On the Taut String Interpretation of the One-dimensional Rudin–Osher–Fatemi Model

Niels Chr Overgaard
Centre for Mathematical Sciences, Lund University, Sölvegatan 18A, 221 00 Lund, Sweden

Keywords: Total Variation Minimization, Taut String, Denoising, Regression Splines, Lewy–Stampacchia Inequality, Sub-modularity, Moreau-Yosida Approximation, Isotonic Regression.

Abstract: A new proof of the equivalence of the Taut String Algorithm and the one-dimensional Rudin–Osher–Fatemi model is presented. Based on duality and the projection theorem in Hilbert space, the proof is strictly elementary. Existence and uniqueness of solutions (in the continuous case) to both denoising models follow as by-products. The standard convergence properties of the denoised signal, as the regularizing parameter tends to zero, are recalled and efficient proofs provided. Moreover, a new and fundamental estimate on the denoised signal is derived. It implies, among other things, the strong convergence (in the space of functions of bounded variation) of the denoised signal to the in-signal as the regularization parameter vanishes.

1 INTRODUCTION

In 2017 it was 25 years ago Leonid Rudin, Stanley Osher and Emad Fatemi proposed their now classical model for edge-preserving denoising of images (Rudin et al., 1992). The present paper will investigate the properties of the one-dimensional version of the Rudin-Osher-Fatemi (ROF) model: To a given (noisy) signal $f \in L^2(I)$, defined on a bounded interval $I = (a,b)$, associate the (ROF) functional

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

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

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

The first term in the ROF-functional is λ times the *total variation* $\int_a^b |u'|\,dx$ of the function u and $BV(I)$ denotes the set of functions on I with finite total variation. Precise definitions will be given below, in Sections 2 and 3.

The one-dimensional ROF model is compared to the *Taut string algorithm*—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. Figure 1):

Algorithm 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

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

(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 $W_\lambda \in T_\lambda$ (the 'Taut string') of the energy

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

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 van de Geer, 1997; Davies and Kovac, 2001) and (Dümbgen and Kovac, 2009). Very recently, using methods from

Overgaard, N.
On the Taut String Interpretation of the One-dimensional Rudin–Osher–Fatemi Model.
DOI: 10.5220/0006720402330244
In *Proceedings of the 7th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2018)*, pages 233-244
ISBN: 978-989-758-276-9

interpolation theory (Peetre's K-functional and the notion of invariant K-minimal sets), Setterqvist has investigated the limits to which taut string methods may be extended (Setterqvist, 2016).

In its original formulation, the Taut string algorithm instruct us to find the solution of the shortest path problem

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

hence the epithet 'taut string'. However, the 'stretched rubber band'-energy E in step 3 of the algorithm is not only easier to handle analytically, it also has precisely the same solution as (3). While this is intuitively clear from our everyday experience with rubber bands and strings, the assertion is, mathematically speaking, not equally self-evident so a proof is offered in Appendix A.

The main purpose of this paper, the first of two, 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 (Mammen and van de Geer, 1997) and in (Davies and Kovac, 2001). In the continuum setting, the equivalence result was explicitly stated and proved in (Grassmair, 2007). There is also an extensive treatment in (Scherzer et al., 2009, Ch. 4). Indeed, a few years earlier (Hintermüller and Kunisch, 2004, p.7), in a brief (but inconclusive) remark, refer to the close relation between the ROF model and the Taut string algorithm.

The second main result of the paper, whose proof we give in Section 6, is the following "fundamental" estimate on the denoised signal:

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

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

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

Just like f', 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. (Rudin, 1986, Sec. 6.6). As an example, compute the derivatives of f and $u_\lambda = f_\lambda$ as shown in Figure 1. The proof of the theorem is based on (an extension of) the Lewi–Stampacchia-inequality (Lewy and Stampacchia, 1970) and uses the Taut String-interpretation of the ROF model (Theorem 1) in an essential way.

A significant consequence of Theorem 2 is that for an in-signal f belonging to $BV(I)$ we get $u_\lambda \to f$ strongly in $BV(I)$ as $\lambda \to 0+$. The usual Moreau–Yosida approximation result, see e.g. (Ambrosio et al., 2000, Ch. 17), only gives the weaker $u_\lambda \to f$ in $L^2(I)$ and $\int_I |u'_\lambda|\,dx \to \int_I |f'|\,dx$ as λ tends to zero.

Further contributions of the paper are: i) The rederivation some known properties of the ROF model (Propositions 2) and ii) proof of some precise results on the rate of convergence of $u_\lambda \to f$ as λ tends to zero Propositions 3 and 4—collecting all such result in one place! iii) A new and slick proof of the (known) fact that u_λ is a semi-group with respect to λ (Proposition 8). iv) Finally we indicated how our method of proof can be modified and applied to the problem of *isotonic regression*.

The paper is based on the author's preprint (Overgaard, 2017) and is completely theoretical. All examples, including the ones in the figures, are computed by hand using the taut string interpretation. However, we predict that the theory developed can be used to construct new fast non-iterative algorithms for denoising using the ROF-model or, at least, can be used to shed new light on existing such algorithms such as (Condat, 2013).

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.

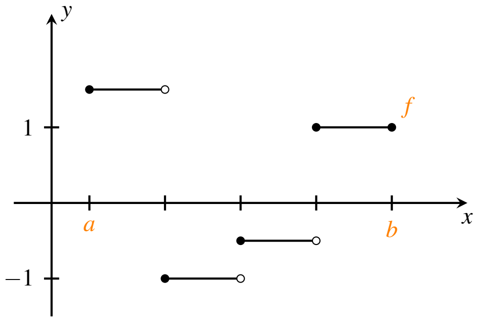
$C_0^1(I)$ denotes the space of continuously differentiable (test-)functions $\xi : I \to \mathbf{R}$ with compact support in I, and $C(\bar{I})$ is the space of continuous functions on the closure of I.

