization becomes

$$\langle u_\lambda, \xi' \rangle \leq \langle u_\lambda, \xi'_\lambda \rangle \quad \text{for all } \xi \in K,$$

where the right-hand side is finite. It follows from the definition of the total variation that $u_\lambda \in \text{BV}(I)$ with $J(u_\lambda) = \langle u_\lambda, \xi'_\lambda \rangle$, as asserted in the theorem. (This reasoning can be reversed; if (19b) is true, then ξ_λ is the minimizer in (22).) Also, if $u_\lambda \neq 0$ then $\|\xi_\lambda\|_\infty < 1$ is not consistent with the maximizing property (19b), and hence, $\|\xi_\lambda\|_\infty = 1$, as claimed.

It remains to be verified that u_λ minimizes E_λ and that equality holds in (21). This follows from a direct calculation:

$$\begin{aligned} \inf_{u \in \text{BV}(I)} E_\lambda(u) &\geq \max_{\xi \in K} \frac{1}{2} \left\{ \|f\|^2 - \|f - \lambda\xi'\|^2 \right\} \\ &= \frac{1}{2} \|f\|^2 - \frac{1}{2} \|u_\lambda\|^2 \\ &= \frac{1}{2} \|f\|^2 + \frac{1}{2} \|u_\lambda\|^2 - \|u_\lambda\|^2 \\ &= \frac{1}{2} \|f\|^2 + \frac{1}{2} \|u_\lambda\|^2 - \langle u_\lambda, f - \lambda\xi'_\lambda \rangle \\ &= \frac{1}{2} \|f - u_\lambda\|^2 + \langle u_\lambda, \lambda\xi_\lambda \rangle \\ &= \frac{1}{2} \|f - u_\lambda\|^2 + \lambda J(u_\lambda) \\ &= E_\lambda(u_\lambda). \end{aligned}$$

So $\inf E_\lambda(u) = E_\lambda(u_\lambda)$, the infimum is attained, and equality holds in (21). The inequality $E_\lambda(u) - E_\lambda(u_\lambda) \geq \frac{1}{2} \|u - u_\lambda\|^2$ implies the uniqueness of u_λ . The converse statement is proved by backtracking the steps of the above proof. \square

The equivalence of the two denoising models can now be established:

Proof of Theorem 1 It follows from Theorem 3 that the minimizer u_λ of the ROF functional is given by $u_\lambda = f - \lambda\xi'_\lambda$ where ξ_λ is the unique solution of

$$\min_{\xi \in K} \frac{1}{2} \|f - \lambda\xi'\|^2_{L^2(I)}. \tag{23}$$

If we introduce the new variable $W := F - \lambda\xi$, where $F \in H_0^1(I)$ is the cumulative signal, then $W \in H_0^1(I)$ and the condition $\|\xi\|_\infty \leq 1$ implies that W satisfies $F(x) - \lambda \leq W(x) \leq F(x) + \lambda$ on I . Therefore, (23) is equivalent to

$$\min_{W \in T_\lambda} \frac{1}{2} \|W'\|^2_{L^2(I)},$$

which is the minimization problem in step 3 of the taut string algorithm whose solution is denoted by W_λ . It follows that $W_\lambda = F - \lambda\xi_\lambda$ and differentiation yields $f_\lambda = W'_\lambda = f - \lambda\xi'_\lambda = u_\lambda$, the desired result. \square

Our proof of Theorem 1 is essentially a change of variables and, as such, becomes almost a “derivation” of the taut string interpretation. We also get the existence and uniqueness of solutions to both models in one stroke. The proof given in [24] first shows that u_λ and W'_λ satisfy the same set of three necessary conditions, and that these conditions admit at most one solution. Then, it proceeds to drive home the point by establishing existence separately for both models. The argument assumes $f \in L^\infty$ and involves a fair amount of measure-theoretic considerations. The proof of equivalence given in [39] is based on a thorough functional analytic study of Meyer’s G -norm and is not elementary.

The last two proofs contain the following useful observations:

Corollary 1 *The optimal dual variable ξ_λ is given by projection in $H_0^1(I)$ onto K ,*

$$\xi_\lambda = Pr_K(\lambda^{-1}F),$$

and the taut string W_λ by the shrinkage map,

$$W_\lambda = Sr_{\lambda K}(F) = \lambda Sr_K(\lambda^{-1}F),$$

where $F \in H_0^1(I)$ is the cumulative signal and $\lambda > 0$.

Denoising according to the ROF model is a mapping $f \mapsto u_\lambda(f)$ which is contractive and continuous with respect to $\lambda > 0$:

Proposition 2 (a) *For signals $f, \bar{f} \in L^2(I)$, we have*

$$\|u_\lambda(\bar{f}) - u_\lambda(f)\|_{L^2(I)} \leq \|\bar{f} - f\|_{L^2(I)}.$$

(b) *For any $f \in L^2(I)$,*

$$\|u_\lambda(f) - u_\nu(f)\|_{L^2(I)} \leq \frac{2|\lambda - \nu|}{\max(\lambda, \nu)} \|f\|_{L^2(I)}$$

for all $\lambda, \nu > 0$,

The first assertion of the proposition is a well-known property of the Moreau–Yosida approximation (or of the proximal map), see [6, Theorem 17.2.1]. Both assertions are easy consequences of the corollary.

Proof (a) We apply the taut string interpretation. Let F and \bar{F} be the cumulative signals corresponding to f and \bar{f} , and let W_λ and \bar{W}_λ be the associated taut strings. The non-expansiveness of the shrinkage map (8) yields

$$\begin{aligned} \|u_\lambda(\bar{f}) - u_\lambda(f)\|_{L^2(I)} &\leq \|\bar{W}_\lambda - W_\lambda\|_{H_0^1(I)} \\ &= \|Sr_{\lambda K}(\bar{F}) - Sr_{\lambda K}(F)\|_{H_0^1(I)} \\ &\leq \|\bar{F} - F\|_{H_0^1(I)} \\ &= \|\bar{f} - f\|_{L^2(I)}, \end{aligned}$$

which is the desired estimate.

(b) This time, we use the non-expansiveness of the projection map to obtain a bound on the difference:

$$\begin{aligned} \|u_\lambda - u_\nu\|_{L^2(I)} &= \|\lambda\xi'_\lambda - \nu\xi'_\nu\|_{L^2(I)} \\ &= \|\lambda\xi_\lambda - \nu\xi_\nu\|_{H_0^1(I)} \\ &= \|\lambda Pr_K(\lambda^{-1}F) - \nu Pr_K(\nu^{-1}F)\|_{H_0^1(I)} \\ &\leq \|\lambda Pr_K(\lambda^{-1}F) - \lambda Pr_K(\nu^{-1}F)\|_{H_0^1} \\ &\quad + \|\lambda Pr_K(\nu^{-1}F) - \nu Pr_K(\nu^{-1}F)\|_{H_0^1} \\ &\leq \lambda\|\lambda^{-1}F - \nu^{-1}F\|_{H_0^1} \\ &\quad + \nu^{-1}|\lambda - \nu|\|F\|_{H_0^1} \\ &\leq 2\frac{|\lambda - \nu|}{\nu}\|f\|_{L^2(I)}. \end{aligned}$$

Here, it was used that $\|Pr_K(F)\|_{H_0^1(I)} \leq \|F\|_{H_0^1(I)}$. This follows from the non-expansiveness of the projection since $Pr_K(0) = 0$. If the roles of λ and ν are interchanged, then we get another bound,

$$\|u_\lambda - u_\nu\|_{L^2(I)} \leq 2\frac{|\lambda - \nu|}{\lambda}\|f\|_{L^2(I)}.$$

Taking the smaller of the two gives the desired result. \square

It is an interesting observation that Theorem 3 associates a *unique* test function (or *dual variable*), $\xi_\lambda \in K$, with the solution u_λ of the ROF model, namely the one which satisfies $J(u_\lambda) = \langle u_\lambda, \xi'_\lambda \rangle_{L^2}$. In particular, as demonstrated in Example 3, because there are functions u for which the supremum in the definition of $J(u)$ is not attained. An explicit example of a ROF minimizer looks as follows:

Example 4 Consider the step function on $I = (-1, 1)$,

$$f(x) = \text{sign}(x), \quad x \in I.$$

An easy calculation, based on the taut string interpretation, shows that

$$u_\lambda = (1 - \lambda)_+ \text{sign}(x) \quad \text{and} \quad \xi_\lambda = \frac{|x| - 1}{\max(1, \lambda)} \in H_0^1(I).$$

In fact, the cumulative signal is $F(x) = |x| - 1$ and, in the case when $0 < \lambda < 1$, it is easy to check that $W_\lambda(x) = (1 - \lambda)F(x)$. It follows from the taut string interpretation that $u_\lambda = W'_\lambda$ and ξ_λ can then be computed from (19a). Notice that the dual variable ξ_λ is not in $C_0^1(I)$, so the extension of the space of test functions from C_0^1 to H_0^1 is essential to our theory. For $\lambda \geq 1$, we find $u_\lambda = 0$ (draw a picture) and $\xi_\lambda = \lambda^{-1}F(x)$. Note that $\|\xi_\lambda\|_\infty = 1$ when $u_\lambda \neq 0$.

This example was considered in Strong and Chan [42], one of the very first papers treating explicit solutions to the ROF model. At the time, they did neither have access to the the duality formulation of the ROF model nor its taut string interpretation, and their solution consequently goes over several pages.

Another observation is the following: if we derive the Euler–Lagrange equation for the ROF functional $E_\lambda(u) = \lambda \int_I |u'| + (1/2) \int_I (f - u)^2$ in a purely formal manner, we get

$$-\lambda \frac{d}{dx} \left(\frac{u'}{|u'|} \right) = f - u.$$

The meaning of the nonlinear differential operator on the left-hand side of this equation is unclear as it stands. However, we know from (19a) in Theorem 3 that the minimizer u_λ satisfies $u_\lambda = f - \lambda\xi'_\lambda$, which, if compared to the formal Euler–Lagrange equation above, leads to the interpretation,

$$-\frac{d}{dx} \left(\frac{u'_\lambda}{|u'_\lambda|} \right) = \xi'_\lambda \tag{24}$$

by homogeneity. In Example 4, we found that the ROF minimizer was proportional to the signal f , namely $u_\lambda = (1 - \lambda)f$ for $0 < \lambda < 1$. Therefore,

$$-\frac{d}{dx} \left(\frac{u'_\lambda}{|u'_\lambda|} \right) = -\frac{d}{dx} \left(\frac{f'}{|f'|} \right).$$

On the other hand, $\xi'_\lambda = \min(1, \lambda^{-1})f = f$ which combined with the above equation and (24) yields

$$-\frac{d}{dx} \left(\frac{f'}{|f'|} \right) = f, \tag{25}$$

that is, f is an eigenfunction of the nonlinear operator $u \mapsto (d/dx)(u'/|u'|)$. Eigenfunctions of this sort have been extensively studied in the two-dimensional setting in Bellettini et al. [9]. There it was shown that if f is any such eigenfunction, then $u_\lambda = (1 - \lambda)_+ f$ minimizes the two-dimensional ROF functional with regularization weight λ (see also [1,8]).

6 Certifying the Quality of Approximate Solutions

If we define the functional

$$L_\lambda(\xi) = \frac{1}{2} \{ \|f\|^2 - \|f - \lambda\xi'\|^2 \}, \tag{26}$$

then Theorem 3 states that $E_\lambda(u) \geq L_\lambda(\xi)$ holds for all $u \in \text{BV}(I)$ and all $\xi \in K$, and that we have equality if and only if $u = u_\lambda$ and $\xi = \xi_\lambda$. In this section, we ask ourselves the following question: suppose (u, ξ) is an approximate solution in the sense that $E_\lambda(u)$ is only slightly bigger than $L_\lambda(\xi)$, does it follow that (u, ξ) is a good approximation to the ROF minimizer in the sense that u is close to u_λ and ξ close to ξ_λ in their respective functional spaces?

To investigate this question, we follow Zhu et al. [44] and introduce the “gap function”

$$\begin{aligned} \text{gap}(u, \xi) &:= E_\lambda(u) - L_\lambda(\xi) \\ &= \lambda J(u) + \frac{1}{2} \|f - u\|_{L^2(\Omega)}^2 \\ &\quad + \frac{1}{2} \|f - \lambda\xi'\|_{L^2(\Omega)}^2 - \frac{1}{2} \|f\|_{L^2(\Omega)}^2, \end{aligned} \tag{27}$$

where $u \in \text{BV}(I)$ and $\xi \in K$. Using this “gap function”, we can prove the following refined version of an estimate found in [44]:

Proposition 3 *For all $u \in \text{BV}(\Omega)$ and $\xi \in K$, the following estimate holds*

$$\text{gap}(u, \xi) \geq \frac{1}{2} \|u - u_\lambda\|_{L^2(\Omega)}^2 + \frac{1}{2} \|\lambda\xi' - \lambda\xi'_\lambda\|_{L^2(\Omega)}^2,$$

where u_λ and ξ_λ are the solution of the ROF problem and its dual.

