SIAM J. Imaging Sciences
Vol. 8, No. 4, pp. 2581–2621

Limiting Aspects of Nonconvex TV^φ Models*

Michael Hintermüller[†], Tuomo Valkonen[‡], and Tao Wu[§]

Abstract. Recently, nonconvex regularization models have been introduced in order to provide a better prior for gradient distributions in real images. They are based on using concave energies φ in the total variation–type functional $\mathrm{TV}^\varphi(u) := \int \varphi(|\nabla u(x)|)\,dx$. In this paper, it is demonstrated that for typical choices of φ, functionals of this type pose several difficulties when extended to the entire space of functions of bounded variation, $\mathrm{BV}(\Omega)$. In particular, if $\varphi(t) = t^q$ for $q \in (0,1)$, and TV^φ is defined directly for piecewise constant functions and extended via weak* lower semicontinuous envelopes to $\mathrm{BV}(\Omega)$, then it still holds that $\mathrm{TV}^\varphi(u) = \infty$ for u not piecewise constant. If, on the other hand, TV^φ is defined analogously via continuously differentiable functions, then $\mathrm{TV}^\varphi \equiv 0$ (!). We study a way to remedy the models through additional multiscale regularization and area strict convergence, provided that the energy $\varphi(t) = t^q$ is linearized for high values. The fact that such energies actually better match reality and improve reconstructions is demonstrated by statistics and numerical experiments.

Key words. total variation, nonconvex, regularization, area-strict convergence, multiscale analysis

AMS subject classifications. 26B30, 49Q20, 65J20

DOI. 10.1137/141001457

1. Introduction.

Recently introduced nonconvex total variation models are based on employing concave energies φ in discrete versions of functionals of the form

$$\mathrm{TV}^\varphi_{\mathrm{c}}(u) := \int_\Omega \varphi(|\nabla u(x)|)\,dx \quad (u \in C^1(\Omega)), \tag{1.1}$$

which we call the *continuous model*, or

$$\mathrm{TV}^\varphi_{\mathrm{d}}(u) := \int_{J_u} \varphi(|u^+(x) - u^-(x)|)\,d\mathcal{H}^{m-1}(x) \quad (u \text{ piecewise constant}), \tag{1.2}$$

*Received by the editors December 27, 2014; accepted for publication (in revised form) August 28, 2015; published electronically November 10, 2015.
http://www.siam.org/journals/siims/8-4/100145.html

†Institute for Mathematics, Humboldt University of Berlin, 10099 Berlin, Germany (hint@math.hu-berlin.de). This author's work was supported by the Austrian FWF SFB F32 "Mathematical Optimization in Biomedical Sciences" and the START-Award Y305.

‡Corresponding author. Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge CB3 0WA, United Kingdom (tuomo.valkonen@iki.fi). A large part of this work was done while this author was at the Center for Mathematical Modeling, Escuela Politécnica Nacional, Quito, Ecuador, where his work was supported by a Prometeo scholarship of the Senescyt. In Cambridge, this author was supported by the King Abdullah University of Science and Technology (KAUST) Award KUK-I1-007-43 and the EPSRC first grant EP/J009539/1 "Sparse & Higher-order Image Restoration."

§Institute for Mathematics and Scientific Computing, University of Graz, 8010 Graz, Austria. Current address: Institute for Mathematics, Humboldt University of Berlin, 10099 Berlin, Germany (wutao.@mathematik.hu-berlin.de). This author's work was supported by the Austrian FWF SFB F32 "Mathematical Optimization in Biomedical Sciences" and the START-Award Y305.

which we call the *discrete model*. Here $\Omega \subset \mathbb{R}^m$ is our image domain, and J_u is the jump set of u, where the one-sided traces $u^\pm$ from different sides of J_u differ. The typical energies include, in particular, $\varphi(t) = t^q$ for $q \in (0,1)$. The models based on discretizations of (1.2) have been proposed for the promotion of piecewise constant (cartoon-like) images [16, 23, 33, 34], whereas models based on discretizations of (1.1) have been proposed for the better modelling of gradient distributions in real-life images [27, 28, 29, 36]. To denoise an image z, one may then solve the nonconvex Rudin–Osher–Fatemi-type problem

$$\min_u \frac{1}{2}\|z-u\|^2_{L^2(\Omega)} + \alpha \mathrm{TV}^\varphi(u) \tag{1.3}$$

for $\mathrm{TV}^\varphi = \mathrm{TV}^\varphi_\mathrm{c}$ or $\mathrm{TV}^\varphi = \mathrm{TV}^\varphi_\mathrm{d}$. Observe that (1.1) is defined rigorously only for differentiable functions. In contrast to (1.2), it is in particular not defined for piecewise constant discretizations or images with discontinuities. The functional has to be extended to the whole space of functions of bounded variation denoted by $\mathrm{BV}(\Omega)$ (see [24] for its definition) in order to obtain a sound model in the nondiscretized setting. Alternatively, we may take (1.2), defined for piecewise constant functions, as the basis and extend it to continuous functions. We will study the extension of both models (1.1) and (1.2) to $\mathrm{BV}(\Omega)$. We demonstrate that (1.1) in particular has severe theoretical difficulties for typical choices of φ. We also demonstrate that some of these difficulties can be overcome by altering the model to better match reality, although we also need additional multiscale regularization in the model for theoretical purposes.

It is worth noting that in the finite-dimensional setting that has mainly been considered in the aforementioned previous works, such problems with well-posedness do not surface. This can also be seen by adaptation of our results, keeping in mind that in a finite-dimensional subspace of $\mathrm{BV}(\Omega)$, weak* and strong convergence are equivalent. Infinite-dimensional models, while more challenging, can be more informative and lead to better understanding of the model—both finite- and infinite-dimensional. Concepts such as smoothness and jumps are, after all, not well defined in the discrete setting. Study of these and other properties in the infinite-dimensional setting can provide valuable information about the behavior of the associated image processing approaches and their solution approaches [1, 7, 10, 13, 14, 20, 46, 47].

Let us consider the specific difficulties with the discrete model $\mathrm{TV}^\varphi_\mathrm{d}$ first. We assume that we have a regularly spaced grid $\Omega_h \subset \Omega \cap h\mathbb{Z}^m$ $(h > 0)$ and a function $u_h : \Omega_h \to \mathbb{R}$. By $\{e_i\}_{i=1}^m$ we denote the canonical orthonormal basis of $\mathbb{R}^m$. Then we identify u_h with a function u that is constant on each cell $k + [0,h]^m$ $(k \in \Omega_h)$. Accordingly, we have

$$\mathrm{TV}^\varphi_\mathrm{d}(u_h) := \sum_{k\in\Omega_h}\sum_{i=1}^m h^{m-1}\varphi(|u_h(k+e_i) - u_h(k)|). \tag{1.4}$$

This discrete expression with $h = 1$ is essentially what is studied in [16, 33, 34], although [33] also studies more general discrete models. In the function space setting, this model has to be extended to all of $\mathrm{BV}(\Omega)$, in particular to smooth functions. The extension naturally has to be lower semicontinuous in a suitable topology, in order to guarantee the existence of solutions to (1.3). Therefore, one is naturally confronted with the question of whether such an extension can be performed meaningfully.

Let us consider a simple motivating example on $\Omega = (0,1)$ with $\varphi(t) = t^q$ for $q \in (0,1)$. We aim to approximate the ramp function

$$u(t) = t$$

by piecewise constant functions. Given $k > 0$, we thus define

$$u^k(t) = i/k \quad \text{for } t \in [(i-1)/k, i/k) \text{ and } i \in \{1, \dots, k\}.$$

Clearly we have that u^k converges strongly to u in $L^1(\Omega)$. Using the discrete model (1.4) with $h = 1/k$, one has

$$\mathrm{TV}^q_{\mathrm{d}}(u^k) = \sum_{i=1}^k (1/k)^0 \cdot (1/k)^q = k^{1-q}.$$

We see that $\lim_{k\to\infty} \mathrm{TV}^q_{\mathrm{d}}(u^k) = \infty$! This suggests that the TV^q model based on the "discrete" functional might allow only piecewise constant functions as solutions. In other words, $\mathrm{TV}^q_{\mathrm{d}}$ would induce pronounced staircasing—a property desirable when restoring piecewise constant images, but less suitable for other applications. In section 3, we will indeed demonstrate that either u is piecewise constant, or $u \notin \mathrm{BV}(\Omega)$.

In order to highlight the inherent difficulties, let us then consider the continuous model $\mathrm{TV}^\varphi_{\mathrm{c}}$, directly given by (1.1) for differentiable functions. In particular, (1.1) also serves as a definition of $\mathrm{TV}^\varphi_{\mathrm{c}}$ for continuous piecewise affine discretizations of $u \in C^1(\Omega)$. We observe that if $u \in C^1(\overline{\Omega})$ on a bounded domain Ω, and we set $u_h(k) = u(k)$ for $k \in \Omega_h$, then

$$\mathrm{TV}^\varphi_{\mathrm{c},h}(u_h) := \sum_{k\in\Omega_h} h^m \varphi(|\nabla_h u_h(k)|), \quad \text{with} \quad [\nabla_h u_h(k)]_i := \big(u_h(k+e_i) - u_h(k)\big)/h, \tag{1.5}$$

satisfies

$$\lim_{h\searrow 0} \mathrm{TV}^\varphi_{\mathrm{c},h}(u_h) = \mathrm{TV}^\varphi_{\mathrm{c}}(u).$$

This approximate model $\mathrm{TV}^\varphi_{\mathrm{c},h}$ with $h = 1$ is essentially what is considered in [27, 28, 36]. On an abstract level, it is also covered by [33]. The question now is whether the definition of $\mathrm{TV}^\varphi_{\mathrm{c}}$ can be extended to functions of bounded variation in a meaningful manner.

To start our investigation, let us try to approximate on $\Omega = (-1,1)$ the step function

$$u(t) = \begin{cases} 0, & t < 0, \\ 1, & t \geq 0. \end{cases}$$

Given $k > 0$, we define

$$u^k(t) = \begin{cases} 0, & t < -1/k, \\ 1, & t \geq 1/k, \\ (kt+1)/2, & t \in [-1/k, 1/k). \end{cases}$$

Then $u^k \to u$ in $L^1(\Omega)$. However, the continuous model (1.1) with $\varphi(t) = t^q$ for $q \in (0,1)$ gives

$$\mathrm{TV}^q_{\mathrm{c}}(u^k) = (2/k)^{q-1}.$$

Thus $\mathrm{TV}_c^q(u^k) \to 0$ as $k \nearrow \infty$. This suggests that any extension of TV_c^q to $u \in \mathrm{BV}(\Omega)$ through weak* lower semicontinuous envelopes will have $\mathrm{TV}_c^q(u) = 0$, and that jumps in u are mapped to 0 in general. In section 4 we will prove this rigorously and detect more striking properties. A weak* lower semicontinuous extension will necessarily satisfy $\mathrm{TV}_c^q \equiv 0$.