For $1 \le p \le \infty$, $L^p(I)$ denotes the Lebesgue space of measurable functions $f : I \to \mathbf{R}$ with finite p-norm; $\|f\|_p := \left(\int_a^b |f(x)|^p\,dx\right)^{1/p} < \infty$, when p is finite, and $\|f\|_\infty = \operatorname{ess\,sup}_{x\in I} |f(x)| < \infty$ 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 $\|f\| := \sqrt{\langle f, f\rangle_{L^2(I)}} = \|f\|_2$.

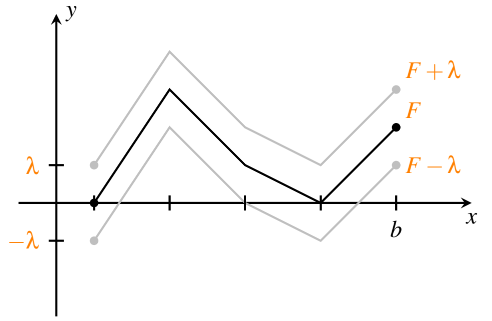
We are going to need the Sobolev spaces over L^2:

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

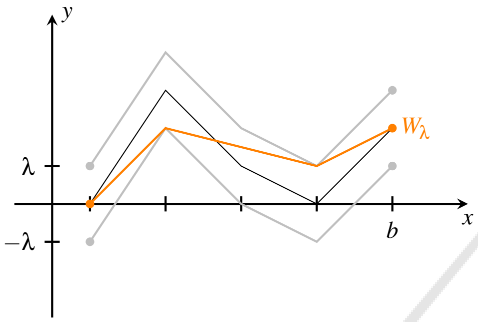
were u' denotes the distributional derivative of u. This is a Hilbert space with inner product $\langle u, v\rangle_{H^1} := \langle u, v\rangle + \langle u', v'\rangle$ and norm $\|u\|_{H^1} = (\|u'\|_2^2 + \|u\|_2^2)^{1/2}$. Any $u \in H^1(I)$ can, after correction on a set of measure zero, be identified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



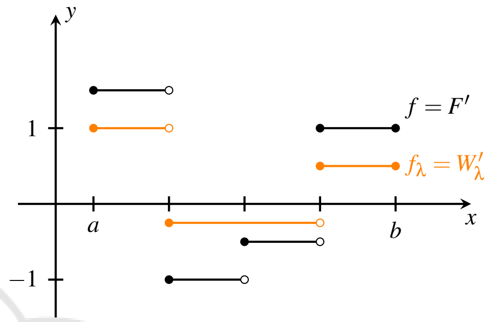
(a) The input signal f.



(b) The cumulative signal F and the tube T_λ.



(c) The taut string W_λ.



(d) The denoised signal f_λ together with the input signal.

Figure 1: A graphical illustrations of the steps in the Taut string algorithm applied to a piecewise constant signal.

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 $H^1(I)$.

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 (Brézis, 1999, Théorème V.2):

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

$$\|\varphi - u\| = \min_{v \in K} \|\varphi - v\|.$$

Moreover, the minimizer u is characterized by the following property:

$$u \in K \quad \text{and} \quad \langle \varphi - u, v - u\rangle \le 0, \text{ for all } v \in K.$$

The point u is called the projection of φ onto K, and is denoted $u = P_K(\varphi)$.

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

$$J(u) = \sup\left\{ \int_a^b u(x)\xi'(x)\,dx : \xi \in C_0^1(I), \|\xi\|_\infty \le 1 \right\}. \tag{5}$$

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 (Chambolle, 2004)). 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 $u \in H^1(I)$ then $J(u) = \int_I |u'|\,dx < \infty$, so $u \in BV(I)$.

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

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

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

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. □

In this example the supremum is attained by 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 3. *Every $u \in BV(I)$ belongs to $L^2(I)$ and*

$$J(u) = \sup_{\xi \in K} \langle u, \xi' \rangle_{L^2(I)}\,, \tag{6}$$

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 (Ambrosio et al., 2000, 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 (6). 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

$$\langle u, \xi' \rangle_{L^2(I)} \leq J(u)\|\xi\|_\infty, \quad \text{for all } \xi \in H_0^1(I), \tag{7}$$

as it implies that the right hand side of (6) cannot exceed $J(u)$. To do this, we first notice that the inequality in holds for all $\zeta \in C_0^1(I)$. This follows by applying homogeneity to the definition of $J(u)$. Secondly, 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 $\zeta_n \to \xi$ in $H_0^1(I)$ (and in $L^\infty(I)$ by the continuous embedding). It follows that

$$\langle u, \xi' \rangle_{L^2(I)} = \lim_{n\to\infty} \langle u, \zeta_n' \rangle_{L^2(I)} \leq J(u) \lim_{n\to\infty} \|\zeta_n\|_\infty = J(u)\|\xi\|_\infty,$$

which establishes (7) and the proof is complete. □

The inequality (7) combined with the Riesz representation theorem (cf. e.g. (Ambrosio et al., 2000, Thm. 1.54)) 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 $E_\lambda : BV(I) \to \mathbf{R}$ given by

$$E_\lambda(u) = \lambda J(u) + \frac{1}{2}\|f - u\|_{L^2(I)}^2\,. \tag{8}$$

Denoising according to the ROF model is the map $L^2(I) \ni f \mapsto u_\lambda \in BV(I)$ defined by (1). To emphasise the role of the in-signal f we sometimes write $E_\lambda(f;u)$ instead of $E_\lambda(u)$. Well-posedness of the ROF model is demonstrated in the next section.