If we assume that the function u and the dual variable ξ satisfy the relation $u = f - \lambda\xi'$, as is the case in many numerical algorithms for TV-minimization, then we get the following estimate as a corollary of the above proposition:

$$\text{gap}(u, \xi) \geq \|u - u_\lambda\|^2. \tag{28}$$

This is the actual result stated in [44].

Proof We first introduce any ROF solution pair u_λ, ξ_λ into the definition of the duality gap and expand:

$$\begin{aligned} \text{gap}(u, \xi) &:= \lambda J(u) + \frac{1}{2} \|f - u_\lambda + u_\lambda - u\|^2 \\ &\quad + \frac{1}{2} \|f - \lambda\xi'_\lambda + \lambda\xi'_\lambda - \lambda\xi'\|^2 - \frac{1}{2} \|f\|^2 \\ &= \lambda J(u) - \langle u, \lambda\xi'_\lambda \rangle + \langle u_\lambda, \lambda\xi'_\lambda \rangle \end{aligned}$$

$$\begin{aligned} &+ \langle u_\lambda, \lambda\xi'_\lambda - \lambda\xi' \rangle \\ &+ \frac{1}{2} \|f - u_\lambda\|^2 + \frac{1}{2} \|f - \lambda\xi'_\lambda\|^2 - \frac{1}{2} \|f\|^2 \\ &+ \frac{1}{2} \|u - u_\lambda\|^2 + \frac{1}{2} \|\lambda\xi' - \lambda\xi'_\lambda\|^2 \\ &\geq \frac{1}{2} \|f - u_\lambda\|^2 + \langle u_\lambda, \lambda\xi'_\lambda \rangle + \frac{1}{2} \|f - u_\lambda\|^2 \\ &\quad - \frac{1}{2} \|f\|^2 + \frac{1}{2} \|u - u_\lambda\|^2 + \frac{1}{2} \|\lambda\xi' - \lambda\xi'_\lambda\|^2 \\ &= \frac{1}{2} \|u - u_\lambda\|^2 + \frac{1}{2} \|\lambda\xi' - \lambda\xi'_\lambda\|^2, \end{aligned}$$

and the proof is complete. □

7 Consequences of Theorem 3 and the Taut String Interpretation

We now prove some known, and some new, properties of the ROF model.

The taut string algorithm suggests that $W_\lambda = 0$, and therefore, $u_\lambda = 0$, when λ is sufficiently large, and that W_λ must touch the sides $F \pm \lambda$ of the tube T_λ when λ is small. These assertions can be made precise:

- Proposition 4** (a) *The denoised signal $u_\lambda = 0$ if and only if $\lambda \geq \|F\|_\infty$, and*
 (b) *if $0 < \lambda < \|F\|_\infty$ then $\|F - W_\lambda\|_\infty = \lambda$.*
 (c) $\|W_\lambda\|_\infty = \max(0, \|F\|_\infty - \lambda)$.

The results (a) and (b) are well known, and proofs, valid in the multidimensional case, can be found in Meyer’s treatise [30]. The natural estimate in (c) seems to be stated here for the first time. Notice that the maximum norm $\|F\|_\infty$ of the cumulative signal F coincides, in one dimension, with the Meyer’s G -norm $\|f\|_*$ of the signal f . Theorem 3 and the taut string interpretation of the ROF model allow us to give very short and direct proofs of all three properties.

Proof (a) By Theorem 1, the denoised signal u_λ is zero if and only if the taut string W_λ is zero. We know that $W_\lambda = F - \lambda\xi_\lambda$ where, as seen from (22), ξ_λ is the projection in $H_0^1(I)$ of $\lambda^{-1}F$ onto the closed convex set K . Therefore, $u_\lambda = 0$ if and only if $\lambda^{-1}F \in K$, that is, if and only if $\|F\|_\infty \leq \lambda$, as claimed.

(b) If $0 < \lambda < \|F\|_\infty$, then $u_\lambda \neq 0$ hence $\|\xi_\lambda\|_\infty = 1$, by Theorem 3. The assertion now follows by taking norms in the identity $\lambda\xi_\lambda = F - W_\lambda$.

(c) The equality clearly holds when $\lambda \geq \|F\|_\infty$ because $W_\lambda = 0$ by (a). When $c := \|F\|_\infty - \lambda > 0$, we use a truncation argument: If W belongs to T_λ , then so does $\hat{W} := \min(c, W)$, in particular $c > 0$ ensures that $\hat{W}(a) = \hat{W}(b) = 0$. Since $E(\hat{W}) \leq E(W)$, and W_λ is the (unique) minimizer of E over T_λ , we conclude that $\max_I W_\lambda \leq c$.

A similar argument gives $-\min_I W_\lambda \leq c$. Thus, $\|W_\lambda\|_\infty \leq \max(0, \|F\|_\infty - \lambda)$. The reverse inequality follows from (b). \square

As an application of the above result and of Theorem 3, we consider the following exactly solvable example which confirms the result of several numerical simulations and which would most likely be out of reach if one had used the methods developed in [42]:

Example 5 On the interval, $I = (0, 2\pi]$ and for some positive integer, n define $f(x) = \cos(nx)$. If, for simplicity, we change the setting a little and impose periodic boundary conditions on the admissible functions u in $E_\lambda(u)$, then we find that

If $\lambda \geq 1/n$, then $u_\lambda = 0$ and for $0 < \lambda < 1/n$, there exists a number $\alpha = \alpha(\lambda)$ such that $0 < \alpha < 1$ and the denoised signal is given by truncation,

$$u_\lambda = Pr_{[-\alpha, \alpha]} \circ f. \tag{29}$$

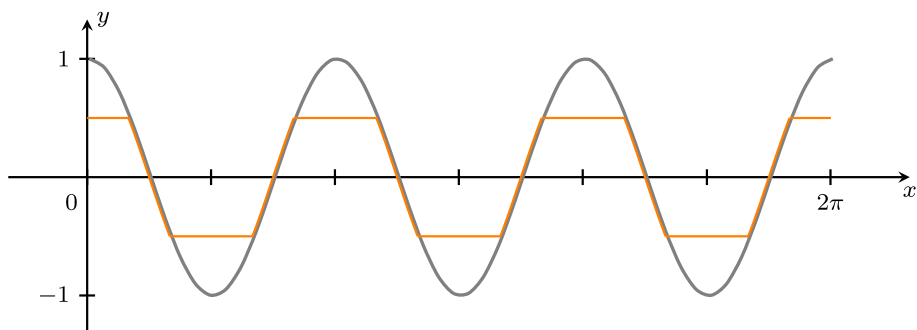
Here, $Pr_{[-\alpha, \alpha]} : \mathbf{R} \rightarrow \mathbf{R}$ denotes projection onto the closed (convex) interval $[-\alpha, \alpha]$. The case with $n = 3$ is illustrated in Fig. 2.

Let us first clarify what we mean by periodic boundary conditions. It is the same as defining and minimizing the ROF functional E_λ for functions defined on the unit circle T rather than over an interval I . The theory developed earlier, in particular Theorem 3, still holds if we replace $H_0^1(I)$ by $H^1(T)$ and define the closed convex set K accordingly.

First, notice that the cumulative function $F(X) = \int_0^x f(t) dt = n^{-1} \sin(nx)$ satisfies $\|F\|_\infty = 1/n$ so, by Proposition 4, $u_\lambda = 0$ if $\lambda \geq 1/n$. Therefore, it is enough to study the case when $0 < \lambda < 1/n$. To deal with that case, we set $u_* = Pr_{[-\alpha, \alpha]} \circ f$ and try to find a dual function $\xi_* \in K$ such that the sufficient conditions for optimality (19a) and (19b) hold for then $u_* = u_\lambda$ by Theorem 3. It follows from the first of these conditions that ξ_* must be unique periodic function which satisfies

$$\xi_*' = \frac{f - u_*}{\lambda},$$

Fig. 2 In-signal $f(x) = \cos(nx)$ with $n = 3$ and the ROF minimizer u_λ with $3\lambda = \sin(\pi/3) - (\pi/3) \cos(\pi/3) = (3\sqrt{3} - \pi)/6$ and $\alpha(\lambda) = \cos(\pi/3) = 1/2$ superimposed



whose mean value over I is zero. It remains to be verified that $\|\xi_*\|_\infty \leq 1$ and that the second of the two conditions holds. We consider the latter first. It is easy to see that

$$J(u_*) = 4n\alpha.$$

we consequently need to show that there exists a number α between zero and one such that $\langle u_*, \xi_*' \rangle = 4n\alpha$. Using periodicity and symmetry of the functions u_* and ξ_*' , it is easy to see that

$$\langle u_*, \xi_*' \rangle = 2n \int_{-a}^a \alpha \frac{\cos(nt) - \alpha}{\lambda} dt$$

where a in the limits of the integral is the smallest positive number such that $\alpha = \cos(na)$. We note that $0 < a < \pi/2n$, see Fig. 2. Evaluation of this integral leads to the following condition on the number a , and therefore, on α :

$$\sin(na) - na \cos(na) = n\lambda \quad (0 < a < \pi/2n). \tag{30}$$

Since the function $\theta \mapsto \sin \theta - \theta \cos \theta$ is strictly increasing on the interval $[0, \pi/2]$ with range $[0, 1]$, it follows that Eq. (30) has a unique solution $a_\lambda \in (0, \pi/2n)$. Thus, we have shown that the condition (19b) holds if we take $\alpha = \alpha(\lambda) = \cos(na_\lambda)$. The proof will be complete if we can verify that ξ_* satisfies $\|\xi_*\|_\infty \leq 1$. But this essentially follows from the work already done and the calculation, which holds for $0 \leq x \leq a_\lambda$:

$$\begin{aligned} 0 \leq \xi_*(x) &= \int_0^x \xi_*'(t) dt \\ &= \int_0^x \frac{\cos(nt) - \cos(na_\lambda)}{\lambda} dt \\ &\leq \int_0^{a_\lambda} \frac{\cos(nt) - \cos(na_\lambda)}{\lambda} dt \\ &= \frac{\sin(na_\lambda) - na_\lambda \cos(na_\lambda)}{n\lambda} = 1, \end{aligned}$$

since a_λ is the solution of (30). Since the function is periodic, it now easy to see that $|\xi_*(x)| \leq 1$ on the entire interval I and the verification is complete.

We continue with some additional properties of the ROF model. Define, for $\lambda > 0$, the *value function*

$$e(\lambda) := \inf_{u \in \text{BV}(I)} E_\lambda(u),$$

that is, $e(\lambda) = E_\lambda(u_\lambda)$. The next two theorems contain the essentially well-known results.

Proposition 5 *The function $e : (0, +\infty) \rightarrow (0, +\infty)$ is non-decreasing and concave, hence continuous, and satisfies $e(\lambda) = \|f\|^2/2$ for $\lambda \geq \|F\|_\infty$. Moreover, for $f \in L^2(I)$*

$$\lim_{\lambda \rightarrow 0+} e(\lambda) = 0.$$

and if $f \in \text{BV}(I)$ then $e(\lambda) = O(\lambda)$ as $\lambda \rightarrow 0+$.

Proof If $\lambda_2 \geq \lambda_1 > 0$, then the inequality $E_{\lambda_2}(u) \geq E_{\lambda_1}(u)$ holds trivially for all u . Taking infimum over the functions in $\text{BV}(I)$ yields $e(\lambda_2) \geq e(\lambda_1)$, so e is non-decreasing.

For any u , the right-hand side of the inequality

$$e(\lambda) \leq E_\lambda(u) = \lambda J(u) + \frac{1}{2} \|u - f\|^2,$$

is an affine and therefore a concave, function of λ . Because the infimum of any family of concave functions is again concave, it follows that $e(\lambda) = \inf_{u \in \text{BV}(I)} E_\lambda(u)$ is concave.

For $\lambda \geq \|F\|_\infty$, we know from the previous theorem that $u_\lambda = 0$, so $e(\lambda) = E_\lambda(0) = \|f\|^2/2$.

To prove the assertion about $e(\lambda)$ as λ tends to zero from the right, we first assume that $f \in \text{BV}(I)$, in which case it follows that $0 < e(\lambda) \leq E_\lambda(f) = \lambda J(f)$, so $e(\lambda) = O(\lambda)$ because $J(f) < \infty$.

If we merely have $f \in L^2(I)$, an approximation argument is needed: For any $\epsilon > 0$, take a function $f_\epsilon \in H_0^1(I)$ such that $\|f - f_\epsilon\|^2/2 < \epsilon$. Then $f_\epsilon \in \text{BV}(I)$ and $0 \leq e(\lambda) \leq E_\lambda(f_\epsilon) < \lambda J(f_\epsilon) + \epsilon$. It follows that $0 \leq \limsup_{\lambda \rightarrow 0+} e(\lambda) < \epsilon$. Since ϵ is arbitrary, we get $\lim_{\lambda \rightarrow 0+} e(\lambda) = 0$. \square

The map $f \rightarrow u_\lambda$ is in fact the Moreau–Fenchel resolvent of the total variation functional J , see [6, Sect. 17.2.1], and the following proposition is a therefore a special case of a much more general result from the theory of Moreau–Fenchel approximations. Notice, however, that the second part of our proposition contains a refined quantification of the rate of convergence of u_λ to f as $\lambda \rightarrow 0$ in that the common $O(\lambda)$ is replaced by $o(\lambda)$. The latter is *not* easily located in the literature.