Despite this discouraging property, after discussing in section 5 the implications of the above-mentioned results, we find appropriate remedies. Our associated principal approach is given in section 6. It utilizes the (stronger) notion of *area-strict convergence* [18, 30], which—as will be shown—can be obtained using the multiscale analysis functional η from [44, 45]. In section 7 we also discuss alternative remedies which are related to compact operators and the space $\mathrm{SBV}(\Omega)$ of special functions of bounded variation. In order to maintain the flow of the paper, the pertinent proofs are relegated to the appendices.

To show existence of solutions to the fixed TV_c^φ model involving area-strict convergence, we require that φ be level coercive, i.e., $\lim_{t\to\infty} \varphi(t)/t > 0$. This induces a linear penalty to edges in the image. Based on these considerations, one arrives at the question of whether gradient statistics, such as those in [29], are reliable in dictating the prior term (regularizer). Our experiments on natural images in section 8 suggest that this is not the case. In fact, the jump part of the image appears to have different statistics from the smooth part. It seems that the conventional total variation regularization [40] provides a model for the jump part, which is superior to the nonconvex total variation model. This statistically validates our model, which is also suitable for a function space setting. Our rather theoretical starting point of making the TV_c^φ model sound in function space therefore leads to improved practical models. Finally, in section 9 we study image denoising with this model and finish with conclusions in section 10. We begin, however, with notation and other preliminary matters in section 2.

2. Notation and preliminaries.

We denote the set of nonnegative reals as $\mathbb{R}^{0,+} := [0, \infty)$. If $\varphi : \mathbb{R}^{0,+} \to \mathbb{R}^{0,+}$, then we write

$$\varphi_0 := \lim_{t \searrow 0} \varphi(t)/t \quad \text{and} \quad \varphi^\infty := \lim_{t \nearrow \infty} \varphi(t)/t,$$

implicitly assuming that the (possibly infinite) limits exist.

The notation $\|x\|$, without explicit specification of the space or type of norm, stands for the L^2-norm (in finite dimensions the 2-norm).

We write the boundary of a set A as ∂A, and the closure as $\overline{A}$. The open ball of radius ρ centered at $x \in \mathbb{R}^m$ is denoted by $B(x, \rho)$.

We now introduce some measure theory. For details of the definitions, we refer the reader to textbooks, including [4, 21, 41]. For $\Omega \subset \mathbb{R}^m$, we denote the space of (signed) Radon measures on Ω by $\mathcal{M}(\Omega)$, and the space of $\mathbb{R}^m$-valued Radon measures by $\mathcal{M}(\Omega; \mathbb{R}^m)$. We use the notation $|\mu|$ for the *total variation measure* of $\mu = (\mu_1, \ldots, \mu_m) \in \mathcal{M}(\Omega; \mathbb{R}^m)$, and define the *total variation (Radon) norm* of μ by

$$\|\mu\|_{\mathcal{M}(\Omega;\mathbb{R}^m)} := |\mu|(\Omega) := \sup \left\{ \sum_{i=1}^m \int_\Omega \varphi_i(x)\, d\mu_i(x) \;\middle|\; \varphi = (\varphi_1, \ldots, \varphi_m) \in C_c^\infty(\Omega; \mathbb{R}^m) \right\}. \tag{2.1}$$

Here $C_c^\infty(\Omega; \mathbb{R}^m)$ denotes the set of $\mathbb{R}^m$-valued smooth, infinitely differentiable functions φ with compact support $\operatorname{supp} \varphi \Subset \Omega$. If $\mu = f\mathcal{L}^m$ corresponds to a function $f \in L^1(\Omega)$, then the definition simplifies to $\|\mu\|_{\mathcal{M}(\Omega)} = \|\mu\|_{L^1(\Omega)}$.

For a measurable set A, we denote by $\mu \llcorner A$ the restricted measure defined by $(\mu \llcorner A)(B) := \mu(A \cap B)$. The restriction of a function u to A is denoted by $u|A$. On any given ambient space $\mathbb{R}^m$ $(k \le m)$, we write $\mathcal{H}^k$ for the k-dimensional *Hausdorff measure*, and $\mathcal{L}^m$ for the *Lebesgue measure*. We also define the *Dirac δ-measure at x* by

$$\delta_x(A) := \begin{cases} 1, & x \in A, \\ 0, & x \notin A. \end{cases}$$

Roughly, in the case $m = 2$ and $k = 1$, which is of most interest for us, $\mathcal{L}^2$ measures the area of sets in $\mathbb{R}^2$, while $\mathcal{H}^1$ measures the length of (collections of) curves embedded in $\mathbb{R}^2$. The Dirac δ measures membership of individual points and is different from $\mathcal{H}^0$, which would count the number of points in a set.

If $J \subset \mathbb{R}^m$ and there exist Lipschitz maps $\gamma_i : \mathbb{R}^{m-1} \to \mathbb{R}$ with

$$\mathcal{H}^{m-1}\left(J \setminus \bigcup_{i=1}^{\infty} \gamma_i(\mathbb{R}^{m-1}) \right) = 0,$$

then we say that J is *countably $\mathcal{H}^{m-1}$-rectifiable*. Again, with $m = 2$, this means roughly that J is contained in a collection of Lipschitz curves, modulo a negligible set. It may, however, happen that J is merely a point cloud contained in the curves and not a full curve itself.

We say that a function $u : \Omega \to \mathbb{R}$ on an open domain $\Omega \subset \mathbb{R}^m$ is of *bounded variation* (see, e.g., [4] for a thorough introduction), denoted $u \in \mathrm{BV}(\Omega)$, if $u \in L^1(\Omega)$, and the distributional gradient Du, given by

$$Du(\varphi) := \int_\Omega \operatorname{div} \varphi(x) u(x)\, dx \quad (\varphi \in C_c^\infty(\Omega)),$$

is a Radon measure. This is equivalent to asking that

$$\mathrm{TV}(u) := |Du|(\Omega) < \infty,$$

where the *total variation of u* is defined by setting $\mu = Du$ in (2.1). We can then decompose Du into

$$Du = \nabla u \mathcal{L}^n + D^j u + D^c u,$$

where $\nabla u \mathcal{L}^n$ is called the *absolutely continuous part*, $D^j u$ the *jump part*, and $D^c u$ the *Cantor part*. We also denote the *singular part* by

$$D^s u := D^j u + D^c u.$$

The density $\nabla u \in L^1(\Omega; \mathbb{R}^m)$ corresponds to the classical gradient if u is differentiable. The jump part may be written as

$$D^j u = (u^+ - u^-) \otimes \nu_{J_u} \mathcal{H}^{m-1} \llcorner J_u,$$

where the *jump set* J_u is countably $\mathcal{H}^{m-1}$-rectifiable, $\nu_{J_u}(x)$ is its normal, and u^+ and u^- are one-sided traces of u on J_u. These are defined by the condition

$$\lim_{\rho \searrow 0} \frac{1}{\rho^m} \int_{B^\pm(x,\rho,\nu)} |u^\pm(x) - u(y)|\, dy = 0$$

being satisfied for some normal vector $\nu \in \mathbb{R}^m$ and the associated half-ball

$$B^{\pm}(x,\rho,\nu) := \{y \in B(x,\rho) \mid \pm\langle y-x,\nu\rangle \geq 0\}.$$

The remaining *Cantor part* $D^c u$ of Du vanishes on any Borel set which is σ-finite with respect to $\mathcal{H}^{m-1}$; in particular, $|D^c u|(J_u) = 0$. We declare u an element of the space $\mathrm{SBV}(\Omega)$ of *special functions of bounded variation* if $u \in \mathrm{BV}(\Omega)$ and $D^c u = 0$.

We define the norm

$$\|u\|_{\mathrm{BV}(\Omega)} := \|u\|_{L^1(\Omega)} + \|Du\|_{\mathcal{M}(\Omega;\mathbb{R}^m)}.$$

We say that a sequence $\{u^i\}_{i=1}^\infty \subset \mathrm{BV}(\Omega)$ *converges weakly** to u in $\mathrm{BV}(\Omega)$, denoted by $u^i \overset{*}{\rightharpoonup} u$, if $u^i \to u$ strongly in $L^1(\Omega)$ and $Du^i \overset{*}{\rightharpoonup} Du$ weakly* in $\mathcal{M}(\Omega;\mathbb{R}^m)$. If, in addition, $|Du^i|(\Omega) \to |Du|(\Omega)$, we say that the convergence is *strict*. The weak* convergence of Du^i to Du may in this case be expressed as

$$\int_\Omega \operatorname{div}\varphi(x)\,du^i(x) \to \int_\Omega \operatorname{div}\varphi(x)u(x)\,dx \quad \text{for all } \varphi \in C_c^\infty(\Omega;\mathbb{R}^m).$$

It means roughly that we have convergence when we "test" the sequence by a sensor—the test function φ. We may, however, not have convergence on those aspects of the image u that cannot be sensed by such sensors. In particular, weak* convergence does not imply strong convergence, or even strict convergence. Strict convergence helps avoid annihilation effects. For example, the step functions $u^i = \chi_{[-1/i,1/i]}$, with $Du^i = \delta_{-1/i} - \delta_{1/i}$ and $|Du^i|(\Omega) = 2$, converge weakly to zero, but not strictly.

Finally, we say that a functional $F : \mathrm{BV}(\Omega) \to \mathbb{R}$ is weak* lower semicontinuous if $u^i \overset{*}{\rightharpoonup} u$ implies $F(u) \leq \liminf_{i\to\infty} F(u^i)$. This is an essential property for showing existence of minimizers of F: If u^i is an infimizing sequence, $F(u^i) \searrow \inf F > -\infty$, then u is a minimizer, provided that F is weak* lower semicontinuous.

3. Limiting aspects of the discrete TV^φ model. We begin by rigorously defining and analyzing the discrete TV^φ model (1.2) in $\mathrm{BV}(\Omega)$. This model is used in the literature to promote piecewise constant solutions to image reconstruction problems. For our analysis we consider the following class of energies φ.

Definition 3.1. *Define $\mathcal{W}_\mathrm{d}$ as the set of increasing, lower semicontinuous, subadditive functions $\varphi : \mathbb{R}^{0,+} \to \mathbb{R}^{0,+}$ that satisfy $\varphi(0) = 0$ and $\varphi_0 = \infty$.*

Example 3.1. Examples of $\varphi \in \mathcal{W}_\mathrm{d}$ include $\varphi(t) = t^q$ for $q \in [0,1)$. Note that this choice is the only one in $\mathcal{W}_\mathrm{d}$ of the classes considered, for example, in [34]. The logistic penalty $\varphi(t) = \log(\alpha t + 1)$ in particular, while subadditive, has $\varphi_0 = \alpha < \infty$. The fractional penalty $\varphi(t) = \alpha t/(1+\alpha t)$ likewise has $\varphi_0 = \alpha < \infty$. As we will later see, these classes are, however, admissible for the continuous model $\mathrm{TV}_\mathrm{c}^\varphi$.