4 EXISTENCE THEORY FOR THE ROF MODEL

We begin with a simple observation: if $u \in BV(I)$ then $J(u + c) = J(u)$ for any real constant c. This property of the total variation has two important consequences. First of all, $E_\lambda(f;u) = E_\lambda(f - c; u - c)$ for any constant c. Taking c to be the mean value of f shows that we may assume, as we do throughout this paper, that the in-signal satisfies $\int_I f\,dx = 0$. This assumption implies that the cumulative signal $F(x)$ satisfies $F(a) = F(b) = 0$, hence $F \in H_0^1(I)$. This plays an important role in our analysis. Secondly, since f has mean value zero, it is enough to minimize E_λ over the subspace of $BV(I)$ consisting of functions with mean value zero. To see this, let P be the orthogonal projection (in $L^2(I)$) onto this subspace. An easy computation yields the identity $E_\lambda(Pu) = E_\lambda(u) - \frac{1}{2}\|u - Pu\|^2$, which shows that u can be a minimizer of E_λ only if it belongs to the range of P.

The following theorem contains the key result of our paper.

Theorem 4. *We have the equality*

$$\min_{u \in BV(I)} E_\lambda(u) = \max_{\xi \in K} \frac{1}{2}\Big\{ \|f\|_{L^2(I)}^2 - \|f - \lambda\xi'\|_{L^2(I)}^2 \Big\}\,, \tag{9}$$

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

$$u_\lambda = f - \lambda\xi_\lambda'\,, \tag{10a}$$

and satisfy

$$J(u_\lambda) = \langle u_\lambda, \xi_\lambda' \rangle_{L^2(I)}\,. \tag{10b}$$

Moreover, if $u_\lambda \neq 0$, then $\|\xi_\lambda\|_\infty = 1$. Conversely, the conditions (10a) and (10b) characterizes the solution; if a pair of functions $\bar{u} \in BV(I)$ 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 a special instance of the Fenchel–Rockafellar theorem, see e.g. (Brézis, 1999, 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 is used in (Hintermüller and Kunisch, 2004) in their analysis of the multidimensional ROF model (with the 'Manhattan metric'). The equality (9) has played an important role in the development of numerical algorithms for total variation minimization, both directly, as for instance in (Zhu et al., 2007) or, indirectly, as in (Chambolle, 2004).

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

$$\inf_{x\in M} \sup_{y\in N} \Phi(x,y) \geq \sup_{y\in N} \inf_{x\in M} \Phi(x,y) , \qquad (11)$$

is always true. The use of inf's and sup's are important, 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 (11) that

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

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:

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

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

$$\inf_{u\in BV(I)} E_\lambda(u) \geq \sup_{\xi\in K} \frac{1}{2}\left\{ \|f\|^2 - \|f-\lambda\xi'\|^2 \right\} \qquad (12)$$

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

$$\begin{aligned} \inf_{\xi\in K} \|f-\lambda\xi'\| &= \inf_{\xi\in K} \|F'-\lambda\xi'\| \\ &= \lambda \inf_{\xi\in K} \|\lambda^{-1}F - \xi\|_{H_0^1(I)} . \qquad (13) \end{aligned}$$

By Proposition 1, this problem has the unique solution $\xi_\lambda = P_K(\lambda^{-1}F) \in K$, so the supremum is attained in (12). Now, let the function u_λ be defined by (10a) in the theorem. A priori, u_λ belongs to $L^2(I)$, but we are going to show that $u_\lambda \in BV(I)$: The characterization of ξ_λ according in the projection theorem states that $\xi_\lambda \in K$ and $\langle f - \lambda\xi_\lambda', \lambda\xi' - \lambda\xi_\lambda'\rangle \leq 0$ 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 $u_\lambda \in BV(I)$ with $J(u_\lambda) = \langle u_\lambda, \xi_\lambda'\rangle$, as asserted in the theorem. (This reasoning can be reversed; if (10b) is true then ξ_λ is the minimizer in (13).) Also, if $u_\lambda \neq 0$ then $\|\xi_\lambda\|_\infty < 1$ is not consistent with the maximizing property (10b), hence $\|\xi_\lambda\|_\infty = 1$, as claimed.

It remains to be verified that u_λ minimizes E_λ and that equality holds in (12). This follows from a direct calculation:

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

So $\inf E_\lambda(u) = E_\lambda(u_\lambda)$, the infimum is attained, and equality holds in (12). 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 back-tracking the steps of the above proof. □

Denoising is a non-expansive mapping:

Corollary. *If f and $\tilde{f}$ are signals in $L^2(I)$ and the corresponding denoised signals are denoted u_λ and $\tilde{u}_\lambda$, respectively, then $\|\tilde{u}_\lambda - u_\lambda\|_{L^2(I)} \leq \|\tilde{f} - f\|_{L^2(I)}$.*

This is a special instance of a more general result about Moreau–Yosida approximation (or of the proximal map), see (Attouch et al., 2015, Theorem 17.2.1). However, the result is easily verified by the reader using the characterization of the ROF-minimzer given in the theorem.