Proposition 6 *For any $f \in L^2(I)$, we have $u_\lambda \rightarrow f$ in L^2 as $\lambda \rightarrow 0+$. Moreover, if $f \in \text{BV}(I)$ then $J(u_\lambda) \rightarrow J(f)$ and $\|u_\lambda - f\|_{L^2(I)} = o(\lambda^{1/2})$ as $\lambda \rightarrow 0+$.*

Proof The obvious inequality $\|f - u_\lambda\|^2/2 \leq e(\lambda)$ and the fact $\lim_{\lambda \rightarrow 0+} e(\lambda) = 0$, proved above, implies the first assertion. When $f \in \text{BV}(I)$, it follows from the inequality $\lambda J(u_\lambda) + \frac{1}{2} \|u_\lambda - f\|_{L^2(I)}^2 = e(\lambda) \leq E_\lambda(f) = \lambda J(f)$ that

$$\|u_\lambda - f\|_{L^2(I)}^2 \leq 2\lambda(J(f) - J(u_\lambda)). \tag{31}$$

Consequently, $\|u_\lambda - f\|_{L^2(I)}^2 = O(\lambda)$ and we also notice that $J(u_\lambda) \leq J(f)$ for all $\lambda > 0$. But we can do slightly better than that. Since $u_\lambda \rightarrow f$ in L^2 as $\lambda \rightarrow 0+$, we have

$$J(f) \leq \liminf_{\lambda \rightarrow 0+} J(u_\lambda),$$

where the lower semi-continuity of the total variation J , cf. [2, Prop. 3.6], was used. Since $J(u_\lambda) \leq J(f)$, we also obtain an estimate from below: $\limsup_{\lambda \rightarrow 0+} J(u_\lambda) \leq J(f)$. We conclude that $\lim_{\lambda \rightarrow 0+} J(u_\lambda) = J(f)$. If this is used in (31), we find that $\|u_\lambda - f\|_{L^2(I)}^2 = o(\lambda)$ as $\lambda \rightarrow 0+$. \square

The next example shows that the convergence rate stated in the proposition is optimal in the sense that the exponent cannot be lowered.

Example 6 On the interval $I = (0, 1)$ and with $\alpha > 1$, let the data be given by $f(x) = \alpha x^{\alpha-1}$ for $x \in I$. Since f is monotone, $J(f) = \lim_{\epsilon \rightarrow 0+} (f(1 - \epsilon) - f(\epsilon)) = \alpha < \infty$, and hence, $f \in \text{BV}(I)$. Applying the taut string interpretation to the cumulative function is $F(x) = x^\alpha$ shows that the denoised signal u_λ is constant near the interval end points and coincides with f in between. In fact, near the left interval end point we have $u_\lambda(x) = \alpha(\lambda/(\alpha - 1))^{(\alpha-1)/\alpha}$ for $0 \leq x \leq (\lambda/(\alpha - 1))^{1/\alpha}$. It follows, by an easy computation, that for each choice of α there exists a positive constant $C = C(\alpha)$ such that $\|f - u_\lambda\|_{L^2(I)}^2 \geq C\lambda^{2-1/\alpha}$. This bounds the rate of convergence from below.

The following result, mentioned in by Burger and Osher [15, Sect. 5], shows that if the data f belong to the space of functions with bounded variation and satisfy an additional regularity condition, then the convergence rate for the limits $u_\lambda \rightarrow f$ and $J(u_\lambda) \rightarrow J(f)$ as $\lambda \rightarrow 0+$ can be improved (considerably) to $O(\lambda)$.

Proposition 7 *Suppose the data $f \in \text{BV}(I)$ satisfy $J(f) = \langle \xi'_0, f \rangle$ for some test function $\xi_0 \in K$, then we have the bounds*

$$\|f - u_\lambda\|_{L^2(I)} \leq 2\|\xi'_0\|_{L^2(I)} \cdot \lambda \tag{32a}$$

and

$$0 < J(f) - J(u_\lambda) \leq 2\|\xi'_0\|_{L^2(I)}^2 \cdot \lambda \tag{32b}$$

for any $\lambda > 0$.

The additional requirement on the data—the supremum in $J(f) = \sup_{\xi \in K} \langle \xi', f \rangle$ being attained for some ξ_0 —is an instance of the so-called *source condition*. The source condition for non-quadratic convex variational regularization of inverse problems was identified in [15] and used to derive convergence rates for the generalized Bregman distances. The authors point out that the above result, which they write in a slightly different way, may be proved along the same lines as their other estimates. Here, we provide the details:

Proof Clearly, $E_\lambda(u_\lambda) \leq E_\lambda(f)$ implies the inequality

$$\lambda J(u_\lambda) + \frac{1}{2} \|f - u_\lambda\|^2 \leq \lambda J(f).$$

Moreover, since $J(u_\lambda) \geq \langle \xi'_0, u_\lambda \rangle$, by the definition of the total variation, we find that the extra assumption on the data f gives yet an inequality:

$$J(f) - J(u_\lambda) \leq \langle \xi'_0, f - u_\lambda \rangle. \tag{33}$$

If these two estimates are combined, together with the Cauchy–Schwarz inequality, then we get

$$\|f - u_\lambda\|^2 \leq 2\lambda \langle \xi'_0, f - u_\lambda \rangle \leq 2\lambda \|\xi'_0\| \|f - u_\lambda\|,$$

which gives (32a). If we apply the Cauchy–Schwarz inequality once more, this time to the right-hand side of (33) and use (32a) then (32b) follows. \square

8 Proof and Applications of Theorem 2

We begin with the proof of the fundamental estimate on the derivative of the denoised signal:

Proof of Theorem 2 The estimate (4) is a consequence of an extension of the original Lewy–Stampacchia inequality [28] to bilateral obstacle problems. The bilateral obstacle problem, in the one-dimensional setting, is to minimize the energy $E(u) := \frac{1}{2} \int_a^b u'(x)^2 dx$ in (3) over the closed convex set $C = \{u \in H_0^1(I) : \phi(x) \leq u(x) \leq \psi(x) \text{ a.e. } I\}$. The obstacles are given by the functions $\phi, \psi \in H^1(I)$ which satisfy the conditions $\phi < \psi$ on I , and $\phi < 0 < \psi$ on $\partial I = \{a, b\}$. The latter ensures that C is non-empty.

Suppose ϕ' and ψ' are in $BV(I)$, such that ϕ'' and ψ'' are signed measures, then the solution u_0 of the minimization problem $\min_{u \in C} E(u)$ satisfies the following inequality (as measures)

$$-(\phi'')^- \leq u_0'' \leq (\psi'')^+. \tag{34}$$

Here, the notation μ^+ and μ^- is used to denote the positive and negative variations, respectively, of a signed measure

μ . The estimate in (34) is the extension of the Lewy–Stampacchia inequality. We prove this result in “Appendix B”. Our proof is based on an abstract argument, valid in much more general settings, given in [23].

The assumption of our theorem that $f \in BV(I)$ implies that $F'' = f'$ is a signed measure. If we apply (34) with $\phi = F - \lambda$ and $\psi = F + \lambda$, then we find that the taut string W_λ satisfies

$$-(F'')^- \leq W_\lambda'' \leq (F'')^+.$$

The estimate (4) follows if we substitute the identities $F' = f$ and $W_\lambda' = u_\lambda$ into the above inequality. \square

Having established Theorem 2, we are able to prove the following result about the strong convergence in $BV(I)$ of the ROF minimizer as the regularization parameter approaches zero.

Proposition 8 *If $f \in BV(I)$ then*

$$J(f - u_\lambda) = J(f) - J(u_\lambda). \tag{35}$$

In particular, both $J(f - u_\lambda)$ and $\|f - u_\lambda\|_{BV}$ tend to zero as $\lambda \rightarrow 0+$.

Proof The measures $(f')^+$ and $(f')^-$ are concentrated on disjoint measurable sets (Hahn decomposition, see [38, Sect. 6.14]), so Proposition 2 implies the pair of inequalities, $0 \leq (u'_\lambda)^+ \leq (f')^+$ and $0 \leq (u'_\lambda)^- \leq (f')^-$. A direct calculation, using the fact that $J(v) = (v')^+(I) + (v')^-(I)$ for any function $v \in BV(I)$, yields

$$\begin{aligned} J(f - u_\lambda) &= (f' - u'_\lambda)^+(I) + (f' - u'_\lambda)^-(I) \\ &= (f')^+(I) - (u'_\lambda)^+(I) + (f')^-(I) - (u'_\lambda)^-(I) \\ &= J(f) - J(u_\lambda), \end{aligned}$$

where the right-hand side tends to zero as $\lambda \rightarrow 0+$, by Proposition 6. \square

Theorem 2 also implies the first part of

Proposition 9 *Suppose f is a piecewise constant function on I . 1) The ROF minimizer u_λ is again piecewise constant for all $\lambda > 0$. 2) There exists a number $\bar{\lambda} > 0$ and a piecewise linear function $\bar{\xi} \in K$ such that $\xi_\lambda = \bar{\xi}$ for all $\lambda, 0 < \lambda \leq \bar{\lambda}$.*

Proof If f is piecewise constant, then there exists nodes $a = x_0 < x_1 < \dots < x_{N-1} < x_N = b$ which partitions the interval $I = (a, b]$ into N subintervals $I_i = (x_{i-1}, x_i]$ such that f equals the constant value $f_i \in \mathbf{R}$ on I_i for $i = 1, \dots, N$. The derivative of the signal becomes

$$f' = \sum_{i=1}^{N-1} b_i \delta_{x_i} \tag{36}$$

where $b_i = f_{i+1} - f_i$ for $i = 1, \dots, N - 1$ and δ_x denotes the Dirac measure supported at x . We may assume that $f_{i+1} \neq f_i$ and therefore $b_i \neq 0$ for all $i = 1, \dots, N - 1$. From (36), it follows that $J(f) = \sum_{i=1}^{N-1} |b_i| < \infty$ so f belongs to $BV(I)$ and Theorem 2 is applicable:

$$\begin{aligned} & \sum_{i=1}^{N-1} \min\{0, b_i\} \delta_{x_i} \\ &= -(f')^- \leq u'_\lambda \leq (f')^+ \\ &= \sum_{i=1}^{N-1} \max\{0, b_i\} \delta_{x_i}. \end{aligned}$$

This estimate shows that there exists numbers $c_i(\lambda)$ such that $u'_\lambda = \sum_{i=1}^{N-1} c_i(\lambda) \delta_{x_i}$ and that the $c_i(\lambda)$'s satisfy $0 \leq c_i(\lambda) \cdot b_i^{-1} \leq 1$ for $i = 1, \dots, N - 1$. Since the derivative is zero except at a finite set of points, we draw the conclusion that u_λ is piecewise constant function, as claimed. Observe that the set of nodes of u_λ is contained in the set of nodes of f . (The latter is the “edge-preserving” property of the one-dimensional ROF model.)

The proof of the second part of the proposition is suggested by the taut string interpretation of the ROF model: When the parameter λ is sufficiently close to zero, the taut string will, at each node of the signal f , meet either the upper obstacle $F + \lambda$ or the lower obstacle $F - \lambda$, and here. The idea of the proof is to guess $\bar{\xi}$ and set

$$\bar{u}_\lambda = f - \lambda \bar{\xi}', \tag{37}$$

as required by (19a), and then verify the identity (19b). Then Theorem 3 implies that $\bar{u}_\lambda = u_\lambda$ and $\bar{\xi} = \xi_\lambda$, as claimed.