Definition 3.2. *Denote by* $\mathrm{pwc}(\Omega)$ *the set of functions* $u \in \mathrm{BV}(\Omega)$ *that are piecewise constant in the sense* $Du = D^j u$. *We then write* $|D^j u| = \theta_u \mathcal{H}^{m-1} \llcorner J_u$, *where*

$$\theta_u(x) := |u^+(x) - u^-(x)|.$$

Definition 3.3. *Given an energy $\varphi \in \mathcal{W}_{\mathrm{d}}$, the "discrete" nonconvex total variation model is defined by*

$$\widetilde{\mathrm{TV}}^{\varphi}_{\mathrm{d}}(u) := \int_{J_u} \varphi(\theta_u(x))\, d\mathcal{H}^{m-1}(x) \quad (u \in \mathrm{pwc}(\Omega)),$$

and we extend this to $u \in \mathrm{BV}(\Omega)$ by defining

$$\mathrm{TV}^{\varphi}_{\mathrm{d}}(u) := \liminf_{\substack{u^i \overset{*}{\rightharpoonup} u,\\ u^i \in \mathrm{pwc}(\Omega)}} \widetilde{\mathrm{TV}}^{\varphi}_{\mathrm{d}}(u^i),$$

with the convergence weakly in $\mathrm{BV}(\Omega)$, in order to obtain a weak* lower semicontinuous functional.*

The functional $\widetilde{\mathrm{TV}}^{\varphi}_{\mathrm{d}}$ in particular agrees with (1.4). Our main result regarding this model is the following.

Theorem 3.4. *Let $\varphi \in \mathcal{W}_{\mathrm{d}}$. Then*

$$\mathrm{TV}^{\varphi}_{\mathrm{d}}(u) = \infty \quad \textit{for} \quad u \in \mathrm{BV}(\Omega) \setminus \mathrm{pwc}(\Omega).$$

The proof is based on the SBV compactness theorem [2]; alternatively, it can be proved via rectifiability results in the theory of currents [49], as used in the study of transportation networks, e.g., in [37, 43].

Theorem 3.5 (SBV compactness [2]). *Let $\Omega \subset \mathbb{R}^m$ be open and bounded. Suppose $\varphi, \psi : \mathbb{R}^{0,+} \to \mathbb{R}^{0,+}$ are lower semicontinuous and increasing with $\varphi^\infty = \infty$ and $\psi_0 = \infty$. Suppose $\{u^i\}_{i=1}^\infty \subset \mathrm{SBV}(\Omega)$ and $u^i \overset{*}{\rightharpoonup} u \in \mathrm{SBV}(\Omega)$ weakly* in $\mathrm{BV}(\Omega)$. If*

$$\sup_{i=1,2,3,\ldots} \left(\int_\Omega \varphi(|\nabla u^i(x)|)\, dx + \int_{J_{u^i}} \psi(\theta_{u^i}(x))\, d\mathcal{H}^{m-1}(x) \right) < \infty,$$

then there exists a subsequence of $\{u^i\}_{i=1}^\infty$, unrelabelled, such that

$$u^i \to u \text{ strongly in } L^1(\Omega), \tag{3.1}$$

$$\nabla u^i \rightharpoonup \nabla u \text{ weakly in } L^1(\Omega; \mathbb{R}^m), \tag{3.2}$$

$$D^j u^i \overset{*}{\rightharpoonup} D^j u \text{ weakly* in } \mathcal{M}(\Omega; \mathbb{R}^m). \tag{3.3}$$

If, moreover, ψ is subadditive, then

$$\int_{J_u} \psi(\theta_u(x))\, d\mathcal{H}^{m-1}(x) \le \liminf_{i\to\infty} \int_{J_{u^i}} \psi(\theta_{u^i}(x))\, d\mathcal{H}^{m-1}(x). \tag{3.4}$$

Proof. For the proof of (3.1)–(3.3), we refer the reader to [2, 4]. As the SBV compactness theorem is typically stated, concavity of ψ is required for the lower semicontinuity result (3.4). The fact that subadditivity and $\psi(0) = 0$ suffice follows from [4, Theorem 5.4]; see also [2]. There we use the fact that

$$\beta := \psi_0 = \varphi^\infty$$

in the application of the theorem to the functional

$$F(u) := \int_\Omega \varphi(|\nabla u(x)|)\,dx + \int_{J_u} \psi(\theta_u(x))\,d\mathcal{H}^{m-1}(x) + \beta|D^c u|(\Omega).$$

An alternative approach for the whole proof of the SBV compactness theorem is to map BV(Ω) into a space of Cartesian currents and use [49]. ■

Proof of Theorem 3.4. Given $u \in \mathrm{BV}(\Omega)$, let $u^i \in \mathrm{pwc}(\Omega)$ satisfy $u^i \overset{*}{\rightharpoonup} u$ weakly* in BV(Ω). Then the SBV compactness theorem shows that $\nabla u = \nabla u^i = 0$ and $D^c u = 0$. Thus $u \in \mathrm{pwc}(\Omega)$. ■

Remark 3.1. The functions $\varphi(t) = \alpha t/(1+\alpha t)$ and $\varphi(t) = \log(\alpha t + 1)$ for $\alpha > 0$, considered in [34] for reconstruction of piecewise constant images, do not have the property $\varphi(t)/t \to \infty$ as $t \searrow 0$. The above result therefore does not apply, and indeed $\mathrm{TV}_{\mathrm{d}}^{\varphi}$ defined using these functions will not force u with $\mathrm{TV}_{\mathrm{d}}^{\varphi}(u) < \infty$ to be piecewise constant, as the following result states.

Proposition 3.6. *Let $\varphi : \mathbb{R}^{0,+} \to \mathbb{R}^{0,+}$ be continuously differentiable and satisfy $\varphi(0) = 0$. Then the following hold.*

(i) *If $\varphi_0 < \infty$ and φ is subadditive, then there exists a constant $C > 0$ such that*

$$\mathrm{TV}_{\mathrm{d}}^{\varphi}(u) \le C\,\mathrm{TV}(u) \quad (u \in \mathrm{BV}(\Omega)).$$

(ii) *If $\varphi_0 > 0$ and φ is increasing, then for every $M > 0$ there exists also a constant $c = c(M) > 0$ such that*

$$c\,\mathrm{TV}(u) \le \mathrm{TV}_{\mathrm{d}}^{\varphi}(u) \quad (u \in \mathrm{BV}(\Omega),\ \|u\|_{L^\infty(\Omega)} \le M).$$

Proof. First we prove the upper bound. To begin with, we observe that

$$\varphi(t) \le \varphi_0 t. \tag{3.5}$$

Indeed, since φ is subadditive, we have

$$\lim_{\delta \searrow 0} \frac{\varphi(t+\delta) - \varphi(t)}{\delta} \le \lim_{\delta \searrow 0} \frac{\varphi(\delta)}{\delta} = \varphi_0 < \infty.$$

Thus $\varphi'(t) \le \varphi_0$. As $\varphi(0) = 0$, it follows that $\varphi(t) \le \varphi_0 t$.

Now, with $u \in \mathrm{BV}(\Omega)$, we pick a sequence $\{u^k\}_{k=1}^\infty$ in pwc(Ω) converging to u strictly in BV(Ω); for details see [12]. Then by (3.5) we have

$$\widetilde{\mathrm{TV}}_{\mathrm{d}}^{\varphi}(u^k) \le \varphi_0 \mathrm{TV}(u^k) \quad (k = 1, \dots, \infty).$$

Then, by the definition of $\mathrm{TV}_{\mathrm{d}}^{\varphi}(u)$ and the strict convergence,

$$\mathrm{TV}_{\mathrm{d}}^{\varphi}(u) \le \liminf_{k\to\infty} \widetilde{\mathrm{TV}}_{\mathrm{d}}^{\varphi}(u^k) \le \liminf_{k\to\infty} \varphi_0 \mathrm{TV}(u^k) = \varphi_0 \mathrm{TV}(u).$$

The claim in (i) follows.

Let us now prove the lower bound in (ii). First of all, we observe the existence of $c > 0$ with

$$\varphi(t) \geq ct \quad (0 \leq t \leq 2M). \tag{3.6}$$

Indeed, by the definition of φ_0, there exists $t_0 > 0$ such that $\varphi(t) > (\varphi_0/2)t$ for $t \in (0, t_0)$. Since φ is increasing, we have $\varphi(t) \geq \varphi(t_0) \geq (\varphi_0/2)t_0$ for $t \geq t_0$. This yields $c = \varphi_0 t_0/(4M)$.

Assuming that $\|u\|_{L^\infty(\Omega)} \leq M < \infty$, we now let $\{u_k\}_{k=1}^\infty \subset \mathrm{pwc}(\Omega)$ approximate u weakly* in $\mathrm{BV}(\Omega)$. We may assume that

$$\|u^k\|_{L^\infty(\Omega)} \leq 2M, \tag{3.7}$$

because if this did not hold, then we could truncate u^k, and the modified sequence $\{u^k_{2M}\}_{k=1}^\infty$ would still converge to u weakly* in $\mathrm{BV}(\Omega)$ with $\widetilde{\mathrm{TV}}{}^\varphi_{\mathrm{d}}(u^k_{2M}) \leq \widetilde{\mathrm{TV}}{}^\varphi_{\mathrm{d}}(u^k)$. Thanks to (3.6) and (3.7), we have

$$c\mathrm{TV}(u^k) \leq \widetilde{\mathrm{TV}}{}^\varphi_{\mathrm{d}}(u^k) \quad (k = 1, \ldots, \infty).$$

By the lower semicontinuity of $\mathrm{TV}(\cdot)$, we obtain

$$c\mathrm{TV}(u) \leq \liminf_{k\to\infty} \mathrm{TV}(u^k) \leq \liminf_{k\to\infty} \widetilde{\mathrm{TV}}{}^\varphi_{\mathrm{d}}(u^k).$$

Since the approximating sequence $\{u^k\}_{k=1}^\infty$ was arbitrary, the claim follows. ■

4. Limiting aspects of the continuous TV$^\varphi$ model.

We now consider the continuous model (1.1) or (1.5). Both are common in works aiming to model real image statistics. We initially restrict our attention to the following energies φ.

Definition 4.1. *We denote by $\mathcal{W}_{\mathrm{c}}$ the class of increasing, subadditive, continuous functions $\varphi : \mathbb{R}^{0,+} \to \mathbb{R}^{0,+}$ with $\varphi^\infty = 0$.*

Example 4.1. Examples of $\varphi \in \mathcal{W}_{\mathrm{c}}$ include in particular $\varphi(t) = t^q$ for $q \in (0, 1)$, as well as the fractional penalty $\varphi(t) = \alpha t/(1 + \alpha t)$ and the logistic penalty $\varphi(t) = \log(\alpha t + 1)$ for $\alpha > 0$.

Definition 4.2. *Given an energy φ, we start with the C^1 model (1.1), which we now denote by*

$$\widetilde{\mathrm{TV}}{}^\varphi_{\mathrm{c}}(u) := \int_\Omega \varphi(|\nabla u(x)|)\, dx \quad (u \in C^1(\Omega)).$$

In order to extend this to $u \in \mathrm{BV}(\Omega)$, we take the weak lower semicontinuous envelope*

$$\mathrm{TV}^\varphi_{\mathrm{c}}(u) := \liminf_{\substack{u^i \xrightarrow{*} u,\\ u^i \in C^1(\Omega)}} \widetilde{\mathrm{TV}}{}^\varphi_{\mathrm{d}}(u^i).$$

In the definition, the convergence is weakly in $\mathrm{BV}(\Omega)$.*

We emphasize that it is crucial to define $\mathrm{TV}^\varphi_{\mathrm{c}}$ through this limiting process in order to obtain weak* lower semicontinuity. This is useful in showing the existence of solutions to variational problems with the regularizer $\mathrm{TV}^\varphi_{\mathrm{c}}$ in $\mathrm{BV}(\Omega)$—or a larger space, as there is no guarantee that $\mathrm{TV}^\varphi_{\mathrm{c}}(u) < \infty$ would imply $u \in \mathrm{BV}(\Omega)$.