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

Proof of Theorem 1. It follows from Theorem 4 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

$$\min_{\xi\in K} \frac{1}{2}\|f-\lambda\xi'\|^2_{L^2(I)} . \tag{14}$$

If we introduce the new variable $W := F - \lambda\xi$, where $F \in H^1_0(I)$ is the cumulative signal, then $W \in H^1_0(I)$ and the condition $\|\xi\|_\infty \le 1$ implies that W satisfies $F(x) - \lambda \le W(x) \le F(x) + \lambda$ on I. Therefore (14) is equivalent to $\min_{W\in T_\lambda}(1/2)\|W'\|^2_{L^2(I)}$, which is the minimization problem in step 3 of the Taut string algorithm whose solution we denoted 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. □

It is interesting to note that Theorem 4 associates a *unique* test function (or 'dual variable') $\xi_\lambda \in K$ with the solution u_λ of the ROF model such that $J(u_\lambda) = \langle u_\lambda, \xi'_\lambda\rangle_{L^2}$, in particular if we compare to the situation in Example 1. A concrete case looks as follows:

Example 2. Let $f(x) = \mathrm{sign}(x)$ be the step function defined on $I = (-1,1)$. An easy calculation, based on the Taut string interpretation, shows that if $0 < \lambda < 1$ then $u_\lambda = (1-\lambda)\,\mathrm{sign}(x)$ and $\xi_\lambda = |x| - 1 \in H^1_0(I)$. Here ξ_λ is not in $C^1_0(I)$, so the extension of the space of test functions from C^1_0 to H^1_0 is essential to our theory. For $\lambda \ge 1$ we find $u_\lambda = 0$ and $\xi_\lambda = \lambda^{-1}(|x|-1)$. Notice that $\|\xi_\lambda\|_\infty = 1$ when $u_\lambda \ne 0$. □

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 (Grassmair, 2007) 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 (Scherzer et al., 2009) is based on a thorough functional analytic study of Meyer's G-norm and is not elementary.

5 CONSEQUENCES OF THE EQUIVALENCE RESULT

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 2. *(a) The denoised signal $u_\lambda = 0$ if and only if $\lambda \ge \|F\|_\infty$, 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 multi-dimensional case, can be found in Meyer's treatise (Meyer, 2000). 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 4 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 (13), ξ_λ is the projection in $H^1_0(I)$ 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 $\|F\|_\infty \le \lambda$, as claimed.

(b) If $0 < \lambda < \|F\|_\infty$ then $u_\lambda \ne 0$ hence $\|\xi_\lambda\|_\infty = 1$, by Theorem 4. The assertion now follows by taking norms in the identity $\lambda\xi_\lambda = F - W_\lambda$.

(c) The equality clearly holds when $\lambda \ge \|F\|_\infty$ 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 $E(\hat{W}) \le E(W)$, and W_λ is the (unique) minimizer of E over T_λ, we conclude that $\max_I W_\lambda \le c$. A similar argument gives $-\min_I W_\lambda \le c$. Thus $\|W_\lambda\|_\infty \le \max(0, \|F\|_\infty - \lambda)$. The reverse inequality follows from (b). □

Now define, for $\lambda > 0$, the *value function*

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

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

Proposition 3. *The function $e : (0,+\infty) \to (0,+\infty)$ is nondecreasing and concave, hence continuous, and satisfies $e(\lambda) = \|f\|^2/2$ for $\lambda \ge \|F\|_\infty$. Moreover, for $f \in L^2(I)$*

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

and if $f \in BV(I)$ then $e(\lambda) = O(\lambda)$ as $\lambda \to 0+$.

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

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 $e(\lambda) = \inf_{u \in BV(I)} E_\lambda(u)$ 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 $f \in BV(I)$, 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 $\varepsilon > 0$ take a function $f_\varepsilon \in H_0^1(I)$ such that $\|f - f_\varepsilon\|^2/2 < \varepsilon$. Then $f_\varepsilon \in BV(I)$ and $0 \leq e(\lambda) \leq E_\lambda(f_\varepsilon) < \lambda J(f_\varepsilon) + \varepsilon$. It follows that $0 \leq \limsup_{\lambda \to 0+} e(\lambda) < \varepsilon$. Since ε is arbitrary, we get $\lim_{\lambda \to 0+} e(\lambda) = 0$. □

The first part of next the proposition is a special instance of a much more general result, see (Attouch et al., 2015, Theorem 17.2.1). The second part contains a quantification of the rate of convergence which is *not* easily located in the literature.

Proposition 4. *For any $f \in L^2(I)$ we have $u_\lambda \to f$ in L^2 as $\lambda \to 0+$. Moreover, if $f \in BV(I)$ then $\|u_\lambda - f\|_{L^2(I)} = o(\lambda^{1/2})$ and $J(u_\lambda) \to J(f)$ as $\lambda \to 0+$.*

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

$$\|u_\lambda - f\|^2_{L^2(I)} \leq 2\lambda(J(f) - J(u_\lambda)) . \tag{15}$$

Consequently $\|u_\lambda - f\|^2_{L^2(I)} = O(\lambda)$ and $J(u_\lambda) \leq J(f)$ for all $\lambda > 0$. But we can do slightly better than that. Since $u_\lambda \to f$ in L^2 as $\lambda \to 0+$, we get $J(f) \leq \liminf_{\lambda \to 0+} J(u_\lambda)$, by the lower semi-continuity of the total variation J, cf. (Ambrosio et al., 2000). Since $J(u_\lambda) \leq J(f)$ we also obtain an estimate from below: $\limsup_{\lambda \to 0+} J(u_\lambda) \leq J(f)$. We conclude that $\lim_{\lambda \to 0+} J(u_\lambda) = J(f)$. If this is used in (15) we find that $\|u - f\|^2_{L^2(I)} = o(\lambda)$ as $\lambda \to 0+$. □

6 PROOF AND APPLICATIONS OF THEOREM 2

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

Proof of Theorem 2. The estimate (4) is a consequence of the extension to bilateral obstacle problems of the original Lewy–Stampacchia inequality (Lewy and Stampacchia, 1970) which we explain here. 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 (2) 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 functions $\phi, \psi \in H^1(I)$ which satisfy the conditions $\phi < \psi$ on I, and $\phi < 0 < \psi$ on $\partial I = \{a, b\}$. This ensures that C is nonempty.