As $\bar{\xi}$ we choose the continuous piecewise linear function which is zero at the end points of the interval I and satisfies

$$\bar{\xi}(x_i) = -\text{sign}(b_i), \quad i = 1, \dots, N - 1,$$

and, as mentioned above, let \bar{u}_λ be defined by (37). It is clear that $\|\bar{\xi}\|_\infty \leq 1$ so that $\bar{\xi} \in K$. Our task is now to verify that the pair $\bar{u}_\lambda, \bar{\xi}$ satisfies $J(\bar{u}_\lambda) = \langle \bar{u}_\lambda, \bar{\xi} \rangle$. First, notice that

$$\bar{u}'_\lambda = f' - \lambda \bar{\xi}'' = \sum_{i=1}^{N-1} (b_i - \lambda d_i) \delta_{x_i},$$

where we have used that $\bar{\xi}$ is piecewise linear such that $\bar{\xi}''$ exists and equals $\sum_{i=1}^{N-1} d_i \delta_{x_i}$ for some real numbers d_1, \dots, d_{N-1} (which could in principle be computed from the definition of $\bar{\xi}$ and knowledge of where f 's node is located.) It is now clear that we may find a positive number $\bar{\lambda}$ such that if $0 \leq \lambda < \bar{\lambda}$ then each of the numbers $b_i - \lambda d_i$, used to represent \bar{u}'_λ , has the same sign as b_i . (In fact $\bar{\lambda} = (\max_{1 \leq i \leq N-1} (d_i/b_i))^{-1}$ works). For any λ smaller than $\bar{\lambda}$, we have

$$\begin{aligned} \langle \bar{u}_\lambda, \bar{\xi}' \rangle &= \langle -\bar{u}'_\lambda, \bar{\xi} \rangle \\ &= \sum_{i=1}^{N-1} -(b_i - \lambda d_i) \langle \delta_{x_i}, \bar{\xi} \rangle \\ &= \sum_{i=1}^{N-1} (b_i - \lambda d_i) \text{sign}(b_i) \\ &= \sum_{i=1}^{N-1} |b_i - \lambda d_i| = J(\bar{u}_\lambda) \end{aligned}$$

as wanted, and the proof of the second part of the proposition is complete. \square

The proposition implies the strongest approximation rate imaginable:

Corollary 2 *If f is a piecewise constant function, then $\|f - u_\lambda\|_{L^2(I)} = O(\lambda)$ and $J(f) - J(u_\lambda) = O(\lambda)$ as $\lambda \rightarrow 0+$.*

Proof Clearly, $f \in BV(I)$ so the result will follow from the bounds in Proposition 7 provided $J(f) = \langle \xi'_0, f \rangle$ for some $\xi_0 \in K$. But this is a consequence of the result just proved: $\xi_\lambda = \bar{\xi}$ when λ is close to zero, hence

$$J(f) = \lim_{\lambda \rightarrow 0+} J(u_\lambda) = \lim_{\lambda \rightarrow 0+} \langle u_\lambda, \bar{\xi}' \rangle_{L^2(I)} = \langle f, \bar{\xi}' \rangle_{L^2(I)},$$

so we may take $\xi_0 = \bar{\xi}$. \square

9 ROF Denoising as a Semi-Group

We now turn our attention to ROF denoising as the map $f \mapsto u_\lambda$ and therefore need to adjust notation to incorporate the dependence of u_λ upon the signal f in a natural manner. Therefore, for each $\lambda \geq 0$, define a mapping $S_\lambda : L^2(I) \rightarrow L^2(I)$ by setting, for any $f \in L^2(I)$,

$$S_\lambda(f) = \begin{cases} u_\lambda & \text{if } \lambda > 0, \text{ and} \\ f & \text{when } \lambda = 0. \end{cases} \tag{38}$$

Then $\{S_\lambda\}_{\lambda \geq 0}$ is a family of nonlinear operators which we claim has the following properties:

1. $S_0 = \text{Id}$, the identity mapping on $L^2(I)$.
2. For any $f \in L^2(I)$, $[0, +\infty) \ni \lambda \mapsto S_\lambda(f) \in L^2(I)$ is continuous.
3. For any $\lambda \geq 0$, S_λ is non-expansive;

$$\|S_\lambda(f_2) - S_\lambda(f_1)\|_{L^2(I)} \leq \|f_2 - f_1\|_{L^2(I)}$$

for all $f_1, f_2 \in L^2(I)$.

4. $S_\lambda \circ S_\nu = S_{\lambda+\nu}$ for all $\lambda, \nu \geq 0$

Property 1) follows, of course, from the definition of S_λ and reflects our intuition associated with the definition of u_λ in (1). This intuition is confirmed by Proposition 6 which states that $u_\lambda \rightarrow f$ in L^2 as $\lambda \rightarrow 0+$. This observation together with part (b) of Proposition 2 implies the second property in the above list. The third property is simply part (a) of Proposition 2 rewritten in terms of S_λ . If the fourth and last property can be proved, then the family $\{S_\lambda\}_{\lambda \geq 0}$ satisfies the axioms of a semi-group of nonlinear operators on the Hilbert space $L^2(I)$, see Barbu [7, Ch. III.1, p.98]. The last property is the content of the following proposition:

Proposition 10 (Semi-group property) *For $f \in L^2(I)$, the formula*

$$S_\lambda(S_\nu(f)) = S_{\lambda+\nu}(f) \tag{39}$$

holds for all $\lambda, \nu \geq 0$.

A proof of the semi-group property can be found in [39]. However, the fundamental estimate in Theorem 2 and the characterization of the ROF minimizer in Theorem 3 allow us to present short and very direct proof of this result.

Proof The assertion holds trivially if either λ or ν is equal zero, so we may assume that $\lambda, \mu > 0$. The idea of the proof is to set

$$\bar{u} = S_\nu(u_\lambda) = S_\nu(S_\lambda(f))$$

and then show that there exists a dual variable $\bar{\xi} \in K$ such that

$$\begin{cases} \bar{u} = f - (\lambda + \mu)\bar{\xi}' & \text{and} \\ J(\bar{u}) = \langle \bar{u}, \bar{\xi}' \rangle. \end{cases}$$

For then, the characterization of the ROF minimizer in Theorem 3 implies that $\bar{u} = S_{\lambda+\nu}(f)$ and (39) holds.

Since u_λ and \bar{u} are both defined as ROF minimizers, they satisfy the conditions (19a) and (19b) of Theorem 3. That is, there exist uniquely determined functions ξ_λ and $\tilde{\xi}_\nu$ in K such that

$$\begin{cases} u_\lambda = f - \lambda\xi'_\lambda, \\ J(u_\lambda) = \langle u_\lambda, \xi'_\lambda \rangle, \end{cases} \quad \text{and} \quad \begin{cases} \bar{u} = u_\lambda - \nu\tilde{\xi}'_\nu, \\ J(\bar{u}) = \langle \bar{u}, \tilde{\xi}'_\nu \rangle. \end{cases}$$

Here, the tilde in $\tilde{\xi}_\nu$ signifies that we are dealing with the dual variable in the denoising of u_λ rather than of f .

If we set

$$\bar{\xi} = \frac{\lambda\xi_\lambda + \nu\tilde{\xi}_\nu}{\lambda + \nu},$$

then $\bar{\xi} \in K$ because it is the convex combination of two elements of K . Using the above characterizations of u_λ and \bar{u} , we find that

$$f - (\lambda + \nu)\bar{\xi}' = (f - \lambda\xi'_\lambda) - \nu\tilde{\xi}'_\nu = u_\lambda - \nu\tilde{\xi}'_\nu = \bar{u},$$

in other words \bar{u} and $\bar{\xi}$ fulfil (19a), by construction. It remains to verify that (19b) is holds as well. Since

$$\begin{aligned} \langle \bar{u}, \bar{\xi}' \rangle &= \frac{\lambda}{\lambda + \nu} \langle \bar{u}, \xi'_\lambda \rangle + \frac{\nu}{\lambda + \nu} \langle \bar{u}, \tilde{\xi}'_\nu \rangle \\ &= \frac{\lambda}{\lambda + \nu} \langle \bar{u}, \xi'_\lambda \rangle + \frac{\nu}{\lambda + \nu} J(\bar{u}), \end{aligned}$$

where we have used the characterization of \bar{u} stated above. We see that (19b) follows if it can be shown that $\langle \bar{u}, \xi'_\lambda \rangle = J(\bar{u})$. This follows from the identity in Proposition 8 applied with u_λ as in-signal:

$$J(\bar{u}) = J(u_\lambda) - J(u_\lambda - \bar{u}).$$

In fact, since $J(\bar{u}) \geq \langle \bar{u}, \xi' \rangle$ for all $\xi \in K$, this identity implies the inequality

$$\begin{aligned} J(\bar{u}) &\leq J(u_\lambda) - \langle u_\lambda - \bar{u}, \xi'_\lambda \rangle \\ &= J(u_\lambda) - J(u_\lambda) + \langle \bar{u}, \xi'_\lambda \rangle \\ &= \langle \bar{u}, \xi'_\lambda \rangle \end{aligned}$$

hence $J(\bar{u}) = \langle \bar{u}, \xi'_\lambda \rangle$ and the proof is complete. □

The last part of the proof yields

Corollary 3 *If $\lambda > 0$ then $J(u_\lambda) = \langle u_\lambda, \xi'_\lambda \rangle_{L^2(I)}$ for all $\nu, 0 < \nu \leq \lambda$.*

That is, the total variation of u_λ can be computed by taking inner product with any of the previous ξ_ν 's.

We now know that ROF denoising defines a family of nonlinear operators $\{S_\lambda\}_{\lambda \geq 0}$ which forms a contractive semi-group under composition. It is natural to seek the infinitesimal generator of this semi-group.

The infinitesimal generator, should it exist, is given by the right-hand derivative of $S_\lambda(f)$ at $\lambda = 0$:

$$\lim_{\lambda \rightarrow 0+} \frac{S_\lambda(f) - f}{\lambda} = \lim_{\lambda \rightarrow 0+} \frac{u_\lambda - f}{\lambda} = - \lim_{\lambda \rightarrow 0+} \xi'_\lambda,$$

where we have used (19a). If the limit $\lim_{\lambda \rightarrow 0+} \xi_\lambda := \xi_0$ exists in $H^1(I)$, then it follows that

$$\frac{d}{d\lambda} S_\lambda(f) \Big|_{\lambda=0} = -\xi'_0. \tag{40}$$

So in order to determine the infinitesimal generator of the semi-group, we first need to show that the limit $\xi_0 :=$

$\lim_{\lambda \rightarrow 0+} \xi_\lambda$ exists for sufficiently many $f \in \text{BV}(I)$ and then characterize ξ_0 in terms of f alone. A first step in this characterization is based on Lemma 3, where the notion of the subgradient of a convex function is used.

We recall the definition of the subdifferential of a convex functional, restricting ourselves to the Hilbert space case. Let H be a real Hilbert space and $\Phi : H \rightarrow (-\infty, \infty]$ a lower semi-continuous convex functional defined in H such that $\text{Dom } \Phi := \{x \in H : \Phi(x) < \infty\}$ is non-empty. Let $x_0 \in \text{Dom } \Phi$ and suppose there is a vector $y \in H$ such that the following inequality holds

$$\Phi(x) - \Phi(x_0) \geq \langle y, x - x_0 \rangle \quad \text{for all } x \in \text{Dom } \Phi.$$

Then y is called a *subgradient* of Φ at x_0 . The set of all such subgradients is called the *subdifferential* of Φ at x_0 and is denoted $\partial\Phi(x_0)$. The map $x \mapsto \partial\Phi(x)$ is a set-valued operator. It is possible that $\partial\Phi(x_0) = \emptyset$. The map $x \mapsto \partial\Phi(x)$ is *monotone* in the sense that if x, x_0 are points satisfying $\partial\Phi(x) \neq \emptyset$ and $\partial\Phi(x_0) \neq \emptyset$, then for any $\xi \in \partial\Phi(x)$ and $\xi_0 \in \partial\Phi(x_0)$ we have $\langle \xi - \xi_0, x - x_0 \rangle \geq 0$. This follows immediately from the definition of the subgradient.

If we take $H = L^2(I)$ and let $\Phi(u) = J(u)$ be the total variation of u , then $\text{Dom } J = \text{BV}(I)$ and we can characterize the subdifferential ∂J in the following manner:

Lemma 3 *Let $u_0 \in \text{Dom } J$ then $\eta \in \partial J(u_0)$ if and only if there exists $\xi_0 \in K$ such that $\eta = \xi'_0$ and $J(u_0) = \langle u_0, \xi'_0 \rangle$.*

The total variation J is considered as a function on $L^2(I)$ so $\eta \in L^2(I)$. Example 3 in Sect. 3 shows that there are cases where $u_0 \in \text{Dom } J$ but $\partial J(u_0) = \emptyset$. The lemma is the one-dimensional instance of a more general multidimensional result, see Alter et al. [1, Lemma 1, p. 335] as well as Bellettini et al. [8]. For completeness of exposition, we provide a proof:

Proof Assume first that the equality $J(u_0) = \langle u_0, \xi'_0 \rangle$ holds for some $\xi_0 \in H_0^1(I)$ and let $u \in \text{BV}(I)$. By the definition of the total variation, we have $J(u) \geq \langle u, \xi'_0 \rangle$. Subtracting the first identity from this inequality yields

$$J(u) - J(u_0) \geq \langle u - u_0, \xi'_0 \rangle, \tag{41}$$

which is precisely the condition for $\xi'_0 \in \partial J(u_0)$.

Conversely, suppose $\eta \in \partial J(u_0)$ such that the inequality (41) holds for all $u \in \text{BV}(I)$. If we set $\xi_0(x) = \int_a^x \eta(t) dt$ then $\xi_0 \in H^1$ with $\eta = \xi'_0$. Clearly, $\xi_0(a) = 0$ so to conclude that $\xi_0 \in H_0^1(I)$ we need to show that $\xi_0(b) = 0$. This, on the other hand, follows if we can show that $\langle \eta, 1 \rangle = 0$. By substituting $u = u_0 + 1$ into the definition of the subgradient of J at u_0 , we get

$$0 = J(u) - J(u_0) \geq \langle \eta, u - u_0 \rangle = \langle \eta, 1 \rangle.$$

Similarly, if $u = u_0 - 1$ we find the reverse inequality and may conclude that $\langle \eta, 1 \rangle = 0$. Consequently, $\xi_0 \in H_0^1(I)$.