We may say the following about the $\mathrm{TV}_{\mathrm{c}}^{\varphi}$ model.

Theorem 4.3. *Let $\varphi \in \mathcal{W}_{\mathrm{c}}$, and suppose that $\Omega \subset \mathbb{R}^m$ has a Lipschitz boundary. Then*

$$\mathrm{TV}_{\mathrm{c}}^{\varphi}(u) = 0 \quad \textit{for} \quad u \in \mathrm{BV}(\Omega).$$

The result is intuitively obvious when one considers the convex envelope of the energy φ, which has to be zero. The rigorous verification of the result will, however, demand some amount of work. We note that this could be done using classical quasi-convexification arguments (see [6, Theorem 11.3.1]), using the integral sense of quasi convexity common in variational analysis, not the simple maximum sense common in optimization. The rigorous computation of the quasi-convex envelope of $x \mapsto \varphi(\|x\|)$, for all of the energies that we consider, would appear to be nearly as much work as a more informative direct argument. We therefore provide a self-contained proof, which will also provide new quantitive estimates for other energies. These will be provided in Proposition 4.6 later in this section.

The main ingredient of the proof of Theorem 4.3 is Lemma 4.5, which will utilize the following simple result.

Lemma 4.4. *Let $\varphi \in \mathcal{W}_{\mathrm{c}}$. Then there exist $a, b > 0$ such that*

$$\varphi(t) \le a + bt \quad (t \in \mathbb{R}^{0,+}).$$

Proof. Since $\varphi^\infty = 0$, we can find $t_0 > 0$ such that $\varphi(t)/t \le 1$ for $t \ge t_0$. Thus, because φ is increasing, we have $\varphi(t) \le \varphi(t_0) + t$ for every $t \in \mathbb{R}^{0,+}$. ■

Lemma 4.5. *Let $\varphi \in \mathcal{W}_{\mathrm{c}}$, and suppose that $\Omega \subset \mathbb{R}^m$ is bounded with Lipschitz boundary. Then*

$$\mathrm{TV}_{\mathrm{c}}^{\varphi}(u) \le \int_\Omega \varphi(|\nabla u(x)|)\, dx \quad (u \in \mathrm{BV}(\Omega)). \tag{4.1}$$

Observe the difference between Lemma 4.5 and Theorem 3.4. The former shows that in the limit of $\widetilde{\mathrm{TV}_{\mathrm{c}}^{\varphi}}$, the singular part is completely free, whereas the latter shows that in the limit of $\widetilde{\mathrm{TV}_{\mathrm{d}}^{\varphi}}$, only the jump part is allowed at all!

Proof. We may assume that

$$\int_\Omega \varphi(|\nabla u(x)|)\, dx < \infty,$$

because otherwise there is nothing to prove. We let $u_0 \in \mathrm{BV}(\mathbb{R}^m)$ denote the zero-extension of u from Ω to $\mathbb{R}^m$. Then

$$Du_0 = Du - \nu_{\partial\Omega} u^- \mathcal{H}^{n-1} \llcorner \partial\Omega$$

for u^-, the interior trace of u on $\partial\Omega$, and $\nu_{\partial\Omega}$, the exterior normal of Ω. In fact [4, section 3.7], there exists a constant $C = C(\Omega)$ such that

$$\|\nu_{\partial\Omega} u^- \mathcal{H}^{n-1} \llcorner \partial\Omega\|_{\mathcal{M}(\mathbb{R}^m;\mathbb{R}^m)} \le C\|u\|_{\mathrm{BV}(\Omega)}.$$

We pick some $\rho \in C_c^\infty(\mathbb{R}^m)$ with $0 \le \rho \le 1$, $\int \rho\, dx = 1$, and $\operatorname{supp} \rho \subset B(0,1)$. We then define the family of mollifiers $\rho_\epsilon(x) := \epsilon^{-n}\rho(x/\epsilon)$ for $\epsilon > 0$, and define by convolution and restriction of domain

$$u_\epsilon := (\rho_\epsilon * u_0)|\Omega.$$

Then $u_\epsilon \in C^\infty(\overline{\Omega})$, and $u_\epsilon \to u$ strongly in $L^1(\Omega)$ as $\epsilon \searrow 0$. As $|Du_\epsilon|(\omega) \le |Du_0|(\Omega)$, it follows that $u_\epsilon \overset{*}{\rightharpoonup} u$ weakly* in $\mathrm{BV}(\Omega)$; see, e.g., [4, Proposition 3.13]. Thus

$$\mathrm{TV}^\varphi_{\mathrm{c}}(u) \le \liminf_{\epsilon\searrow 0} \widetilde{\mathrm{TV}^\varphi_{\mathrm{c}}}(u_\epsilon).$$

In order to obtain the conclusion of the theorem, we just have to calculate the right-hand side.

We have

$$\begin{aligned} |\nabla u_\epsilon(x)| &= \left|\int_{\mathbb{R}^m} \rho_\epsilon(x-y)\, dDu_0(y)\right| \le \int_{\mathbb{R}^m} \rho_\epsilon(x-y)\, d|Du_0|(y) \\ &\le \int_{\mathbb{R}^m} \rho_\epsilon(x-y)|\nabla u_0(y)|\, dy + \int_{\mathbb{R}^m} \rho_\epsilon(x-y)\, d|D^s u_0|(y). \end{aligned} \tag{4.2}$$

We approximate the terms for the absolutely continuous and singular parts differently. Starting with the absolutely continuous part, we let K be a compact set such that $\Omega + B(0,1) \subset K$, and define

$$g_0(x) := |\nabla u_0(x)| \quad \text{and} \quad g_\epsilon(x) := \int_{\mathbb{R}^m} \rho_\epsilon(x-y)|\nabla u_0(y)|\, dy.$$

Then $g_\epsilon \to g_0$ in $L^1(K)$, and $g_\epsilon|(\mathbb{R}^m \setminus K) = 0$ for $\epsilon \in (0,1)$. By the L^1 convergence, we can find a sequence $\epsilon^i \searrow 0$ such that $g_{\epsilon^i} \to g_0$ almost uniformly. Consequently, given $\delta > 0$, we may find a set $E \subset K$ with $\mathcal{L}^m(K \setminus E) < \delta$ and $g_{\epsilon^i} \to g_0$ uniformly on E. We may assume that each ϵ^i is small enough such that

$$\|g_{\epsilon^i} - g_0\|_{L^1(K)} \le \delta. \tag{4.3}$$

Lemma 4.4 provides for some $a, b > 0$ the estimate

$$\varphi(t) \le a + bt. \tag{4.4}$$

From the uniform convergence on E, it follows that for large enough i, we have

$$\varphi(g_{\epsilon^i}(x)) \le \varphi(1 + g_0(x)) \le v(x) := a + b(1 + g_0(x)) \quad (x \in E).$$

Since $E \subset K$ is bounded, $v \in L^1(E)$. The reverse Fatou inequality on E gives the estimate

$$\limsup_{i\to\infty} \int_E \varphi(g_{\epsilon^i}(x))\, dx \le \int_E \limsup_{i\to\infty} \varphi(g_{\epsilon^i}(x))\, dx \le \int_E \varphi(g_0(x))\, dx. \tag{4.5}$$

On $K \setminus E$, we obtain the estimate

$$\begin{aligned} \int_{K\setminus E} \varphi(g_{\epsilon^i}(x))\, dx &\le \int_{K\setminus E} \varphi(g_0(x)) + \varphi(|g_{\epsilon^i}(x) - g_0(x)|)\, dx && \text{(by subadditivity)} \\ &\le \int_{K\setminus E} \varphi(g_0(x))\, dx + a\mathcal{L}^m(K \setminus E) + b\|g_{\epsilon^i} - g_0\|_{L^1(K)} && \text{(by (4.4))} \\ &\le \int_{K\setminus E} \varphi(g_0(x))\, dx + (a+b)\delta && \text{(by (4.3))}. \end{aligned} \tag{4.6}$$

Combining the estimates (4.5) and (4.6), we have

$$\limsup_{i\to\infty}\int_\Omega \varphi(g_{\epsilon^i}(x))\,dx \le \int_K \varphi(g_0(x))\,dx + (a+b)\delta.$$

Since $\delta > 0$ was arbitrary, and we may always find an almost uniformly convergent subsequence of any subsequence of $\{g_\epsilon\}_{\epsilon>0}$, we conclude that

$$\limsup_{\epsilon\searrow 0}\int_\Omega \varphi(g_\epsilon(x))\,dx \le \int_K \varphi(|\nabla u_0(x)|)\,dx = \int_\Omega \varphi(|\nabla u(x)|)\,dx. \tag{4.7}$$

Let us then consider the singular part in (4.2). We observe that $\int_{\mathbb{R}^m} \rho_\epsilon(x-y)\,d|D^s u_0(y)| = 0$, for $x \in \mathbb{R}^m \setminus K$. If we define

$$f_\epsilon(x) := \epsilon^{-m}|D^s u_0|(B(x,\epsilon)) \quad (x \in K),$$

then by Fubini's theorem

$$\begin{aligned}\int_K f_\epsilon(x)\,dx &= \epsilon^{-m}\int_K\int_K \chi_{B(x,\epsilon)}(y)\,d|D^s u_0|(y)\,dx = \epsilon^{-m}\int_K\int_K \chi_{B(y,\epsilon)}(x)\,dx\,d|D^s u_0|(y)\\ &\le \omega_m |D^s u_0|(K).\end{aligned} \tag{4.8}$$

Here ω_m is the volume of the unit ball in $\mathbb{R}^m$. Moreover, by the Besicovitch derivation theorem (discussed, e.g., in [4, 32]), we have

$$\lim_{\epsilon\searrow 0} f_\epsilon(x) = 0 \quad (\mathcal{L}^m\text{-a.e. } x \in K).$$

Because $\mathcal{L}^m(K) < \infty$, Egorov's theorem shows that $f_\epsilon \to 0$ almost uniformly. Thus, for any $\delta > 0$, there exists a set $K_\delta \subset K$ with $\mathcal{L}^m(K \setminus K_\delta) \le \delta$ and $f_\epsilon \to 0$ uniformly on K_δ.