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

$$-(\phi'')^- \leq u_0'' \leq (\psi'')^+ . \tag{16}$$

Here the notation μ^+ and μ^- is used to denote the positive and negative variation, respectively, of a signed measure μ. This is the generalization of the Lewy–Stampacchia inequality, proof of which can be found in Appendix B. This proof is based on the abstract proof, valid in a much more general setting, given in (Gigli and Mosconi, 2015). The assumption of our theorem, that $f \in BV(I)$, implies that $F'' = f'$ is a signed measure. If we apply (16) 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 use the identities $F' = f$ and $W_\lambda' = u_\lambda$ into the above inequality. □

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 weight approaches zero.

Proposition 5. *If $f \in BV(I)$ then*

$$J(f - u_\lambda) = J(f) - J(u_\lambda). \tag{17}$$

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

Proof. The measures $(f')^+$ and $(f')^-$ are concentrated on disjoint measurable sets (Hahn decomposition, see (Rudin, 1986, Sec. 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 $\lambda \to 0+$, by Proposition 4. □

Theorem 2 also implies the first part of

Proposition 6. *If f is piecewise constant function on I, then so is u_λ for all $\lambda > 0$. Moreover, there exists a number $\bar{\lambda} > 0$ and a piecewise linear function $\bar{\xi} \in K$ such that $\xi_\lambda = \bar{\xi}$ for all λ, $0 < \lambda \le \bar{\lambda}$.*

We only give the proof of the first part of this theorem, which is simple, and omit the proof of the second part, which is rather lengthy.

Proof. If f is piecewise constant then there exists nodes $a = x_0 < x_1 < \ldots < x_{N-1} < x_N = b$ 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 $i = 1,\ldots,N$. That is,

$$f = \sum_{i=1}^{N} f_i \chi_{I_i},$$

where, as usual, χ_A denotes the characteristic function of the set A. The derivative of the signal becomes $f' = \sum_{i=1}^{N-1}(f_{i+1} - f_i)\delta_{x_i}$, with δ_x denoting the Dirac measure supported at x, and therefore $J(f) = \sum_{i=1}^{N-1} |f_{i+1} - f_i| < \infty$. Therefore f belongs to $BV(I)$ and Theorem 2 may be applied:

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

This estimate shows that $u'_\lambda = \sum_{i=1}^{N-1} c_i(\lambda)\delta_{x_i}$ where the real numbers $c_i(\lambda)$ satisfy $0 \le c_i(\lambda)\cdot(f_{i+1} - f_i)^{-1} \le 1$ for $i = 1,\ldots,N-1$. Since the derivative is zero except at a finite set of points we draw the conclusion that u_λ is piecewise constant $u_\lambda = \sum_{i=1}^{N}(u_\lambda)_i \chi_{I_i}$ with nodes contained in the node set of f. (The latter is the "edge-preserving" property of the one-dimensional ROF model.) Once the $c_i(\lambda)$'s are known the N real numbers $(u_\lambda)_i$ may be determined from the N linear equations $(u_\lambda)_{i+1} - (u_\lambda)_i = c_i(\lambda)$, $i = 1,\ldots,N-1$, and $\sum_{i=1}^{N}(u_\lambda)_i(x_i - x_{i-1}) = \int u_\lambda\,dx = \int f\,dx = 0$. □

The latter half of the above proposition can be proved by "guessing" the the dual variable ξ_λ—it must be a continuous piecewise linear function with the same nodes as f—and then use the characterization of solutions from Theorem 4. This result is mentioned because it implies what is possibly the strongest imaginable approximation result:

Proposition 7. *If f is piecewise constant function, then $\|f - u_\lambda\|_{L^2(I)} = O(\lambda)$, $\lambda \to 0+$.*

Proof. We know from (15) that $(1/2)\|f - u_\lambda\|^2_{L^2(I)} \le \lambda(J(f) - J(u_\lambda))$ so an estimate of the difference $J(f) - J(u_\lambda)$ is needed. By Theorem 4, $J(u_\lambda) = \langle u_\lambda, \xi'_\lambda\rangle_{L^2(I)}$. Since $\xi_\lambda = \bar{\xi}$ when λ is close to zero it follows that

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

Moreover, the scalar product of $u_\lambda = f - \lambda\xi'_\lambda$ and $\xi'_\lambda = \bar{\xi}'$ is $J(u_\lambda) = \langle u_\lambda, \bar{\xi}'\rangle = \langle f - \lambda\bar{\xi}', \bar{\xi}'\rangle = J(f) - \lambda\|\bar{\xi}'\|^2$. Hence $J(f) - J(u_\lambda) = O(\lambda)$, $\lambda \to 0+$. □

Our interest in the various limits as $\lambda \to 0+$ is motivated by the fact that $\lambda \mapsto u_\lambda$ is a semi-group; statements about limits at $\lambda = 0$ can be translated to limits at any $\lambda > 0$.