Our next aim is to show that ξ_0 belongs to K . For any $v \in \text{BV}(I)$, we set $u = u_0 + v$. An application of the triangle inequality for J and the inequality (41) implies

$$J(v) \geq J(u_0 + v) - J(u_0) \geq \langle v, \xi'_0 \rangle.$$

Since this inequality holds for all $v \in \text{BV}(I)$, we conclude that $\|\xi_0\|_\infty \leq 1$, and hence, $\xi_0 \in K$. (In fact, any $v \in H^1$ belongs to BV with $J(v) = \|v'\|_{L^1}$ and may therefore be used in the above estimate;

$$\|v'\|_{L^1} = J(v) \geq \langle v, \xi'_0 \rangle = \langle -v', \xi_0 \rangle,$$

where integration by parts was used in the last step. This shows that $\mathbb{L}^2 \ni f \mapsto \langle f, \xi_0 \rangle \in \mathbf{R}$ extends by continuity to a bounded linear on functional L^1 . Riesz' theorem then implies that $\xi_0 \in L^\infty$ with $\|\xi_0\|_\infty \leq 1$.)

Now, if we take $u = (1/2)u_0$ in (41) and use the positive homogeneity of J then this inequality becomes

$$-\frac{1}{2}J(u_0) \geq -\frac{1}{2}\langle u_0, \xi'_0 \rangle$$

or

$$J(u_0) \leq \langle u_0, \xi'_0 \rangle,$$

which, in view of the definition of J , implies $J(u_0) = \langle u_0, \xi'_0 \rangle$, and the proof is complete. \square

In view of Lemma 3, the necessary and sufficient conditions (19a) and (18) in Theorem 3 can be reformulated as $\lambda^{-1}(f - u_\lambda) \in \partial J(u_\lambda)$, i.e. Fermat's rule for a minimum (for convex functions.) We note in passing that the equation $u + \partial J(u) \ni f$ has a unique solution (namely u_λ) for each $f \in L^2(I)$. Hence, ∂J is a maximal monotone (or $-\partial J$ is a maximal dissipative) set-valued operator on L^2 , cf. Barbu [7, p. 71].

Now, suppose that $f \in L^2(I)$ is such that the limit $\xi_\lambda \rightarrow \xi_0$ exists in $H_0^1(I)$ as $\lambda \rightarrow 0+$. Here, as usual, ξ_λ is the unique element in K associated with the ROF minimizer $u_\lambda = S_\lambda(f)$. Clearly, $\xi_0 \in K$. Since the identity

$$J(u_\lambda) = \langle u_\lambda, \xi'_\lambda \rangle$$

holds for all $\lambda > 0$ and $u_\lambda \rightarrow f$ in $L^2(I)$ as $\lambda \rightarrow 0+$ by Proposition 6, the lower semi-continuity of the total variation implies

$$J(f) \leq \lim_{\lambda \rightarrow 0+} \langle u_\lambda, \xi'_\lambda \rangle = \langle f, \xi'_0 \rangle.$$

We conclude from this that $f \in \text{Dom } J = \text{BV}(I)$ with $J(f) = \langle f, \xi'_0 \rangle$. Moreover, Lemma 3 implies that $\xi'_0 \in \partial J(f)$ so, by (40), we arrive at

$$-\frac{d}{d\lambda} S_\lambda(f) \Big|_{\lambda=0} = \xi'_0 \in \partial J(f).$$

Notice that the derivative $(d/d\lambda)S_\lambda(f)|_{\lambda=0}$ exists for all f in a dense subset of $L^2(I)$. To see this, we use that the limit $\lim_{\lambda \rightarrow 0+} \xi_\lambda := \xi_0$ exists whenever f is a piecewise constant function (by Proposition 9) and that the piecewise constant functions on I are dense in $L^2(I)$. We have proved (Cf. [7, Th. 1.2, p.175]):

Theorem 4 *The infinitesimal generator of the nonlinear contractive semi-group $\{S_\lambda\}_{\lambda \geq 0}$ is the nonlinear set-valued mapping $f \mapsto -\partial J(f)$.*

Using the formal notation introduced in Sect. 5, this result may be expressed by saying that $u(x, \lambda) := u_\lambda(x)$ solves the following Cauchy problem:

$$\partial_\lambda u = \partial_x \left(\frac{\partial_x u}{|\partial_x u|} \right) \text{ on } I \times [0, \infty) \text{ and } u(\cdot, 0) = f.$$

The nonlinear parabolic PDE above is known as the total variation flow. The total variation flow in higher dimensions has been extensively studied in Andreu et al. [3,4], Bellettini et al. [8] and Alter et al. [1]. They all start with the construction of the minimizing gradient flow associated with the total variation functional and then go on to derive its various properties including its relation to the ROF model. Here, we have started at the other end: using the nice properties of the one-dimensional case, it is proved that the ROF model gives rise to a semi-group of nonlinear contractive operators. We then proceed to derive its infinitesimal generator, which turns out to be the minus, the subdifferential of the total variation.

In view of the characterization of J 's subgradient given in Lemma 3, it follows from Example 3 that there exist $f \in L^2(I)$ such that $\partial J(f)$ consists of more than one element. If f is such a function and the derivative $-(d/d\lambda)S_\lambda(f)|_{\lambda=0}$ exists, then it is reasonable to ask which element of $\partial J(f)$ this derivative corresponds to. The answer is provided by the following result.

Proposition 11 *Suppose $f \in L^2(I)$ is such that the derivative $(d/d\lambda)S_\lambda(f)|_{\lambda=0} := -\xi'_0$ exists. Then ξ'_0 is the element in $\partial J(f)$ with the smallest L^2 -norm.*

Proof The assumption on f implies that the limit $\xi_\lambda \rightarrow \xi_0$ exists in $H^1_0(I)$ as $\lambda \rightarrow 0+$. For each $\lambda > 0$, ξ_λ is characterized as the unique member of K which solves

$$\max_{\xi \in K} \left\{ \|f\|^2 - \|f - \lambda \xi'\|^2 \right\}.$$

It follows that

$$2\langle f, \xi'_\lambda \rangle - \lambda \|\xi'_\lambda\|^2 \geq 2\langle f, \xi' \rangle - \lambda \|\xi'\|^2$$

for all $\xi \in K$. If we pick ξ in $\partial J(f)$, then $\langle f, \xi' \rangle = J(f)$, by the Lemma, and we find

$$\|\xi'\|^2 - \|\xi'_\lambda\|^2 \geq \frac{2}{\lambda} \left\{ J(f) - \langle f, \xi'_\lambda \rangle \right\} \geq 0,$$

which is the desired result. \square

A very detailed and well-written analysis of the one-dimensional total variation flow can be found in the paper by Bonforte and Figalli [11]. Here, the theory of the flow is developed as a limit of a time-discretized problem (the Crandall–Liggett approach) which leads them to study certain properties of the ROF functional, some of which are close to ours (e.g. Lemma 2.3 in [11] seems to contain the same insight as our Theorem 2.)

10 Application to Fused Lasso

In this section and the next, we briefly analyze two TV-minimization problems that are related to the ROF model. First out is a generalization of the ROF model obtained by adding a positive multiple of the L^1 -norm of the variable u to the ROF functional (14):

$$E_{\lambda, \mu}(u) = \lambda J(u) + \mu \|u\|_{L^1(I)} + \frac{1}{2} \|f - u\|_{L^2(I)}^2, \quad (42)$$

where $\lambda, \mu \geq 0$ are regularization parameters. Suppose f is a piecewise constant function on $I = (a, b]$ with equidistant nodes $a = x_0 < x_1 < \dots < x_{N-1} < x_N = b$ and with the constant value f_i on each subinterval $(x_{i-1}, x_i]$ for $i = 1, \dots, N$, i.e.

$$f = \sum_{i=1}^N f_i \chi_{(x_{i-1}, x_i]}.$$

Also, restrict minimization to the set of functions u which are piecewise constant with the same nodes as f and the constant value u_i on the i th subinterval; $u = \sum_{i=1}^N u_i \chi_{(x_{i-1}, x_i]}$. Substitution of such f and u into (42) leads to the minimizing the following discretized functional,

$$E_{\lambda, \mu}(u) = \lambda \sum_{i=1}^{N-1} |u_{i+1} - u_i| + \mu \sum_{i=1}^N |u_i| + \frac{1}{2} \sum_{i=1}^N |f_i - u_i|^2.$$

This minimization problem, known as the *fused lasso model* (in Lagrange form), was introduced in Tibshirani et al. [43]. The functional is strictly convex and has compact sub-level sets and therefore possesses a unique minimum $u^* = \sum_{i=1}^N u_i^* \chi_{(x_{i-1}, x_i]}$ which is called the *fused lasso signal approximator*. The idea of the fused lasso model is to simultaneously promote sparsity in u^* and its (discrete) derivative.

As shown by Friedman et al. [22], there is a close relationship between the fused lasso model and the discrete ROF model (i.e. $E_{\lambda,\mu}$ with $\mu = 0$): the fused lasso signal approximator u^* can be obtained from the discrete ROF minimizer $u_\lambda = \sum_{i=1}^N u_{\lambda,i} \chi_{(x_{i-1}, x_i]}$ by soft thresholding at level μ ,

$$u_i^* = \begin{cases} u_{\lambda,i} - \mu \operatorname{sign}(u_{\lambda,i}) & \text{for } |u_{\lambda,i}| > \mu \\ 0 & \text{otherwise,} \end{cases}$$

for $i = 1, \dots, N$.

The purpose of this section is to extend this result to the continuous fused lasso model (42). In the continuous case, the fused lasso signal approximator is defined by

$$u_{\lambda,\mu} = \arg \min_{u \in \operatorname{BV}(I)} E_{\lambda,\mu}(u).$$

and the claimed relation to the ROF model is:

Theorem 5 *The fused lasso signal approximator $u_{\lambda,\mu}$ can be computed, for all $\lambda, \mu \geq 0$, from the ROF minimizer u_λ by the formula*

$$u_{\lambda,\mu} = Sr_{[-\mu,\mu]} \circ u_\lambda, \tag{43}$$

where the function $Sr_{[-\mu,\mu]} : \mathbf{R} \rightarrow \mathbf{R}$ given by

$$Sr_{[-\mu,\mu]}(t) = t - \frac{\mu t}{\max(\mu, |t|)}, \quad t \in \mathbf{R}, \tag{44}$$

is the soft threshold map (or shrinkage map) at level μ (cf. Sect. 2).

Proof The total variation is again expressed in terms of a dual variable ξ as $J(u) = \sup_{\xi \in K} \langle u, \xi' \rangle$. We follow the same pattern and represent the L^1 -norm by $\|u\|_{L^1} = \sup_{\eta \in B} \langle u, \eta \rangle$ where $B := \{\eta \in L^\infty(I) : \|\eta\|_\infty \leq 1\}$ denotes the closed unit ball in L^∞ (which is a subset of $L^2(I)$ because I is bounded.) This formula holds because L^∞ is the dual of L^1 , cf. [38, Thm. 6.16]. It follows that (42) may be written as

$$E_{\lambda,\eta}(u) = \sup_{\eta \in B, \xi \in K} \left\{ \lambda \langle u, \xi' \rangle + \mu \langle u, \eta \rangle + \frac{1}{2} \|f - u\|^2 \right\}.$$

Now, the convex sets $K' := \{\zeta = \xi' : \xi \in K\}$ and B are both closed in $L^2(I)$. The same is true for the dilated sets $\lambda K'$ and μB and for their Minkowski sum

$$C := \lambda K' + \mu B.$$

We can therefore rewrite the fused lasso energy as

$$E_{\lambda,\eta}(u) = \sup_{\zeta \in C} \left\{ \langle u, \zeta \rangle + \frac{1}{2} \|f - u\|^2 \right\}.$$

By following the proof of Theorem 3 step by step, we can show that

$$\min E_{\lambda,\eta}(u) = \max_{\zeta \in C} \frac{1}{2} \left\{ \|f\|^2 - \|f - \zeta\|^2 \right\} \tag{45}$$

where equality is attained for a unique pair of function $u_* \in \operatorname{BV}(I)$ and $\zeta_* \in C$. Moreover, this pair is characterized by the condition that

$$u_* = f - \zeta_* \quad \text{and} \quad \langle f - \zeta_*, \zeta - \zeta_* \rangle \leq 0, \quad \forall \zeta \in C.$$

By definition, $u_* = u_{\lambda,\mu}$, the fused lasso estimator.