Next we study $K \setminus K_\delta$. We pick an arbitrary $\sigma > 0$. Because $\varphi(t)/t \to 0$ as $t \to \infty$, there exists $t_0 > 0$ such that $\varphi(t) \le \sigma t$ for $t \ge t_0$. In fact, because φ is lower semicontinuous and $\varphi(0) = 0$, if we choose

$$t_0 := \inf\{t \ge 0 \mid \varphi(t) < \sigma t\},$$

then $\varphi(t_0) = \sigma t_0$. Thus, because φ is increasing,

$$\varphi(t) \le \widetilde{\varphi}(t) := \sigma(t_0 + t) \quad (t \in \mathbb{R}^{0,+}). \tag{4.9}$$

Choosing $\epsilon \in (0,1)$ such that $f_\epsilon \le \delta$ on K_δ, and using $\rho_\epsilon \le \epsilon^{-m}\chi_{B(0,\epsilon)}$, we may approximate

$$\begin{aligned}\int_{\mathbb{R}^m} \varphi\left(\int_{\mathbb{R}^m} \rho_\epsilon(x-y)\,d|D^s u_0|(y)\right)dx &\le \int_{\mathbb{R}^m} \varphi\left(f_\epsilon(x)\right)\,dx\\ &\le \int_{K_\delta} \varphi\left(f_\epsilon(x)\right)\,dx + \int_{K\setminus K_\delta} \widetilde{\varphi}\left(f_\epsilon(x)\right)\,dx\\ &\le \int_{K_\delta} \varphi(\delta)\,dx + \int_{K\setminus K_\delta} \sigma\left(t_0 + f_\epsilon(x)\right)\,dx\\ &\le \mathcal{L}^m(K)\varphi(\delta) + \delta\sigma t_0 + \sigma\omega_m|D^s u_0|(K) \qquad (\text{by } (4.8)).\end{aligned} \tag{4.10}$$

Thus

$$\liminf_{\epsilon\searrow 0}\int_{\mathbb{R}^m}\varphi\left(\int_{\mathbb{R}^m}\rho_\epsilon(x-y)\,d|D^s u_0|(y)\right)dx \le \mathcal{L}^m(K)\varphi(\delta)+\delta\sigma t_0+\sigma\omega_m|D^s u_0|(K).$$

Observe that the choices of σ and t_0 are independent of δ. Therefore, because $\delta > 0$ was arbitrary, using the continuity of φ we deduce that we may set $\delta = 0$ above. But then, because $\sigma > 0$ was also arbitrary, we deduce that

$$\lim_{\epsilon\searrow 0}\int_{\mathbb{R}^m}\varphi\left(\int_{\mathbb{R}^m}\rho_\epsilon(x-y)\,d|D^s u_0|(y)\right)dx = 0. \tag{4.11}$$

Finally, combining the estimate (4.7) for the absolutely continuous part and the estimate (4.11) for the singular part with (4.2), we deduce that

$$\limsup_{\epsilon\searrow 0}\widetilde{\mathrm{TV}_\mathrm{c}^\varphi}(u_\epsilon)=\limsup_{\epsilon\searrow 0}\int_{\mathbb{R}^m}\varphi\left(\int_{\mathbb{R}^m}\rho_\epsilon(x-y)\,d|Du_0|(y)\right)dx \le \int_\Omega \varphi(|\nabla u(x)|)\,dx.$$

This concludes the proof of (4.1). ■

Proof of Theorem 4.3. We employ the bound (4.1) of Lemma 4.5, but still have to extend it to a possibly unbounded domain Ω. For this purpose, we let $R > 0$ be arbitrary and apply the lemma to $u_R := u|B(0,R)$. Then

$$\mathrm{TV}_\mathrm{c}^\varphi(u_R)\le\int_\Omega\varphi(|\nabla u_R(x)|)\,dx\le\int_\Omega\varphi(|\nabla u(x)|)\,dx.$$

But $u_R \overset{*}{\rightharpoonup} u$ weakly* in $\mathrm{BV}(\Omega)$ as $R \nearrow \infty$; indeed, L^1 convergence is obvious, and for any $\varphi \in C_c^\infty(\Omega;\mathbb{R}^m)$, we have $\operatorname{supp}\varphi \in B(0,R)$ for large enough R, so that $Du_R(\varphi) = Du(\varphi)$. Therefore, because $\mathrm{TV}_\mathrm{c}^\varphi$ is weakly* lower semicontinuous by construction, we conclude that

$$\mathrm{TV}_\mathrm{c}^\varphi(u)\le\int_\Omega\varphi(|\nabla u(x)|)\,dx. \tag{4.12}$$

Given any $u \in C^1(\Omega)$, we may find $u_h \in \mathrm{pwc}(\Omega)$ $(h > 0)$ strictly convergent to u in $\mathrm{BV}(\Omega)$ [12]. But (4.12) shows that

$$\mathrm{TV}_\mathrm{c}^\varphi(u_h)=0.$$

By the weak* lower semicontinuity of $\mathrm{TV}_\mathrm{c}^\varphi$ we conclude that

$$\mathrm{TV}_\mathrm{c}^\varphi(u)\le\liminf_{h\searrow 0}\mathrm{TV}_\mathrm{c}^\varphi(u_h)=0\quad(u\in C^1(\Omega)).$$

Another referral to lower semicontinuity now shows that $\mathrm{TV}_\mathrm{c}^\varphi(u) = 0$ for any $u \in \mathrm{BV}(\Omega)$. ■

Similarly to Proposition 3.6 for $\mathrm{TV}_\mathrm{d}^\varphi$, we have the following more positive result.

Proposition 4.6. *Let $\varphi : \mathbb{R}^{0,+} \to \mathbb{R}^{0,+}$ be lower semicontinuous and satisfy $\varphi(0) = 0$. Then the following hold.*

(i) *If $\varphi_0 < \infty$ and φ is subadditive, then there exists a constant $C > 0$ such that*

$$\mathrm{TV}_\mathrm{c}^\varphi(u)\le C\,\mathrm{TV}(u)\quad(u\in\mathrm{BV}(\Omega)).$$

(ii) *If $\varphi_0, \varphi^\infty > 0$ and φ is increasing, then there also exists a constant $c > 0$ such that*

$$c\,\mathrm{TV}(u) \le \mathrm{TV}_{\mathrm{c}}^{\varphi}(u) \quad (u \in \mathrm{BV}(\Omega)).$$

Remark 4.1. If we assume that φ is concave, the condition $\varphi_0 > 0$ in (ii) follows from the other assumptions.

Proof. The proof of the upper bound follows in exactly the same way as the upper bound in Proposition 3.6, just replacing approximation in $\mathrm{pwc}(\Omega)$ by $C^1(\Omega)$.

For the lower bound, first we observe that there exists $t^\infty > 0$ such that $\varphi(t) \ge (\varphi^\infty/2)t$ $(t \ge t^\infty)$. Second, there exists $t_0 > 0$ such that $\varphi(t) \ge (\varphi_0/2)t$ $(0 \le t \le t_0)$. Since φ is increasing, $\varphi(t) \ge \varphi(t_0) \ge t\varphi(t_0)/t_0$ $(t_0 \le t \le t_\infty)$. Consequently

$$\varphi(t) \ge ct \quad (t \ge 0) \quad \text{for } c := \min\{\varphi^\infty/2, \varphi(t_0)/t_\infty, \varphi^0/2\}.$$

Therefore

$$c\mathrm{TV}(u') \le \widetilde{\mathrm{TV}}_{\mathrm{c}}^{\varphi}(u') \quad (u' \in C^1(\Omega)).$$

The claim now follows from the weak* lower semicontinuity of TV as in the proof of Proposition 3.6. ■

In fact, in cases of interest to us we may prove a slightly stronger result.

Theorem 4.7. *Let $\varphi : \mathbb{R}^{0,+} \to \mathbb{R}^{0,+}$ be concave with $\varphi(0) = 0$ and $0 < \varphi^\infty < \infty$. Suppose that $\Omega \subset \mathbb{R}^m$ has a Lipschitz boundary. Then*

$$\mathrm{TV}_{\mathrm{c}}^{\varphi}(u) = \varphi^\infty \mathrm{TV}(u) \quad (u \in \mathrm{BV}(\Omega)). \tag{4.13}$$

Proof. We first suppose that Ω is bounded. The proof of the upper bound,

$$\mathrm{TV}_{\mathrm{c}}^{\varphi}(u) \le \int_\Omega \varphi(|\nabla u(x)|)\,dx + \varphi^\infty |D^s u|(\Omega), \tag{4.14}$$

is then a modification of Lemma 4.5. The estimate (4.7) for the absolutely continuous part follows as before. For the singular part, we observe that (4.9) holds for any $\sigma > \varphi^\infty$. Therefore, proceeding as before, we obtain in place of (4.11) the estimate

$$\lim_{\epsilon \searrow 0} \int_{\mathbb{R}^m} \varphi\left(\int_{\mathbb{R}^m} \rho_\epsilon(x-y)\,d|D^s u_0|(y)\right) dx \le \sigma |D^s u|(\Omega). \tag{4.15}$$

Letting $\sigma \searrow \varphi^\infty$ and combining (4.7) with (4.15), we get (4.14). As in Proposition 4.6, we may extend this bound to a possibly unbounded Ω.

If $u \in C^1(\Omega)$, we may again approximate u strictly in $\mathrm{BV}(\Omega)$ by piecewise constant functions $\{u^i\}_{i=1}^\infty$. By the lower semicontinuity of $\mathrm{TV}_{\mathrm{c}}^{\varphi}$ and (4.14), we then have

$$\mathrm{TV}_{\mathrm{c}}^{\varphi}(u) \le \liminf_{i\to\infty} \varphi^\infty |D^s u|(\Omega) = \varphi^\infty |Du|(\Omega). \tag{4.16}$$

Finally, we observe that by concavity

$$\varphi(t) \ge \varphi^\infty t.$$

Thus $\widetilde{\mathrm{TV}}_{\mathrm{c}}^{\varphi}(u) \ge \varphi^\infty |Du|(\Omega)$. We immediately obtain (4.13) for $u \in C^1(\Omega)$. By strictly convergent approximation, we then extend the result to $u \in \mathrm{BV}(\Omega)$. ■

5. Discussion. Theorems 4.3 and 4.7 show that we cannot hope to have a simple weakly* lower semicontinuous nonconvex total variation model as a prior for image gradient distributions. In fact, it follows from [9] (see also [4, section 5.1] and [22, Theorem 5.14]) that lower semicontinuity of the continuous $\mathrm{TV}_\mathrm{c}^\varphi$ model is possible only for convex φ. The problem is that if φ^∞ is less than $\varphi'(t)$, then image edges are always cheaper than smooth transitions. If $\varphi^\infty = 0$, they are so cheap that we get a zero functional at the limit for a general class of functions. If $\varphi^\infty > 0$ and φ is concave, then we get a factor of TV as the result. If φ is not concave, we still have the upper bound (4.16); it may, however, be possible that some gradients are cheaper than jumps. This would in particular be the case with Huber regularization of $\varphi(t) = t$. More about the jump set of solutions to Huber-regularized as well as nonconvex total variation models may be found in [47].

In fact, in [27] Huber regularization was used with $\varphi(t) = t^q$ for $q \in (0,1)$ for algorithmic reasons. For small $\gamma > 0$, this is defined as

$$\widetilde{\varphi}(t) := \begin{cases} t^q - \frac{2-q}{2}\gamma^q, & t > \gamma, \\ \frac{q}{2}\gamma^{q-2}t^2, & t \in [0,\gamma]. \end{cases} \tag{5.1}$$

Then $\widetilde{\varphi}(t) \leq \varphi(t)$, so that

$$\mathrm{TV}_\mathrm{c}^{\widetilde{\varphi}} \leq \mathrm{TV}_\mathrm{c}^\varphi = 0.$$

The asymptotic behavior of the regularizer at infinity is the crucial feature here, and it also cannot be altered without changing the edge behavior of the regularizer. Huber regularization does not alter the asymptotic behavior, and as long as alternative smoothing strategies, such as those considered in [15], do not, they also provide no change in the results.