Proposition 8 (Semi-group property)**.** *Let $f \in L^2(I)$. With the convention (mentioned above) that $u_0 = f$ the formula*

$$(u_\lambda)_\mu = u_{\lambda+\mu}$$

holds for all $\lambda, \mu \ge 0$.

Here we have tweaked the notation slightly to make the statement more compact: By using the letter u in place of f for the in-signal, the operation of denoising, for some $\lambda > 0$, is indicated by adding the subscript 'λ' to u, thus obtaining u_λ. This makes sense even for $\lambda = 0$ if we agree to set $u_0 = u$.

A proof of the semi-group property can be found in (Scherzer et al., 2009). However, the fundamental estimate in Theorem 2 and the characterization of the ROF-minimizer in Theorem 4 allow us to present short and very direct proof of this result:

Proof. The assertion holds trivially if either λ or μ equals zero, so we may assume that $\lambda, \mu > 0$. The idea of the proof is then to set $\bar{u} = (u_\lambda)_\mu$ and show that there exists a function $\bar{\xi} \in K$ such that

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

The characterization of solutions to the ROF model in Theorem 4 then implies that $\bar{u}$ equals $u_{\lambda+\mu}$. Since u_λ and $\bar{u}$ are the ROF-minimizers of $E_\lambda(f;\cdot)$ and $E_\mu(u_\lambda;\cdot)$, respectively, they both satisfy the conditions (10a) and (10b), that is

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

for a uniquely determined pair of functions ξ_λ and $\bar{\xi}_\mu$ in K. Now, if we set

$$\xi = \frac{\lambda\xi_\lambda + \mu\bar{\xi}_\mu}{\lambda+\mu}$$

then $\bar{\xi} \in K$ because it is the convex combination of two elements of K. Using what is known about u_λ and $\bar{u}$, the following calculation reveals why we make this definition of $\bar{\xi}$, in fact $f - (\lambda+\mu)\bar{\xi}' = f - \lambda\xi_\lambda' - \mu\bar{\xi}_\mu = u_\lambda - \mu\bar{\xi}_\mu = \bar{u}$, hence $\bar{u}$ and $\bar{\xi}$ fulfil the condition (10a) by construction. It remains to verify that (10b) is fulfilled as well. Since

$$\langle \bar{u}, \bar{\xi}' \rangle = \frac{\lambda}{\lambda+\mu}\langle \bar{u}, \xi_\lambda' \rangle + \frac{\mu}{\lambda+\mu}\langle \bar{u}, \bar{\xi}_\mu' \rangle$$

$$= \frac{\lambda}{\lambda+\mu}\langle \bar{u}, \xi_\lambda' \rangle + \frac{\mu}{\lambda+\mu} J(\bar{u}).$$

we see that the second condition follows if it can show that $\langle \bar{u}, \xi_\lambda' \rangle = J(\bar{u})$. This essentially follows from the identity in Proposition 5 which states that $J(\bar{u}) = J(u_\lambda) - J(u_\lambda - \bar{u})$. In fact, using this identity we get the inequality

$$J(\bar{u}) \le J(u_\lambda) - \langle u_\lambda - \bar{u}, \xi_\lambda' \rangle$$

$$= J(u_\lambda) - J(u_\lambda) + \langle \bar{u}, \xi_\lambda' \rangle = \langle \bar{u}, \xi_\lambda' \rangle$$

But $J(\bar{u}) \ge \langle \bar{u}, \xi' \rangle$ for all $\xi \in K$, so $J(\bar{u}) = \langle \bar{u}, \xi_\lambda' \rangle$, and the proof is complete. □

The last part of the proof yields

Corollary. *If* $\lambda > 0$ *then* $J(u_\lambda) = \langle u_\lambda, \xi_\mu' \rangle_{L^2(I)}$ *for all* μ, $0 < \mu \le \lambda$.

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

7 APPLICATION TO ISOTONIC REGRESSION

We illustrates the usefulness of our approach by briefly outlining (without proofs) how the theory developed earlier can be modified in order to derive the so-called "lower convex envelope" interpretation of the solution to the problem of isotonic regression. Isotonic regression is a method from mathematical statistics used for non-parametric estimation of probability distributions, see for instance (Anevski and Soulier, 2011). It is a least-squares problem with a monotonicity constraints: given $f \in L^2(I)$, find the non-decreasing function $u_\uparrow \in L^2(I)$ which solves the minimization problem,

$$\min_{u \in L^2_\uparrow(I)} \frac{1}{2}\|u - f\|^2_{L^2(I)}, \tag{18}$$

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

The idea is to re-formulate (18) as an unconstrained optimization problem by replacing the total variation term J of the ROF functional by regularization term $J_\uparrow$ which can distinguish between functions that are non-decreasing or not. To achieve this we set $K_+ = \{\xi \in H_0^1(I) : \xi(x) \ge 0 \text{ for all } x \in I\}$ and define $J_\uparrow(u) = \sup_{\xi \in K_+} \langle u, \xi' \rangle_{L^2(I)}$. It can be shown that

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

The isotonic regression problem (18) now becomes equivalent to finding the minimizer $u_\uparrow$ in $L^2(I)$ of the functional

$$E_\uparrow(u) := J_\uparrow(u) + \frac{1}{2}\|u - f\|^2_{L^2(I)}. \tag{19}$$

Notice that there is no need for a positive weight in this functional because the regularizer assumes only the values zero and infinity.