Now, the second member ζ_* may be written as a sum $\zeta_* = \lambda \xi_*' + \mu \eta_*$ for some $\xi_* \in K$ and $\eta_* \in B$. This decomposition of ζ_* is usually not unique, but the following argument holds for any such decomposition. The necessary and sufficient condition for optimality in (45) may be expressed as

$$\begin{cases} \langle f - \lambda \xi_*' - \mu \eta_*, \zeta - \lambda \xi_*' - \mu \eta_* \rangle \leq 0, \\ \text{for all } \zeta \in \lambda K' + \mu B. \end{cases} \tag{46}$$

This, in turn, can be split into a pair of conditions by choosing first $\zeta = \lambda \xi' + \mu \eta$ for $\xi \in K$ and then $\zeta = \lambda \xi_*' + \mu \eta_*$ in (46):

$$\langle f - \lambda \xi_*' - \mu \eta_*, \xi' - \xi_*' \rangle \leq 0, \quad \forall \xi \in K, \tag{47}$$

$$\langle f - \lambda \xi_*' - \mu \eta_*, \eta - \eta_* \rangle \leq 0, \quad \forall \eta \in B. \tag{48}$$

These conditions together are both necessary and sufficient for $\zeta_* = \lambda \xi_*' + \mu \eta_*$ to be the maximizer of the right-hand side of (45). By the projection theorem, (48) implies that $\mu \eta_* = Pr_{\mu B}(f - \lambda \xi_*')$. This means that the fused lasso solution can be written in terms of ξ_* as

$$\begin{aligned} u_{\lambda,\mu} &= f - \zeta_* \\ &= f - \lambda \xi_*' - \mu \eta_* \\ &= (f - \lambda \xi_*') - Pr_{\mu B}(f - \lambda \xi_*') \\ &= Sr_{[-\mu,\mu]}(f - \lambda \xi_*'). \end{aligned} \tag{49}$$

It remains to be proved that we can take ξ_* to be the optimal dual variable ξ_λ in the ROF model. This is a consequence of the following argument: From Sect. 2, we know that the projection $\mu \eta_*$ of $f - \lambda \xi_*'$ onto μB can be computed explicitly by the formula

$$\mu \eta_* = \frac{\mu (f - \lambda \xi_*')}{\max(\mu, |f - \lambda \xi_*'|)}. \tag{50}$$

If this expression is substituted into (47), it follows that ξ_* satisfies:

$$\langle Sr_{[-\mu,\mu]}(f - \lambda \xi_*'), \xi' - \xi_*' \rangle \leq 0 \quad \forall \xi \in K. \tag{51}$$

where $Sr_{[-\mu, \mu]}$ is the soft threshold (44). Now, let the function $H : \mathbf{R} \rightarrow \mathbf{R}$ be defined by

$$H(t) = \begin{cases} \frac{1}{2}(t - \mu)^2 & \text{for } t \geq \mu, \\ 0 & \text{when } -\mu \leq t \leq \mu \text{ and} \\ \frac{1}{2}(t + \mu)^2 & \text{for } t \leq -\mu. \end{cases} \quad (52)$$

This function is convex and differentiable, and it is easy to check that its derivative is $H'(t) = Sr_{[-\mu, \mu]}(t)$. Thus, (51) can be rewritten as

$$\langle H'(f - \lambda \xi'_*), \xi' - \xi'_* \rangle \leq 0 \quad \forall \xi \in K,$$

which is the necessary and sufficient condition for $\xi_* \in K$ to be a solution to the minimization problem

$$\inf_{\xi \in K} L_H(f - \lambda \xi') \quad \text{where } L_H(W) := \int_I H(W') dx. \quad (53)$$

In summary, for any decomposition of ζ_* into a sum $\lambda \xi'_* + \mu \eta_*$ where $\xi_* \in K$ and $\eta_* \in B$, the first function ξ_* solves (53) and the second one η_* can be computed from ξ_* using (50). Conversely, if we find a solution ξ_* of (53) and define η_* by the explicit formula in (50) then $\zeta_* = \lambda \xi'_* + \mu \eta_*$ maximizes the right-hand side in (45). Now, Lemma 1 implies that the ROF minimizer ξ_λ is a solution to (53). Since $u_\lambda = f - \lambda \xi'_\lambda$, the formula (43) follows immediately from (49). \square

11 Application to Isotonic Regression

As a second example, we briefly outline, mostly without proofs, how the theory developed for the ROF model can be modified in order to derive the so-called “lower convex envelope” interpretation of the isotonic regression estimator. Isotonic regression is a method from mathematical statistics

used for nonparametric estimation of probability distributions, see for instance, [5]. It is a least-squares problem with a monotonicity constraint: for $f \in L^2(I)$, determine a non-decreasing function $u_\uparrow \in L^2(I)$ which solves the minimization problem,

$$\min_{u \in L^2_\uparrow(I)} \frac{1}{2} \|u - f\|_{L^2(I)}^2, \quad (54)$$

where $L^2_\uparrow(I)$ denotes the set of all non-decreasing functions in $L^2(I)$. The “lower convex envelope” interpretation is shown for a piecewise constant signal f in Fig. 3.

The idea is to reformulate (54) as an unconstrained optimization problem by replacing the total variation term J of the ROF functional by regularization term J_\uparrow that can distinguish between functions that are non-decreasing or not. To achieve this, we set

$$K_+ = \{\xi \in H^1_0(I) : \xi(x) \geq 0 \text{ for all } x \in I\}$$

and define the functional

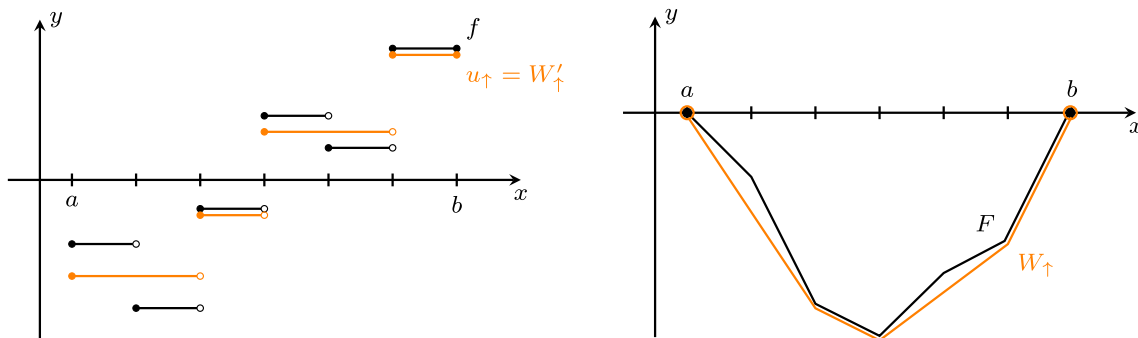
$$J_\uparrow(u) = \sup_{\xi \in K_+} \langle u, \xi' \rangle_{L^2(I)}.$$

It can be shown that

$$J_\uparrow(u) = \begin{cases} 0 & \text{if } u \in L^2_\uparrow(I), \\ +\infty & \text{otherwise.} \end{cases}$$

In the terminology of convex analysis, J_\uparrow is the indicator function of the closed convex cone $L^2_\uparrow(I)$. It follows that the isotonic regression problem (54) is equivalent to finding the minimizer u_\uparrow in $L^2(I)$ of the functional

$$E_\uparrow(u) := J_\uparrow(u) + \frac{1}{2} \|u - f\|_{L^2(I)}^2. \quad (55)$$



(a) The piecewise constant input signal f and the monotonic solution u_\uparrow to the isotonic regression problem. (b) The cumulative signal F and the corresponding lower convex envelope (or taut string) W_\uparrow .

Fig. 3 Graphical illustrations of the taut string interpretation of isotonic regression

Notice that, there is no need for a positive weight (such as the λ in the ROF functional) in this functional because the regularizer assumes only the values zero and infinity.

Again, we may assume that f mean value equal to zero so that the cumulative function $F(x) := \int_a^x f(t) dt$ belongs to $H_0^1(I)$. This will be used below.

Mimicking the proof of Theorem 3, we get:

$$\min_{u \in L^2(I)} E_{\uparrow}(u) = \max_{\xi \in K_+} \frac{1}{2} \left\{ \|f\|^2 - \|f - \xi'\|_{L^2(I)}^2 \right\}, \quad (56)$$

where the minimum on the left-hand side is attained by a unique function $u_{\uparrow} \in L^2_{\uparrow}(I)$ and the maximum on the right-hand side by a unique $\xi_{\uparrow} \in K_+$. These pair of functions satisfy

$$u_{\uparrow} = f - \xi'_{\uparrow} \quad \text{and} \quad \langle u_{\uparrow}, \xi'_{\uparrow} \rangle = 0. \quad (57)$$

These conditions are also sufficient for a pair of functions, $u_{\uparrow} \in L^2_{\uparrow}(I)$ and $\xi_{\uparrow} \in K_+$ to be the optimizers in (56).

Notice that if the two conditions of (57) are combined, the solution to the isotonic regression problem (54) can be characterized by the conditions $u_{\uparrow} \in L^2_{\uparrow}(I)$ and $f - u_{\uparrow} \in K'_+ := \{\xi' : \xi \in K_+\}$ and $\langle u_{\uparrow}, f - u_{\uparrow} \rangle = 0$. Thus, K'_+ is the dual cone of $L^2_{\uparrow}(I)$ and the pair $u_{\uparrow}, f - u_{\uparrow}$ is the Moreau decomposition of f .

Example 7 For $f(x) = 1 - |x|$, $-1 < x < 1$, the solution of the isotonic regression problem is

$$u_{\uparrow}(x) = \begin{cases} f(x) & \text{for } -1 < x < \alpha - 1, \\ \alpha & \text{for } \alpha - 1 \leq x < 1, \end{cases}$$

where $\alpha = 2 - \sqrt{2}$. To verify this, observe that u_{\uparrow} is clearly in $L^2_{\uparrow}(-1, 1)$ and if we define $\xi'_{\uparrow} = f - u_{\uparrow}$, then

$$\langle u_{\uparrow}, \xi'_{\uparrow} \rangle = \int_{-1}^{\alpha-1} 0 dx + \int_{\alpha-1}^1 \alpha(f(x) - \alpha) dx = 0,$$

so that the conditions (57) are both fulfilled. Since it is easy to verify that $\xi_{\uparrow}(x) := \int_{-1}^x \xi'_{\uparrow}(t) dt \geq 0$ for all $-1 < x < 1$, so that $\xi_{\uparrow} \in K_+$, it follows from the above characterization that u_{\uparrow} is indeed the isotonic regressor associated with f .

Now, to prove the lower envelope interpretation we introduce the new variable $W = F - \xi$, where $\xi \in K_+$, and set

$$T = \{W \in H_0^1(I) : W(x) \leq F(x), x \in I\}.$$

Maximization of the right-hand side of (55) is seen to be equivalent to solving the obstacle problem

$$\min_{W \in T} \frac{1}{2} \|W'\|_{L^2(I)}^2,$$

which admits a unique solution W_{\uparrow} by the projection theorem. It follows that (55) also has the unique solution $u_{\uparrow} = W'_{\uparrow}$ (distributional derivative) which belongs to $L^2_{\uparrow}(I)$ because $E_{\uparrow}(u_{\uparrow})$ is finite and therefore zero.

The solution W_{\uparrow} of the obstacle problem satisfies $W''_{\uparrow} \geq 0$ (this is the “easy” part of the original Lewy–Stampacchia inequality, $0 \leq W''_{\uparrow} \leq (F'')^+$) and is therefore automatically a convex function. In fact, by optimality, W_{\uparrow} is the maximal convex function lying below F , i.e. it is the *lower convex envelope* of F . Similar problems are considered in the multidimensional case, using higher-order methods (the space of functions with bounded Hessians), in Hinterberger and Scherzer [25].

12 Higher-Order Total Variation Regularization

In this section, we briefly consider the analogue of the ROF model and its taut string interpretation for denoising using higher-order total variation regularization. The one-dimensional n th order ROF model is defined as minimization of the functional

$$E_{\lambda}^n(u) := \lambda \int_a^b |u^{(n)}(x)| dx + \frac{1}{2} \int_a^b (f(x) - u(x))^2 dx,$$

where $u^{(n)}$ denotes the n th derivative of $u \in C^n(a, b)$ and $\lambda > 0$ is the regularization parameter. Here, we shall treat only the case $n = 2$, i.e. the second-order ROF model. In the multidimensional setting, second- and higher-order regularizations have been considered early on by Pöschl and Scherzer [36] as well as by Breides et al. [12]. They have subsequently found applications in restoration of MRI [27] and image inpainting [35], to mention just two examples. A detailed account of the second-order regularization in image restoration, as well as additional references, can be found in Begrounioux [10]. The one-dimensional case has been studied in a purely discrete setting in Steidl et al. [41], and one-dimensional examples are found in [34,36], among others.