In contrast to the continuous $\mathrm{TV}_\mathrm{c}^\varphi$ model, according to Theorem 3.4, the discrete model works correctly for $\varphi(t) = t^q$ and generally $\varphi \in \mathcal{W}_\mathrm{d}$ if the desire is to force piecewise constant solutions to (1.3). As we saw in the comments preceding Proposition 3.6, it does not, however, force piecewise constant solutions for some of the energies φ typically employed in this context. Generally, what causes piecewise constant solutions is the property $\varphi^0 = \infty$. If one does not desire piecewise constant solutions, one can therefore use Huber regularization or linearize φ for $t < \delta$. The latter employs

$$\widetilde{\varphi}(t) = \begin{cases} \varphi(t) - \varphi(\delta) + \varphi'(\delta)\delta, & t > \delta, \\ \varphi'(\delta)t, & t \leq \delta. \end{cases}$$

Then $\varphi(t) \leq Ct$ for some $C > 0$, so that $\mathrm{TV}_\mathrm{d}^\varphi(u) < \infty$ for every $u \in \mathrm{BV}(\Omega)$. We also note that although this approach defines a regularization functional on all of $\mathrm{BV}(\Omega)$, it cannot be used for modelling the distribution of gradients in real images, the purpose of the $\mathrm{TV}_\mathrm{c}^\varphi$ model. In fact, as in the $\mathrm{TV}_\mathrm{d}^\varphi$ model, we cannot control the penalization of ∇u beyond a constant factor.

In summary, the $\mathrm{TV}_\mathrm{d}^\varphi$ model works as intended for $\varphi \in \mathcal{W}_\mathrm{d}$—it enforces piecewise constant solutions. The $\mathrm{TV}_\mathrm{c}^\varphi$ model, however, is not theoretically sound in function spaces. We will therefore seek ways to fix it in section 6.

6. Multiscale regularization and area-strict convergence.

The problem with the $\mathrm{TV}_{\mathrm{c}}^{\varphi}$ model is that weak* lower semicontinuity is too strong a requirement. We need a weaker type of lower semicontinuity or, in other words, a stronger type of convergence. Norm convergence in $\mathrm{BV}(\Omega)$ is too strong; it would not be at all possible to approximate edges. Strict convergence is also still too weak, as can be seen from the proof of Lemma 4.5. Strong convergence in L^2, which we could, in fact, obtain from strict convergence for $\Omega \subset \mathbb{R}^2$ (see [31, 39]), is also not enough, as a stronger form of gradient convergence is the important part. A suitable mode of convergence is the so-called *area-strict convergence* [18, 30]. For our purposes, the following definition is the most appropriate one.

Definition 6.1. *Suppose $\Omega \subset \mathbb{R}^n$ with $n \geq 2$. The sequence $\{u^i\}_{i=1}^{\infty} \subset \mathrm{BV}(\Omega)$ converges to $u \in \mathrm{BV}(\Omega)$* area-strictly *if the sequence $\{U^i\}_{i=1}^{\infty}$ with $U^i(x) := (x/|x|, u^i(x))$ converges strictly in $\mathrm{BV}(\Omega; \mathbb{R}^{n+1})$ to $U(x) := (x/|x|, u(x))$.*

In other words, $\{u^i\}_{i=1}^{\infty}$ converges to u area-strictly if $u^i \to u$ strongly in $L^1(\Omega)$, $Du^i \overset{*}{\rightharpoonup} Du$ weakly* in $\mathcal{M}(\Omega; \mathbb{R}^n)$, and $\mathcal{A}(u^i) \to \mathcal{A}(u)$ for the *area functional*

$$\mathcal{A}(u) := \int_{\Omega} \sqrt{1 + |\nabla u(x)|^2}\, dx + |D^s u|(\Omega).$$

It can be shown that area-strict convergence is stronger than strict convergence, but weaker than norm convergence. Here we recall from section 2 the definitions of strict convergence and the singular part $D^s u$ of Du.

In order to state a continuity result with respect to area-strict convergence, we need a few definitions. Specifically, we denote the Sobolev conjugate

$$1^* := \begin{cases} n/(n-1), & n > 1, \\ \infty, & n = 1, \end{cases}$$

and define

$$u^{\theta}(x) := \begin{cases} \theta u^+(x) + (1-\theta) u^-(x), & x \in J_u, \\ \widetilde{u}(x), & x \notin S_u. \end{cases}$$

In [39] (see also [30]), the following result is proved.

Theorem 6.2. *Let Ω be a bounded domain with Lipschitz boundary, $p \in [1, 1^*]$ if $n \geq 2$ and $p \in [1, 1^*)$ if $n = 1$. Let $f \in C(\Omega \times \mathbb{R} \times \mathbb{R}^n)$ satisfy*

$$|f(x, y, A)| \leq C(1 + |y|^p + |A|) \quad ((x, y, A) \in \Omega \times \mathbb{R} \times \mathbb{R}^n),$$

and assume the existence of $f^{\infty} \in C(\Omega \times \mathbb{R} \times \mathbb{R}^n)$, defined by

$$f^{\infty}(x, y, A) := \lim_{\substack{x' \to x \\ y' \to y \\ A' \to A \\ t \to \infty}} \frac{f(x', y', tA')}{t}. \tag{6.1}$$

Then the functional

$$\mathcal{F}(u) := \int_{\Omega} f(x, u(x), \nabla u(x))\, dx + \int_{\Omega} \int_0^1 f^{\infty}\left(x, u^{\theta}(x), \frac{dD^s u}{d|D^s u|}(x)\right) d|D^s u|(x)$$

is area-strictly continuous on $\mathrm{BV}(\Omega)$.

We now apply this mode of convergence to nonconvex total variation, restricting our attention to the following class of functions.

Definition 6.3. *We denote by* $\mathcal{W}_{\mathrm{as}}$ *the set of functions* $\varphi \in C(\mathbb{R}^{0,+})$ *such that* φ^∞ *exists, and for some* $c, C > 0$ *and* $b \geq 0$ *the following estimates hold true:*

$$ct - b \leq \varphi(t) \leq C(1+t) \quad (t \in \mathbb{R}^{0,+}). \tag{6.2}$$

Example 6.1. Let φ be any of the functions in Example 4.1. They do not satisfy the lower bound in (6.2). If, however, we pick some cut-off $M > 0$, then $\varphi_M \in \mathcal{W}_{\mathrm{as}}$ for the high-value linearization

$$\varphi_M(t) := \begin{cases} \varphi(t), & t \leq M, \\ \varphi(M) + \varphi'(M)(t-M), & t > M. \end{cases}$$

Corollary 6.4. *Suppose* $\varphi \in \mathcal{W}_{\mathrm{as}}$. *Then the functional*

$$\mathrm{TV}^\varphi_{\mathrm{as}}(u) := \int_\Omega \varphi(|\nabla u(x)|)\,dx + \varphi^\infty |D^s u|(\Omega) \quad (u \in \mathrm{BV}(\Omega))$$

is area-strictly continuous on $\mathrm{BV}(\Omega)$.

Proof. Letting $f(x,y,A) := \varphi(A)$, we verify (6.1). The claim is therefore immediate from Theorem 6.2 with $p = 1$. Note that we do not need the lower bound in the definition of $\mathcal{W}_{\mathrm{as}}$ just yet. ∎

But how could we obtain area-strict convergence of an infimizing sequence of a variational problem? In [44, 45] the following multiscale analysis functional η was introduced for scalar-valued measures $\mu \in \mathcal{M}(\Omega)$. Given $\eta_0 > 0$ and $\{\rho_\epsilon\}_{\epsilon>0}$, a family of mollifiers satisfying the semigroup property $\rho_{\epsilon+\delta} = \rho_\epsilon * \rho_\delta$, η can be defined as

$$\eta(\mu) := \eta_0 \sum_{\ell=1}^{\infty} \int_{\mathbb{R}^n} (|\mu| * \rho_{2^{-i}})(x) - |\mu * \rho_{2^{-i}}|(x)\,dx \quad (\mu \in \mathcal{M}(\Omega)).$$

If the sequence of measures $\{\mu^i\}_{i=1}^\infty \subset \mathcal{M}(\Omega)$ satisfies $\sup_i \eta(\mu) < \infty$ and $\mu^i \overset{*}{\rightharpoonup} \mu$ weakly* in $\mathcal{M}(\Omega)$, then we have $|\mu^i|(\Omega) \to |\mu|(\Omega)$. In essence, the functional η penalizes the type of complexity of measures such as two approaching δ-spikes of different sign, which prohibits strict convergence. In Appendix A, we extend the strict convergence results of [44, 45] to vector-valued $\mu \in \mathcal{M}(\Omega; \mathbb{R}^N)$, in particular, the case $\mu = DU$ for U the lifting of u as defined in Definition 6.1.

In order to bound in $\mathrm{BV}(\Omega)$ an infimizing sequence of problems using $\mathrm{TV}^\varphi_{\mathrm{as}}$ as a regularizer, we require slightly stricter assumptions on φ. These can usually, and particularly in the interesting case $\varphi(t) = t^q$, be easily satisfied by linearizing φ above a cut-off point M with respect to the function value. This will force $\varphi^\infty > 0$, which is not required for continuity with respect to area-strict convergence in its own right. We will later see that such a cut-off can be justified by real gradient distributions and also argued in numerical experiments.

Now we may prove the following result, which shows that area-strict convergence and the multiscale analysis functional η provide a remedy for the theoretical difficulties associated with the $\mathrm{TV}^{\varphi}_{c}$ model.

Theorem 6.5. *Suppose $\Omega \subset \mathbb{R}^n$ is bounded with Lipschitz boundary, and $\varphi \in \mathcal{W}_{\mathrm{as}}$. Define $U(x) := (1, u(x))$. Then the functional*

$$F(u) := \mathrm{TV}^{\varphi}_{\mathrm{as}}(u) + \eta(DU)$$

is weak lower semicontinuous on $\mathrm{BV}(\Omega)$, and any sequence $\{u^i\}_{i=1}^{\infty} \subset L^1(\Omega)$ with*

$$\sup_i F(u^i) < \infty$$

admits an area-strictly convergent subsequence.

Proof. Suppose $\{u^i\}_{i=1}^{\infty}$ converges weakly* to $u \in \mathrm{BV}(\Omega)$. Then it follows that $\{U^i\}_{i=1}^{\infty}$ converge weakly* to $U \in \mathrm{BV}(\Omega; \mathbb{R}^{m+1})$. If $\liminf_{i\to\infty} \eta(DU^i) = \infty$, we clearly have lower semicontinuity of F. By switching to an unrelabelled subsequence, we may therefore assume that $\sup_i \eta(DU^i) < \infty$. It follows from Theorem A.4 in Appendix A that $|DU^i|(\Omega) \to |DU|(\Omega)$. In other words, $\{u^i\}_{i=1}^{\infty}$ converges area-strictly to u. Applying Corollary 6.4 and the weak* lower semicontinuity of η, we now see that

$$F(u) \leq \liminf_{i\to\infty} F(u^i).$$

Thus weak* lower semicontinuity holds true.