Again we may assume the mean value f to be zero so that the cumulative function F belongs to $H_0^1(I)$. Mimicking the proof of Theorem 4 we get:

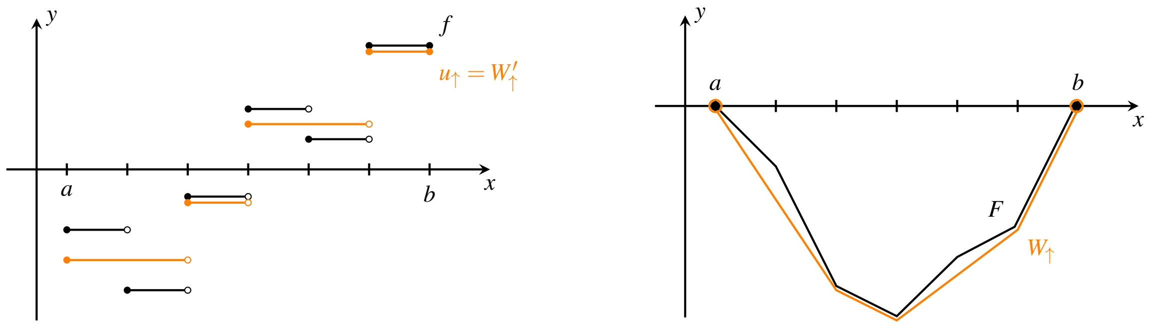
$$\min_{u \in L^2(I)} E_\uparrow(u) = \max_{W \in T} \frac{1}{2}\left\{\|f\|^2 - \frac{1}{2}\|W'\|^2_{L^2(I)}\right\}$$

where $W = F - \xi$, $\xi \in K_+$, and $T = \{W \in H_0^1(I) : W(x) \le F(x), x \in I\}$. The minimization of (19) is equivalent to the obstacle problem $\min_{W \in T} \frac{1}{2}\|W'\|^2_{L^2(I)}$ which admits a unique solution $W_\uparrow$ by the Projection theorem. It follows that (19) 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.

The solution $W_\uparrow$ of the obstacle problem satisfies $W_\uparrow'' \ge 0$ (this is the 'easy' part of the original Lewy-Stampacchia inequality, $0 \le W_\uparrow'' \le (F'')^+$) 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 higer-order methods (the space of functions with bounded Hessians), in Hinterberger and Scherzer (Hinterberger and Scherzer, 2006).

8 CONCLUDING REMARKS

We have developed the theory for the one-dimensional ROF model in the continuous setting for a quite general class of signals and proved several properties of solution of the model including 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 in



(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) $W_{\uparrow}$.

Figure 2: A graphical illustrations of the taut string interpretation of isotonic regression.

Proposition 6 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"). This observation, together with the semi-group property, immediately suggests a (perhaps not so efficient) non-iterative algorithm for the computation of the denoising of a piecewise constant signal which is different from the one in (Condat, 2013). If very fast non-iterative schemes for finding the solution to the one-dimensional ROF model can be devised then, as already indicated by L. Condat, efficient iterative algorithms for image denoising using the two-dimensional ROF model (with the 'Manhattan metric' as measure of the image gradient magnitude) may be constructed as well. We hope to return to this topic in the future.

ACKNOWLEDGEMENTS

The author whishes to thank Viktor Larsson at the Centre for Mathematical Sciences, Lund University, for reading and commenting on several drafts of this paper.

REFERENCES

Ambrosio, L., Fusco, N., and Pallara, D. (2000). *Functions of Bounded Variation and Free Discontinuity Problems*. Clarendon Press, Oxford.

Anevski, D. and Soulier, P. (2011). Monotone spectral density estimation. *Ann. Statistics*, 39:418–438.

Attouch, H., Buttazzo, G., and Michaille, G. (2015). *Variational Analysis in Sobolec and BV Spaces; Applications to PDEs and Optimization*. SIAM, New York, 2nd edition.

Brézis, H. (1999). *Analyse fonctionelle—Théorie et applications*. Dunod, Paris.

Chambolle, A. (2004). An algorithm for total variation minimization and applications. *J. Math. Imaging Vis.*, 20:89–97.

Condat, L. (2013). A direct algorithm for 1d total variation denoising. *IEEE Signal Proc. Letters*, 20(11):1054–1057.

Davies, P. and Kovac, A. (2001). Local extremes, runs, strings and multiresolution. *Annals of Statistics*, 29:1–65.

Dümbgen, L. and Kovac, A. (2009). Extension of smoothing via taut strings. *Electronic Journal of Statistics*, 3:41–75.

Gigli, N. and Mosconi, S. (2015). The abstract lewystampacchia inequality and applications. *J. Math. Pures Appl.*, 104:258–275.

Grassmair, M. (2007). The equivalence of the taut string algorithm and BV-regularization. *J. Math. Imaging Vis.*, 27:56–66.

Hinterberger, W. and Scherzer, O. (2006). Variational methods on the space of functions of bounded Hessian for convexification and denoising. *Computing*, 76:109–133.

Hintermüller, W. and Kunisch, K. (2004). Total bounded variation regularization as a bilaterally constrained optimization problem. *SIAM J. Appl. Math.*, 64(4):1311–1333.

Lewy, H. and Stampacchia, G. (1970). On the smoothness of superharmonics which solve a minimum problem. *J. Analyse Math.*, 23:227–236.

Mammen, E. and van de Geer, S. (1997). Locally adaptive regression splines. *Annals of Statistics*, 25(1):387–413.