We first give a proper definition of the second-order total variation, J_2 , which will replace the formal expression $\int_a^b |u''| dx$ used in the definition of the second-order ROF functional $E_{\lambda}^2(u)$. For $u \in L^1(I)$, where $I = (a, b)$, we define

$$J_2(u) = \sup \int_a^b u \xi'' dx$$

with the supremum being taken over all $\xi \in C_c^2(I)$ which satisfies $|\xi(x)| \leq 1$ for all $x \in I$. The set of functions for which J_2 has a finite value is denoted $BV_2(I)$. It is clear from this definition that if $u \in H^1(I)$ then $J_2(u) = J(u')$;

just apply integration by parts in the above integral and recall the definition of J . It can be shown that any $u \in BV_2(I)$ is automatically a member of $H^1(I)$ (in fact $u \in W^{1,\infty}(I)$ holds). In particular, $u \in L^2(I)$ and therefore the definition of the second-order total variation J_2 may be rephrased as

$$J_2(u) = \sup_{\xi \in K_2} \langle u, \xi'' \rangle_{L^2(I)},$$

where $K_2 := \{\xi \in H_0^2(I) : \|\xi\|_\infty \leq 1\}$. Recall that $H_0^2(I)$ is the closure of $C_c^2(I)$ in the Hilbert space $H^2(I)$, in fact

$$H_0^2(I) = \{\xi \in H^2(I) : \xi = \xi' = 0 \text{ at } \partial I = \{a, b\}\},$$

see Brezis [13, p. 134, Remarque 18]. Moreover, the map $u \mapsto \|u''\|_{L^2(I)}$ defines a Hilbert space norm on $H_0^2(I)$, i.e. the norm induced by the bilinear form $\langle u'', v'' \rangle_{L^2(I)}$, which defines an inner product on $H_0^2(I)$. The precise definition of the second-order ROF functional is

$$E_\lambda^2(u) := \lambda J_2(u) + \frac{1}{2} \|f - u\|_{L^2(I)}^2.$$

The function $u_\lambda : I \rightarrow \mathbf{R}$ given by

$$u_\lambda := \arg \min_{u \in BV_2(I)} E_\lambda^2(u),$$

is the denoising of $f \in L^2(I)$ using the second-order total variation regularization with weight λ .

The function u_λ , if it exists, is the solution of the unconstrained minimization problem for E_λ^2 over $BV_2(I)$. By considering variations of the form $u = u_\lambda + a_0 + a_1x$, where $a_0, a_1 \in \mathbf{R}$, then the optimality of u_λ implies that $\partial E_\lambda^2 / \partial a_0 = \partial E_\lambda^2 / \partial a_1 = 0$. Since $J_2(u + a_0 + a_1x) = J_2(u)$ for all $u \in BV_2(I)$, this leads to the two conditions: $\int_a^b u_\lambda = \int_a^b f$ and $\int_a^b x u_\lambda = \int_a^b x f$. We may therefore assume, without loss of generality, that the data f satisfies

$$\int_a^b f(x) dx = 0 \quad \text{and} \quad \int_a^b x f(x) dx = 0.$$

This assumption is imposed throughout the rest of this section, unless otherwise stated.

Using the methods introduced in the study of the (ordinary) ROF model, it is possible to prove the following result:

Theorem 6 *We have the identity*

$$\min_{u \in BV_2(I)} E_\lambda^2(u) = \max_{\xi \in K_2} \frac{1}{2} \left(\|f\|_{L^2(I)}^2 - \|f - \lambda \xi''\|_{L^2(I)}^2 \right), \tag{58}$$

with the minimum achieved by a unique $u_\lambda \in BV_2(I)$ and the maximum by a unique $\xi_\lambda \in K_2$, the two functions are related by

$$u_\lambda = f - \lambda \xi_\lambda'', \tag{59a}$$

and

$$J_2(u_\lambda) = \langle u_\lambda, \xi_\lambda'' \rangle_{L^2(I)}. \tag{59b}$$

Moreover, if $u_\lambda \neq 0$, then $\|\xi_\lambda\|_\infty = 1$. Conversely, the conditions (59a) and (59b) are sufficient for u_λ to be a solution of the second-order ROF model.

It follows from this theorem that the minimizer u_λ of the second-order ROF functional exists and can be obtained by finding the solution ξ_λ of

$$\min_{\xi \in K_2} \frac{1}{2} \|f - \lambda \xi''\|_{L^2(I)}^2$$

and the set $u_\lambda = f - \lambda \xi_\lambda''$.

Now, let us define $F'(x) = \int_a^x f(t) dt$ and $F(x) = \int_a^x F'(t) dt$ such that $F'' = f$. Then $F(a) = F'(a) = 0$ by construction and $F(b) = F'(b) = 0$ by our assumption of the signal f . Since $f \in L^2(I)$, we see that $F \in H_0^2(I)$. If we introduce the new variable $W = F - \lambda \xi$, where $\xi \in K_2$, then the above minimization problem becomes:

$$u_\lambda = W_\lambda'' \text{ where } W_\lambda = \arg \min_{W \in T_\lambda^2} \frac{1}{2} \|W''\|_{L^2(I)}^2.$$

Here, the set of admissible functions

$$T_\lambda^2 = \{W \in H_0^2(I) : F - \lambda \leq W \leq F + \lambda\}$$

is the second-order λ -tube. The condition on $W := F - \lambda \xi$ follows from the condition $\|\xi\|_\infty \leq 1$ on the functions ξ in K_2 .

The above procedure for finding u_λ has a mechanical interpretation: The second-order regularized denoising u_λ of f is the second (weak) derivative of the function W_λ which in turn corresponds to the energy minimizing shape of an ideal elastic beam (a cubic spline), clamped at the end points of I and forced to lie between a pair of parallel walls at a uniform distance λ from the graph of the bi-cumulative signal F . Denoising using n th order total variation regularization may be analyzed in a similar manner, but for $n > 2$ there is no obvious mechanical interpretation.

This ‘‘restricted spline’’ interpretation of the second-order ROF model can be used to ‘‘guess’’ the analytical solution of the second-order ROF model for simple f . The following example is probably the simplest imaginable non-trivial example.

Example 8 On the interval $I = (-1, 1)$, let the signal $f : I \rightarrow \mathbf{R}$ be given by $f(x) = |x| - \frac{1}{2}$. Notice that $\int_I f = 0$ and $\int_I x f = 0$. We can use the necessary and sufficient condition formulated in Theorem 6 to prove that the restored signal is

$$u_\lambda = (1 - 12\lambda)_+ f, \tag{60}$$

for any $\lambda > 0$. In particular it follows that $u_\lambda = 0$ if $\lambda \geq 1/12$ and that $\|f - u_\lambda\|_{BV_2(I)} := J_2(f - u_\lambda) + \|f - u_\lambda\|_{L^1(I)} = O(\lambda)$ as $\lambda \rightarrow 0+$. The proof runs as follows: If (60) is substituted into (59a), we obtain

$$\xi_\lambda'' = \min(12, \lambda^{-1})f.$$

Since the signal f satisfies $\int_I f = \int_I xf = 0$ it follows that $\xi_\lambda = \min(12, \lambda^{-1})F \in H_0^2(I)$. Moreover, $\|\xi_\lambda\|_\infty \leq 1$ with equality if and only if $0 \leq \lambda \leq 1/12$. It is now easy to verify that (59b) holds:

$$\begin{aligned} \langle \xi_\lambda'', u_\lambda \rangle &= \langle \min(12, \lambda^{-1})f, (1 - 12\lambda)_+ f \rangle \\ &= 12(1 - 12\lambda)_+ \|f\|_{L^2(I)}^2 \\ &= 2(1 - 12\lambda)_+ = J_2(u_\lambda), \end{aligned}$$

the last equality being a consequence of the fact that $J_2(f) = J_2(f') = 2$. This completes the verification.

The above explicit example was first found by Papafitsoros and Breides [34], see their Fig. 8d, but the derivation presented here is considerably shorter. Their paper focuses on a regularization term which is a kind of weighted combination of the first- and second-order total variation and contains only the latter as a special case. The example may be viewed as the second-order analogue to Example 4 where which considered denoising of a simple piecewise constant signal in the ROF model; in both cases, the restored signal is obtained by a simple scaling of the data.

Theorem 6 and its “restricted spline” interpretation make it clear that certain results for the (ordinary) ROF model carries over to the second-order case. For instance, parts (a) and (b) of Proposition 4 still hold: $u_\lambda = 0$ if and only if $\|F\|_\infty < \lambda$ and that $\|W_\lambda - F\|_\infty = \lambda$ also holds in that case, where F is the bi-cumulative function and W_λ the optimal spline.

13 Concluding Remarks

We have developed the theory for the one-dimensional ROF model for a quite general class of signals and given a thorough investigation of the properties of its solutions. This includes a useful fundamental estimate, Theorem 2, on the denoised signal. The theory may find practical applications in signal processing and image analysis alike. Indeed, by using the fundamental estimate, we saw how application of the ROF model to a piecewise constant signal leads to a piecewise constant denoised signal with the same nodes, i.e. the model is “edge-preserving” (Proposition 9). Our theory can be modified to cover the one-dimensional ROF model defined on the real line \mathbf{R} or the half-line \mathbf{R}_+ . To do this, the Sobolev spaces

have to be replaced by certain Beppo Levi spaces. The theory can also be modified to handle regularization with a weighted total variation term, something like

$$J_w(u) := \int_{\mathbf{R}} w(x)|u'(x)| dx,$$

where $w(x) > 0$. These two extensions would allow us to analyze the higher-dimensional ROF functional for spherically symmetric signals. The theory for the N -dimensional ROF model can be developed along almost the same lines, it seems, with exception for the uniqueness of the dual variables and the fundamental estimate; when $N \geq 2$ the dual variables are vector fields $\xi = (\xi_1, \dots, \xi_N)$ whose magnitudes are bounded by one. As is well known, only the divergence of this vector field is uniquely determined, not the vector field itself. The natural generalization of the fundamental estimate does not seem to hold for $N \geq 2$. The reason why it fails is that if f is the characteristic function of the unit square in the plane, then the denoising u_λ has level curves which looks like squares with the corners rounded of, see [18, Sect. 2.2.3]. This means that the support of the gradient of denoised signal is not contained in the support of the gradient of the original signal (rather, the support shifts inward, inside the square), and this is incompatible with a bound like the fundamental estimate. It would be interesting to know if there exists some alternative estimate on the denoised signal which could replace the fundamental estimate in higher dimensions.

Acknowledgements Open access funding provided by Lund University. The author wishes to thank Viktor Larsson, now at the Department of Computer Science at ETH Zürich, for reading and commenting on early drafts of this paper.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Appendix A

As promised in the introduction we now give a proof of Lemma 1. The idea of the proof is to verify that the function

$$W_\lambda := \arg \min_{W \in T_\lambda} E(W),$$

i.e. the “stretched rubber band”, also solves the variational inequality:

$$\int_I h(W'_\lambda)(W' - W'_\lambda) dx \geq 0, \text{ for all } W \in T_\lambda, \tag{61}$$

where $h = H'$. This condition is both necessary and sufficient for W_λ to be a minimizer of L_H over T_λ . If H happens to be strictly convex, then the condition also ensures uniqueness of the minimizer.

Being the minimizer of E over T_λ , $W_\lambda \in T_\lambda$ satisfies the following variational inequality (which is a special case of (61) with $H(s) = s^2/2$):

$$\int_I W'_\lambda(W' - W'_\lambda) dx \geq 0, \text{ for all } W \in T_\lambda. \tag{62}$$

Set $C_+ = \{x \in I : W_\lambda(x) = F(x) + \lambda\}$ and $C_- = \{x \in I : W_\lambda(x) = F(x) - \lambda\}$. These are the sets where the graph of W_λ touches the upper and the lower obstacles, respectively. Since F and W_λ are continuous functions, both sets are closed. In fact, C_+ and C_- are compact because $\lambda > 0$ implies that they do not reach the boundary of I . Clearly $C_+ \cap C_- = \emptyset$ and the union, $C = C_+ \cup C_-$, is the (so-called) contact set for W_λ .

For any non-negative function $\xi \in C_0^1(I)$ with $\text{supp } \xi \subset I \setminus C_-$, there exists a number $\epsilon > 0$ such that $W := W_\lambda - \epsilon\xi$ belongs to T_λ . If this W is substituted into (62) we find the condition

$$-\int_I W'_\lambda \xi' dx \geq 0, \quad \forall 0 \leq \xi \in C_0^1(I), \text{ supp } \xi \subset I \setminus C_-.$$

It follows from Riesz' lemma that W''_λ is a positive measure on $I \setminus C_-$, hence W'_λ is non-decreasing on each connected component of $I \setminus C_-$ (but not necessarily on the set as a whole). Similarly one proves that W'_λ is non-increasing on each connected component of $I \setminus C_+$. This means, in particular, that W'_λ is constant on each connected component of $(I \setminus C_-) \cap (I \setminus C_+) = I \setminus C$, i.e. away from the contact set.

Since H is convex, h is non-decreasing and so it follows that the composite function $h(W'_\lambda)$ has the same monotonicity properties as W'_λ . Therefore, the distributional derivative $h(W'_\lambda)'$ is a positive measure μ^+ on $I \setminus C_-$ and minus a positive measure, $-\mu^-$, on $I \setminus C_+$. Clearly $\text{supp } \mu^+ \subset C_+$ and $\text{supp } \mu^- \subset C_-$, so $h(W'_\lambda)'$ is a signed measure μ with the Jordan decomposition $\mu = \mu^+ - \mu^-$. The following calculation now verifies (61): For any $W \in T_\lambda$ we have

$$\begin{aligned} \int_I h(W'_\lambda)(W' - W'_\lambda) dx &= - \int_I W - W_\lambda d\mu \\ &= - \int_I W - W_\lambda d\mu^+ + \int_I W - W_\lambda d\mu^- \geq 0 \end{aligned}$$

which holds because $W - W_\lambda \leq 0$ on $\text{supp } \mu^+ \subset C_+$ and $W - W_\lambda \geq 0$ on $\text{supp } \mu^- \subset C_-$. This completes the proof.