Next suppose that $\{u^i\}_{i=1}^{\infty} \subset L^1(\Omega)$ with $\sup_i F(u^i) < \infty$. Since $ct - b \leq \varphi(t)$ and Ω is bounded, it follows that $\sup_i \mathrm{TV}(u^i) < \infty$. The sequence therefore admits a subsequence, unrelabelled without loss of generality, which converges weakly* to some $u \in \mathrm{BV}(\Omega)$. Hence, the fact that $\{u^i\}_{i=1}^{\infty}$ admits an area-strictly convergent subsequence now follows as in the previous paragraph. ■

We immediately deduce the following corollary.

Corollary 6.6. *Suppose $\Omega \subset \mathbb{R}^n$ is bounded with Lipschitz boundary, $\varphi \in \mathcal{W}_{\mathrm{as}}$, $J : \mathrm{BV}(\Omega) \to \mathbb{R}$ is convex, proper, and weakly* lower semicontinuous, and J satisfies for some $C > 0$ the coercivity condition*

$$J(u) \geq C(\|u\|_{L^1(\Omega)} - 1).$$

Then the functional

$$G(u) := J(u) + \alpha \mathrm{TV}^{\varphi}_{\mathrm{as}}(u) + \eta(DU) \quad (u \in \mathrm{BV}(\Omega)) \tag{6.3}$$

admits a minimizer $u \in \mathrm{BV}(\Omega)$.

Remark 6.1. We can, for example, take $J(u) = \frac{1}{2}\|z - u\|^2_{L^2(\Omega)}$.

Observe that

$$\eta(DU) = \eta_0 \sum_{\ell=1}^{\infty} \eta_\ell(DU),$$

where, for $\epsilon_\ell > 0$,

$$\eta_\ell(DU) := \int_{\mathbb{R}^n} (\rho_{\epsilon_\ell} * |DU|)(x) - |\rho_{\epsilon_\ell} * DU|(x)\,dx$$
$$= |D^s u|(\Omega) + \int_{\mathbb{R}^n} \sqrt{1+|\nabla u(x)|^2} - \sqrt{1+|(\rho_{\epsilon_\ell} * Du)(x)|^2}\,dx.$$

In particular, if $u \in W^{1,1}(\Omega)$, then we obtain with $\nabla_\epsilon u := \rho_\epsilon * \nabla u$ the expression

$$\eta_\ell(DU) = \int_{\mathbb{R}^n} \sqrt{1+|\nabla u(x)|^2} - \sqrt{1+|\nabla_{\epsilon_\ell} u(x)|^2}\,dx$$

and the estimate

$$\eta_\ell(DU) \le \int_{\mathbb{R}^n} \sqrt{\big||\nabla u(x)|^2 - |\nabla_{\epsilon_\ell} u(x)|^2\big|}\,dx.$$

The following proposition shows that, in infimizing sequences, we may ignore terms from η. This justifies the associated numerical approximation.

Proposition 6.7. *Suppose that $\Omega \subset \mathbb{R}^n$ is bounded with Lipschitz boundary, $\varphi \in \mathcal{W}_{\mathrm{as}}$, and that $J : \mathrm{BV}(\Omega) \to \mathbb{R}$ is as in Corollary 6.6. Let $K_i \in \mathbb{N}^+$ and $\epsilon_i > 0$, $i = 1, 2, 3, \dots,$ satisfy*

$$\lim_{i\to\infty} K_i = \infty \quad \textit{and} \quad \lim_{i\to\infty} \epsilon_i = 0.$$

Suppose further that $\{u^i\}_{i=1}^\infty \subset \mathrm{BV}(\Omega)$ satisfies

$$J(u^i) + \alpha \mathrm{TV}^\varphi_{\mathrm{as}}(u^i) + \sum_{\ell=1}^{K_i} \eta_\ell(DU^i) \le \inf_{u\in\mathrm{BV}(\Omega)} G(u) + \epsilon_i \quad (i = 1, 2, 3, \dots).$$

Then we can find $\hat{u} \in \mathrm{BV}(\Omega)$ and a subsequence of $\{u^i\}_{i=1}^\infty$, unrelabelled, such that $u^i \to \hat{u}$ area-strictly, and $\hat{u}$ minimizes G.

Proof. Let $L := \inf_{u\in\mathrm{BV}(\Omega)} G(u)$. Since there is nothing to prove if $L = \infty$, we may assume $L < \infty$. Then we have

$$c\mathrm{TV}(u^i) - b\mathcal{L}^n(\Omega) \le \mathrm{TV}^\varphi_{\mathrm{as}}(u^i).$$

This yields

$$J(u^i) + \alpha c \mathrm{TV}(u^i) \le \alpha b \mathcal{L}^n(\Omega) + L + \epsilon_i.$$

It follows for a subsequence, unrelabelled, that $u^i \xrightarrow{*} \hat{u}$ weakly* for some $\hat{u} \in \mathrm{BV}(\Omega)$. By the weak* lower semicontinuity of η_ℓ (see Theorem A.4), we then have

$$\sum_{\ell=1}^{K_j} \eta_\ell(D\hat{U}) \le \liminf_{i\to\infty} \sum_{\ell=1}^{K_j} \eta(DU^i) \le L \quad (j = 1, 2, 3, \dots).$$

It follows that

$$\eta(D\hat{U}) = \sum_{\ell=1}^{\infty} \eta_\ell(D\hat{U}) \le L.$$

Using Lemma A.3 with $\mu^i = DU^i$ and $\mu = DU$, we see that $u^i \to \hat{u}$ area-strictly and that

$$u \mapsto J(u) + \alpha \mathrm{TV}^{\varphi}_{\mathrm{as}}(u) + \sum_{\ell=1}^{K_j} \eta_\ell(DU)$$

is area-strictly lower semicontinuous for each fixed $j = 1, 2, 3, \ldots$. This shows that $G(\hat{u}) \leq L$. As a consequence, $\hat{u}$ minimizes G. ■

7. Remarks on alternative remedies.

We now discuss two alternative approaches to make the $\mathrm{TV}^{\varphi}_{\mathrm{c}}$ model work in the limit. These are based on compactifying the differential operator and on working in $\mathrm{SBV}(\Omega)$, respectively. As we only intend to demonstrate alternative possibilities, we keep our remarks brief here. Hence the proofs have been placed in the appendices.

Remark 7.1 (compact operators). Area-strict convergence is not the only possibility for making the $\mathrm{TV}^{\varphi}_{\mathrm{c}}$ model function; another way to understand the problems with the basic $\mathrm{TV}^{\varphi}_{\mathrm{c}}$ model is that the operator ∇ is not compact. One way to obtain a compact operator is by convolution, as shown in the following result.

Theorem 7.1. *Let $\{\rho_\epsilon\}_{\epsilon>0}$ be a family of mollifiers, let $\Omega \subset \mathbb{R}^n$ be open, and let $\varphi : \mathbb{R}^{0,+} \to \mathbb{R}^{0,+}$ be increasing, subadditive, and continuous with $\varphi(0) = 0$. Fix $\epsilon > 0$, and define $D_\epsilon : L^1(\Omega) \to L^1(\mathbb{R}^n; \mathbb{R}^n)$ by*

$$D_\epsilon u := \rho_\epsilon * Du.$$

Then

$$\mathrm{TV}^{\varphi,\epsilon}_{\mathrm{c}}(u) := \int_{\mathbb{R}^n} \varphi(|D_\epsilon u(x)|)\, dx \quad (u \in \mathrm{BV}(\Omega))$$

is lower semicontinuous with respect to weak convergence in* $\mathrm{BV}(\Omega)$. *Moreover,*

$$\lim_{\epsilon \searrow 0} \mathrm{TV}^{\varphi,\epsilon}_{\mathrm{c}}(u) = \widetilde{\mathrm{TV}}^{\varphi}_{\mathrm{c}}(u) \quad (u \in C^1(\Omega)). \tag{7.1}$$

We relegate the proof of this theorem to Appendix B. It should be noted that any $u \in L^1(\Omega)$ satisfies $\mathrm{TV}^{\varphi,\epsilon}_{\mathrm{c}}(u) < \infty$. In particular,

$$\sup_i \frac{1}{2}\|z - u^i\|^2_{L^2(\Omega)} + \alpha \mathrm{TV}^{\varphi,\epsilon}_{\mathrm{c}}(u^i) < \infty$$

does not guarantee weak* convergence of a subsequence. For that, an additional $\mathrm{TV}(u^i)$ term (with small factor) is required in a $\mathrm{TV}^{\varphi,\epsilon}_{\mathrm{c}}$-based variational model in image processing.

Remark 7.2 (the space $\mathrm{SBV}(\Omega)$ and η). If we apply the η functional of [44, 45] to a bounded sequence of functions $g^i \in L^1(\Omega; \mathbb{R}^m)$, then we get strict convergence in this space. It remains to find whether we get convergence. Then we could regularize ∇u this way, and, working in the space $\mathrm{SBV}(\Omega)$, penalize the jump part separately. It turns out that this is possible if we state the modification $\bar{\eta}$ of η in $L^p(\mathbb{R}^n; \mathbb{R}^m)$ for $p \in (1, \infty)$. Then strict convergence is equivalent to strong convergence.

With $\epsilon_\ell \searrow 0$, $\eta_0 > 0$, and $p \in (1, \infty)$, we define

$$\bar{\eta}(g) := \eta_0 \sum_{\ell=1}^{\infty} \bar{\eta}_\ell(g), \quad \bar{\eta}_\ell(g) := \|g\|_{L^p(\mathbb{R}^n;\mathbb{R}^m)} - \|g * \rho_{\epsilon_\ell}\|_{L^p(\mathbb{R}^n;\mathbb{R}^m)} \quad (g \in L^p(\Omega; \mathbb{R}^m)). \tag{7.2}$$

Then we have the following result, whose proof is relegated to Appendix C.

Theorem 7.2. *Let $\Omega \subset \mathbb{R}^m$ be open and bounded. Suppose $\psi \in \mathcal{W}_\mathrm{d}$, $\varphi : \mathbb{R}^{0,+} \to \mathbb{R}^{0,+}$ is lower semicontinuous and increasing with $\varphi^\infty = \infty$, and*

$$\|g\|_{L^p(\Omega;\mathbb{R}^m)} \le C\left(1 + \int \varphi(|g(x)|)\,dx\right) \quad (g \in L^p(\Omega;\mathbb{R}^m)) \tag{7.3}$$

for some $C > 0$, where $p \in (1,\infty)$ is as in (7.2). Let

$$F(u) := \int_{\mathbb{R}^n} \varphi(|\nabla u(x)|)\,dx + \bar\eta(\nabla u) + \int_{J_u} \psi(\theta_u(x))\,d\mathcal{H}^{m-1}(x).$$

Then F is lower semicontinuous with respect to weak convergence in $\mathrm{BV}(\Omega)$. In fact, any sequence $\{u^i\}_{i=1}^\infty$ with $\sup_i F(u^i) < \infty$ admits a subsequence convergent weakly* and in the sense of (3.1)–(3.3).*

8. Image statistics and the jump part.

Our studies in the preceding sections have pointed us to the following questions: Are the statistics of [29] valid when we split the image into smooth and jump parts? What are the statistics for jump heights, and does splitting the gradient into these two parts alter the distribution for the absolutely continuous part? When calculating statistics from discrete images, we do not have the excuse that the jumps would be negligible, i.e., $\mathcal{L}^m(J_u) = 0$!