Meyer, Y. (2000). *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations*. American Mathematical Society, New York.

Overgaard, N. C. (2017). On the taut string interpretation of the one-dimensional Rudin-Osher-Fatemi model: A new proof, a fundamental estimate and some applications. arXiv:1710.10985 [eess.IV]:1–19.

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

Rudin, W. (1986). *Real and Complex Analysis*. McGraw-Hill, New York, 3rd edition.

Scherzer, O., Grasmair, M., Grossauer, H., Haltmeier, M., and Lenzen, F. (2009). *Variational Methods in Imaging*. Springer, New York.

Setterqvist, E. (2016). Taut strings and real interpolation. *Linkoping Studies in Science and Technology Dissertations*, 1801.

Zhu, M., Wright, S., and Chan, T. (2007). Duality-based algorithms for total variation image restoration. *Comput. Optim. Appl.*, 47:377–400.

APPENDIX A

As promised in the introduction, we are going to prove that the solution of the minimization problem (2) in STEP 3 of the Taut string algorithm coincides with the solution of the shortest path problem (3). In fact we prove the slightly more general statement:

Proposition. *Let H denote any strictly convex C^1-function defined on $\mathbf{R}$ and set*

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

Then the problem $\min_{W\in T_\lambda} L_H(W)$ has precisely the same solution as the minimization problem $\min_{W\in T_\lambda} E(W)$ in (2).

The case we need in our analysis follows by taking $H(s) = (1+s^2)^{1/2}$.

Proof. The idea of the proof is to verify that $W_\lambda := \arg\min_{W\in T_\lambda} E(W)$ solves the variational inequality:

$$\int_I h(W_\lambda'(x))(W'(x) - W_\lambda(x)')\,dx \geq 0\,,\ \text{for all } W \in T_\lambda, \tag{20}$$

where $h = H'$. This condition is both necessary and sufficient for W_λ to be a minimizer of L_H over T_λ, and since L_H is a strictly convex functional, there is at most one such minimizer.

Being the minimizer of E over T_λ, $W_\lambda \in T_\lambda$ satisfies the variational inequality (which is a special case of (20) if we take $H(s) = s$):

$$\int_I W_\lambda'(W' - W_\lambda')\,dx \geq 0,\ \text{for all } W \in T_\lambda. \tag{21}$$

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 solution touches the upper and the lower obstacles, respectively. Since F and W_λ are continuous, both sets are closed. In fact, C_+ and C_- are compact because $\lambda > 0$ implies that they do not reach the boundary of I. They are disjoint, $C_+ \cap C_- = \emptyset$, and their union, $C = C_+ \cup C_-$, is the contact set for W_λ.

For any non-negative $\xi \in C_0^1(I\backslash C_+)$ there exists an $\varepsilon > 0$ such that $W := W_\lambda + \varepsilon\xi$ belongs to T_λ. If this W is substituted into (21) we find that

$$\int_I W_\lambda'\xi'\,dx \geq 0 \quad \text{for all } \xi \in C_0^1(I\backslash C_+) \text{ with } \xi \geq 0.$$

It follows that $-W_\lambda''$ is a positive measure on $I\backslash C_+$, hence $-W_\lambda'$ is non-decreasing on each connected component of $I\backslash C_+$. Similarly one proves that $-W_\lambda'$ is non-increasing on each connected component of $I\backslash C_-$. This means, in particular, that W_λ' constant on each connected component of $I\backslash C$.

Since h is non-decreasing, the composite function $-h(W_\lambda')$ has the same monotonicity properties as $-W_\lambda'$. Therefore the distributional derivative $-h(W_\lambda')'$ is a positive measure μ^+ on $I\backslash C_+$ and minus a positive measure $-\mu^-$ on $I\backslash C_-$. Clearly $\operatorname{supp}\mu^+ \subset C_-$ and $\operatorname{supp}\mu^- \subset C_+$, so $-h(W_\lambda')'$ is a signed measure μ with the Jordan decomposition $\mu = \mu^+ - \mu^-$. The following calculation now verifies (20): 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 $W - W_\lambda \geq 0$ on C_- and $W - W_\lambda \leq 0$ on C_+. □

APPENDIX B

We prove the inequality (16) used in the proof of Theorem 2. With the notation introduced in this proof we can formulate this result as follows:

Proposition (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

$$E(v) - E(u) \geq \int_a^b u'(x)(v'(x) - u'(x))\,dx, \tag{22}$$

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 sufficien condition:

$$\int_a^b u_0'(v' - u_0')\,dx \geq 0 \quad \text{for all } v \in C.$$

Proposition 8 was first proved for the unilateral obstacle problem (in multiple dimensions) in (Lewy and Stampacchia, 1970) 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 $u, v \in H^1(I)$,

$$E(u \wedge v) + E(u \vee v) \leq E(u) + E(v), \qquad (23)$$

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 (Gigli and Mosconi, 2015) 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

$$\min_{u \in H_0^1(I): u \geq u_0} \hat{E}(u). \qquad (24)$$

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

$$\hat{E}(u \wedge \psi) \leq \hat{E}(u). \qquad (25)$$

Since $\psi > 0$ on ∂I and $u \in H_0^1(I)$ we get $u \wedge \psi \in H_0^1(I)$ which is therefore admissible for the min-problem above. The claim is proved using the sub-modularity (23) of E and the identity $u \wedge v + u \vee v = u + v$ with v replaced by ψ:

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

where the last inequality follows from (22).

The claim (25) 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 (24).

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 $u \in H_0^1(1)$ 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. □