Appendix B

We prove the inequality (34) used in the proof of Theorem 2. With the notation introduced in this proof we can formulate this result as follows:

Proposition 12 (Lewy–Stampacchia inequality) *Suppose ϕ'' and ψ'' are signed measures. Then the minimizer u_0 of E over $C = \{u : \phi \leq u \leq \psi\}$ satisfies*

$$-(\phi'')^- \leq u''_0 \leq (\psi'')^+$$

where μ^+ and μ^- denote the positive and negative variations, respectively, of the signed measure μ .

Note that the functional E is defined, convex and differentiable on $H^1(I)$ and therefore satisfies the inequality

$$E(v) - E(u) \geq \int_a^b u'(x)(v'(x) - u'(x)) dx, \tag{63}$$

for all $u, v \in H^1(I)$, as is easily checked. We also know that $\min_C E$ has a unique solution $u_0 \in C$ (use the projection theorem) which satisfies the necessary and sufficient condition:

$$\int_a^b u'_0(v' - u'_0) dx \geq 0 \text{ for all } v \in C.$$

Proposition 12 was first proved for the unilateral obstacle problem (in multiple dimensions) in [28] and since extended to bilateral obstacle problems. Here we present a proof based on the *sub-modularity* of the functional E , that is, for all $u, v \in H^1(I)$,

$$E(u \wedge v) + E(u \vee v) \leq E(u) + E(v), \tag{64}$$

where $u \wedge v := \max(u, v)$ and $u \vee v := \min(u, v)$ both belong to $H^1(I)$. In fact, the functional E is so simple that equality holds for all u, v . This method of proof was invented recently by [23] and used to prove a very general version of the bilateral Lewy–Stampacchia inequality. We use their approach.

Proof We prove the rightmost inequality, $u''_0 \leq (\psi'')^+$, the leftmost one then follows by a symmetry argument: $-u_0$ minimizes E over the $-C = \{u \in H_0^1(I) : -\psi \leq u \leq -\phi\}$.

To simplify notation we set $\psi'' = \mu$. Define the new functional $\hat{E}(u) = E(u) + \langle \mu^+, u \rangle$ and consider the minimization problem

$$\min_{u \in H_0^1(I) : u \geq u_0} \hat{E}(u). \tag{65}$$

The goal is to prove that u_0 itself solves this problem. This will imply the desired inequality, as we shall see below.

First the following claim is proved: For any $u \in H_0^1(I)$ satisfying $u \geq u_0$ we have

$$\hat{E}(u \wedge \psi) \leq \hat{E}(u). \tag{66}$$

Since $\psi > 0$ on ∂I and $u \in H_0^1(I)$ we get $u \wedge \psi \in H_0^1(I)$, so $u \wedge \psi$ is admissible for the min-problem above. The claim is proved using the sub-modularity (64) of E and the identity $u \wedge v + u \vee v = u + v$ with v replaced by ψ :

$$\begin{aligned} \hat{E}(u) - \hat{E}(u \wedge \psi) &= E(u) - E(u \wedge \psi) + \langle \mu^+, u - u \wedge \psi \rangle \\ &\geq E(u \vee \psi) - E(\psi) + \langle \mu^+, u \vee \psi - \psi \rangle \\ &\geq E(u \vee \psi) - E(\psi) + \langle \psi'', u \vee \psi - \psi \rangle \\ &\geq E(u \vee \psi) - E(\psi) - \langle \psi', (u \vee \psi)' - \psi' \rangle \geq 0, \end{aligned}$$

where the last inequality follows from (63).

The claim (66) shows that the minimum of \hat{E} over $\{u \geq u_0\}$ coincides with the minimum over $\{\psi \geq u \geq u_0\}$. Therefore, since $u - u_0 \geq 0$, we find

$$\begin{aligned} \hat{E}(u) - \hat{E}(u_0) &= E(u) - E(u_0) + \langle \mu^+, u - u_0 \rangle \\ &\geq E(u) - E(u_0) \geq 0, \end{aligned}$$

where the last estimate follows from the observation that $\{u : u_0 \leq u \leq \psi\} \subset C$ and that u_0 minimizes E over C . That is, u_0 is the solution of (65).

The necessary condition for u_0 to be a minimizer for \hat{E} reads:

$$\frac{d}{d\alpha} \hat{E}((1 - \alpha)u_0 + \alpha u) \Big|_{\alpha=0} \geq 0$$

for all $u \in H_0^1(I)$ satisfying $u \geq u_0$. That is,

$$\int_a^b u'_0(u' - u'_0) dx + \langle \mu^+, u - u_0 \rangle \geq 0$$

for all $u \in H_0^1(I)$ satisfying $u \geq u_0$. This implies $\langle -u''_0 + \mu^+, \varphi \rangle \geq 0$ for all $\varphi \in C_0^1(I)$ such that $\varphi \geq 0$ on I . Therefore, $-u''_0 + \mu^+$ is a positive measure, hence $u''_0 \leq \mu^+$, which is the desired result. \square

References

1. Alter, F., Caselles, V., Chambolle, A.: A characterization of convex calibrable sets in \mathbf{R}^N . *Math. Ann.* **332**, 329–366 (2005)
2. Ambrosio, L., Fusco, N., Pallara, D.: *Functions of Bounded Variation and Free Discontinuity Problems*. Clarendon Press, Oxford (2000)
3. Andreu, F., Ballester, C., Caselles, V., Mazón, J.M.: Minimizing total variation flow. *Differ. Integral Equ.* **14**(3), 321–360 (2001)

4. Andreu, F., Caselles, V., Diaz, J.I., Mazón, J.M.: Qualitative properties of the total variation flow. *J. Funct. Anal.* **188**(2), 516–547 (2002)
5. Anevski, D., Soulier, P.: Monotone spectral density estimation. *Ann. Stat.* **39**, 418–438 (2011)
6. Attouch, H., Buttazzo, G., Michaille, G.: *Variational Analysis in Sobolev and BV Spaces; Applications to PDEs and Optimization*, 2nd edn. SIAM, New York (2015)
7. Barbu, V.: *Nonlinear Semigroups and Differential Equations in Banach Spaces*. Nordhoff International Publishing, Leyden (1976)
8. Bellettini, G., Caselles, V., Novaga, M.: The total variation flow in \mathbf{R}^N . *J. Differ. Geom.* **144**, 475–525 (2002)
9. Bellettini, G., Caselles, V., Novaga, M.: Explicit solution of the eigenvalue problem $-\text{div}(du/|du|) = u$ in \mathbf{R}^2 . *SIAM J. Math. Anal.* **4**, 1095–1129 (2005)
10. Bergounioux, M.: Second order variational models for image texture analysis. *Adv. Imaging Electron Phys.* **181**, 35–124 (2014). <https://doi.org/10.1016/B978-0-12-800091-5.00002-1>
11. Bonforte, M., Figalli, A.: Total variation flow and sign fast diffusion in one dimension. *J. Differ. Equ.* **252**, 4455–4480 (2012). <https://doi.org/10.1016/j.jde.2012.01.003>
12. Bredies, K., Kunisch, K., Pock, T.: Total generalized variation. *SIAM J. Imaging Sci.* **3**(3), 492–526 (2010). <https://doi.org/10.1137/090769521>
13. Brézis, H.: *Analyse fonctionnelle—Théorie et applications*. Dunod, Paris (1999)
14. Briani, A., Chambolle, A., Novaga, M., Orlandi, G.: On the graient flow of a one-homogeneous functional. *Conflu. Math.* **3**(4), 617–635 (2011). <https://doi.org/10.1016/j.jde.2012.01.003>
15. Burger, M., Osher, S.: Convergence rates of convex variational regularization. *Inverse Probab.* **20**, 1411–1421 (2004)
16. Caselles, V., Chambolle, A., Novaga, M.: The discontinuity set of solutions of the TV denoising problem and some extensions. *Multiscale Model. Simul.* **6**(3), 879–894 (2007)
17. Chambolle, A.: An algorithm for total variation minimization and applications. *J. Math. Imaging Vis.* **20**, 89–97 (2004)
18. Chambolle, A., Caselles, V., Cremers, D., Novaga, M., Pock, T.: An introduction to total variation for image analysis. *Theor. Found. Numer. Methods Sparse Recovery* **9**, 263–340 (2010)
19. Chambolle, A., Lions, P.-L.: Image recovery via total variation minimization and related problems. *Numer. Math.* **76**, 167–188 (1997)
20. Davies, P., Kovac, A.: Local extremes, runs, strings and multiresolution. *Ann. Stat.* **29**, 1–65 (2001)
21. Dümbgen, L., Kovac, A.: Extension of smoothing via taut strings. *Electron. J. Stat.* **3**, 41–75 (2009)
22. Friedman, J., Hastie, T., Höfling, H., Tibshirani, R.: Sparsity and smoothness via the fused lasso. *Ann. Appl. Stat.* **1**(2), 302–332 (2007)
23. Gigli, N., Mosconi, S.: The abstract Lewy–Stampacchia inequality and applications. *J. Math. Pures Appl.* **104**, 258–275 (2015)
24. Grassmair, M.: The equivalence of the taut string algorithm and BV-regularization. *J. Math. Imaging Vis.* **27**, 56–66 (2007)
25. Hinterberger, W., Scherzer, O.: Variational methods on the space of functions of bounded Hessian for convexification and denoising. *Computing* **76**, 109–133 (2006)
26. Hintermüller, W., Kunisch, K.: Total bounded variation regularization as a bilaterally constrained optimization problem. *SIAM J. Appl. Math.* **64**(4), 1311–1333 (2004)
27. Knoll, F., Bredies, K., Pock, T., Stollberger, R.: Second order total generalized variation (TGV) for MRI. *Magn. Reson. Med.* **65**, 480–480 (2011)
28. Lewy, H., Stampacchia, G.: On the smoothness of superharmonics which solve a minimum problem. *J. Anal. Math.* **23**, 227–236 (1970)

29. Mammen, E., van de Geer, S.: Locally adaptive regression splines. *Ann. Stat.* **25**(1), 387–413 (1997)
30. Meyer, Y.: *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations*. American Mathematical Society, New York (2000)
31. Niyobuhungiro, J.: *Exact minimizers in real interpolation–characterization and applications*. Linköping Studies in Science and Technology Dissertations, vol. 1650 (2014)
32. Overgaard, N.C.: On the taut string interpretation of the one-dimensional Rudin–Osher–Fatemi model: a new proof, a fundamental estimate and some applications, pp. 1–19. [arXiv:1710.10985](https://arxiv.org/abs/1710.10985) [eess.IV] (2017)
33. Overgaard, N.C.: On the taut-string interpretation of the one-dimensional Rudin–Osher–Fatemi model. *Proc. ICPRAM* **2018**, 233–244 (2018)
34. Papafitsoros, K., Bredies, K.: A study of the one dimensional total generalised variation regularisation problem
35. Papafitsoros, K., Schönlieb, C.B., Sengul, B.: Combined first and second order total variation inpainting using split Bregman. *Image Process. On Line* **3**, 112–135 (2013)
36. Pöschl, C., Scherzer, O.: Characterization of minimizers of convex regularization functionals. *Contemp. Math.* (2008). <https://doi.org/10.1090/conm/451/08784>
37. Rudin, L., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. *Physica D* **60**, 259–268 (1992)
38. Rudin, W.: *Real and Complex Analysis*, 3rd edn. McGraw-Hill, New York (1986)
39. Scherzer, O., Grasmair, M., Grossauer, H., Haltmeier, M., Lenzen, F.: *Variational Methods in Imaging*. Springer, New York (2009)
40. Setterqvist, E.: *Taut strings and real interpolation*. Linköping Studies in Science and Technology Dissertations, vol. 1801 (2016)
41. Steidl, G., Didas, S., Neumann, J.: Splines in higher order TV regularization. *Int. J. Comput. Vis.* **70**(3), 241–255 (2006)
42. Strong, D., Chan, T.: Exact solutions to total variation regularization problems. *CAM Rep.* **96–41**, 259–268 (1996)
43. Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., Knight, K.: Sparsity and smoothness via the fused lasso. *J. R. Stat. Soc. Ser. B* **67**(1), 91–108 (2005)
44. Zhu, M., Wright, S.J., Chan, T.F.: Duality-based algorithms for total variation image restoration. *Comput. Optim. Appl.* **47**(3), 377–400 (2010)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Niels Chr Overgaard Grew up in Denmark where he earned his M.Sc. in Mathematics and Physics from Aalborg University. He moved to Sweden where he received a Ph.D. in Mathematics from Lund University in 2002. His mathematical interests include application of variational calculus to image analysis problems and PDE models for bacterial colony growth. In his spare time, he enjoys cycling and has taken part in several cyclosportives such as the local 300 km Vätternrundan or the 430 km Jotunheimen Runt in Norway.