In order to gain some insight, here we did a few experiments with real photographs, displayed in Figures 1–3. These three photographs represent images with different types of statistics. The pier photo of Figure 1 is very simple, with large smooth areas and some fine structures. The parrot test image in Figure 2 has a good balance of features. The summer lake scene in Figure 3 is somewhat more complex, with plenty of fine features.

We split the pixels of each image into edge and smooth parts by a simple threshold on the norm $\|\nabla u(k)\|$ of the discrete gradient at each pixel k. Then we find optimal α and $q \in (0,2)$ for the distribution

$$P_t(t) := C_t \exp(-\alpha\varphi(t)),$$

to match the experimental distribution. This in turn gives rise to the prior

$$P_u(u) = C_u \exp\left(-\alpha \int_\Omega \varphi(|\nabla u(x)|)\,dx\right).$$

Both $C_t, C_u > 0$ are normalizing factors. In practice we do the fitting of P_t to the experimental distribution by a simple least squares fit on the logarithms of the distributions. We will comment on the suitability of this approach later in this section. In the least squares fit we keep C as a free (unnormalized) parameter and recalculate it after the fit. Observe that the normalization constant does not affect the denoising problem (here in the finite-dimensional setting)

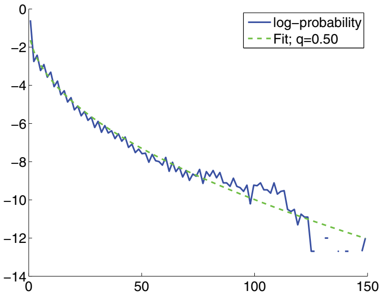
$$\max_f P_{u|f}(u|f)P_u(u) \propto \max_u \exp\left(-\frac{\sigma^2}{2}\|z-u\|_2^2 - \alpha\widetilde{\mathrm{TV}}{}_\mathrm{c}^\varphi(u)\right).$$

Here we have the Gaussian noise distribution

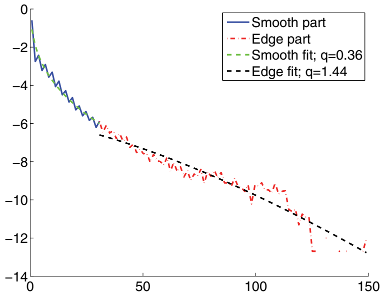
$$P_{f|u}(f|u) = C' \exp\left(-\frac{\sigma^2}{2}\|z-u\|_2^2\right)$$

(a) Pier photo

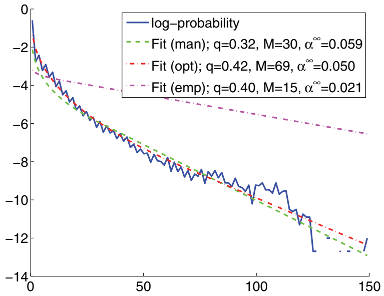
(b) Detected edge pixels (red).



(c) Log-histogram and t^q model.



(d) Separate t^q model for edge part.



(e) Linearized models.

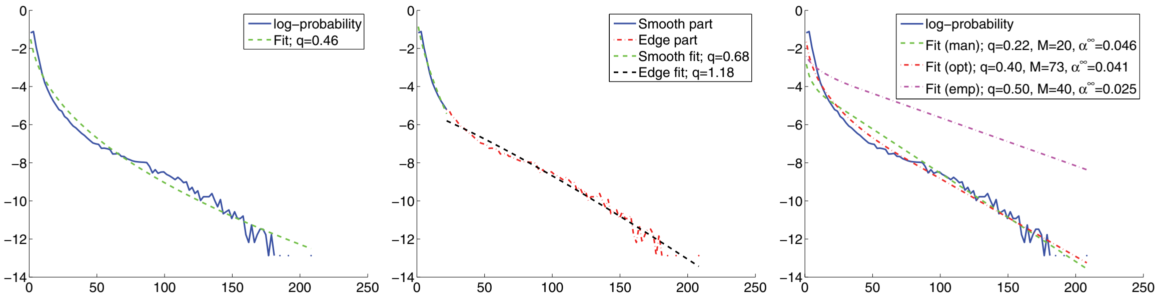
Figure 1. *Pier photo: Discrete gradient histogram and least squares models. The image intensity in* (a) *is in the range* $[0, 255]$*, and we have chosen pixels* k *with* $|\nabla u(k)| \geq 30$ *as edges* (b)*. The log-histogram of* $|\nabla u(k)|$ *with optimal fit of* $t \mapsto t^q$ *is displayed in* (c)*. This is done separately for the edge pixels in* (d)*. The linearized model is fitted in* (e) *for the cut-off point* $M = 30$ *(manual edge detection),* $M = 69$ *(optimal least squares fit). Moreover, we show the empirically best linearized model.*

for σ the noise level. This gives the statistical interpretation of the denoising model, that of a maximum a posteriori (MAP) estimate.

Finally, in the matter of statistics, we note that the TV^φ prior attempts to correctly model only gradient statistics; the modelling of histogram statistics with Wasserstein distances was recently studied in [38, 42] together with the conventional total variation gradient prior. It is also worth remarking that our approach of improving the prior based on the statistics of the ground-truth is different from recent approaches that optimize the prior based on the denoising result [8, 11, 17, 25, 26]. These approaches can provide improved results in practice, but no longer have the simple MAP interpretation. It is definitely possible to optimize the parameters of the TV^φ model in this manner, but this is beyond the scope of the present, already lengthy, paper.

Our experiments confirm the findings of [29] that some $q \in (0, 1)$ is generally a good fit for the entire distribution, as well as for the smooth part. However, optimal q for the edge part varies. In Figure 1, actually $q = 1.44$—larger than one! We have to admit that the number of

(a) Parrot photo **(b)** Detected edge pixels (red).



(c) Log-histogram and t^q model. **(d)** Separate t^q model for edge part. **(e)** Linearized models.

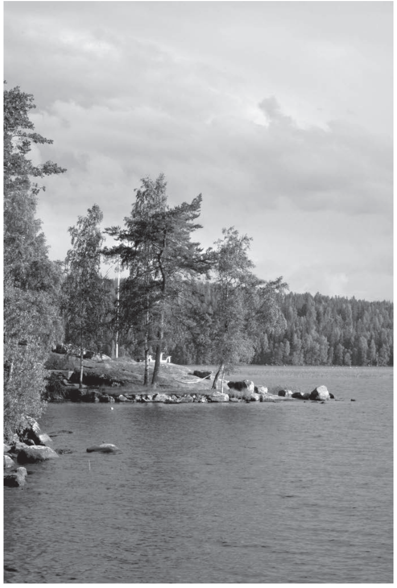
Figure 2. *Parrot photo: Discrete gradient histogram and least squares models. The image intensity in (a) is in the range $[0, 255]$, and we have chosen pixels k with $|\nabla u(k)| \geq 20$ as edges (b). The log-histogram of $|\nabla u(k)|$ with optimal fit of $t \mapsto t^q$ is displayed in (c). This is done separately for the edge pixels in (d). The linearized model is fitted in (e) for the cut-off point $M = 20$ (manual edge detection), $M = 73$ (optimal least squares fit). Moreover, we show the empirically best linearized model.*

edge pixels in this image is quite small, so statistically the result may be considered unreliable. In Figure 3, with a significant proportion of edge pixels, we still have $q = 1.05$. These findings also suggest that *on average* fitting a single $q \in (0, 1)$ to the entire statistic (not split into edge and smooth parts) may be right, but there is significant variation between images in the shape of the distribution for the edge part. The smooth parts generally look roughly similar among our test images.

In order to suggest an improved model for image gradient statistics, in each of Figures 1(e), 2(e), and 3(e), we also fit to the statistics of the linearized distribution $P_t(t) = C \exp(-\alpha\varphi(t))$ for

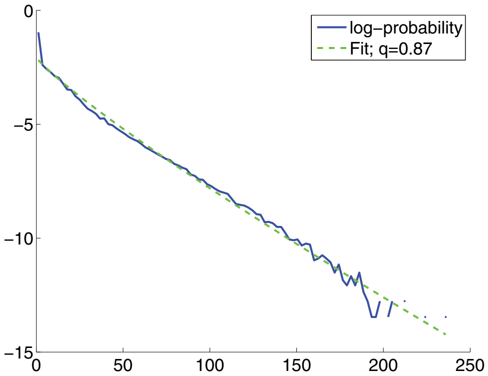
$$\varphi(t) := \begin{cases} t^q, & 0 \leq t \leq M, \\ (1-q)M^q + qM^{q-1}t, & t > M. \end{cases} \tag{8.1}$$

This is again done by a least squares fit on the logarithm of the distribution. For the "Fit (man)" curve, we fix the cut-off point M to a hand-picked (manual) edge threshold and

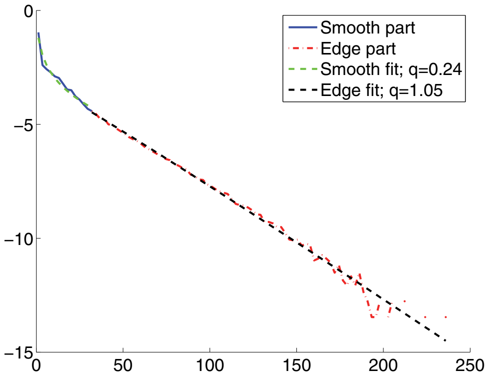


(a) Summer photo

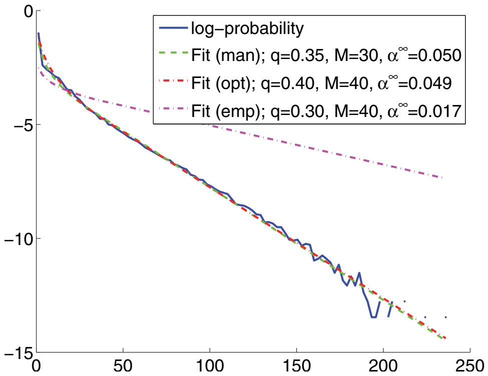
(b) Detected edge pixels (red).



(c) Log-histogram and t^q model.



(d) Separate t^q model for edge part.



(e) Linearized models.

Figure 3. *Summer photo: Discrete gradient histogram and least squares models. The image intensity in* (a) *is in the range* $[0, 255]$*, and we have chosen pixels* k *with* $|\nabla u(k)| \geq 30$ *as edges* (b)*. The log-histogram of* $|\nabla u(k)|$ *with optimal fit of* $t \mapsto t^q$ *is displayed in* (c)*. This is done separately for the edge pixels in* (d)*. The linearized model is fitted in* (e) *for the cut-off point* $M = 30$ *(manual edge detection),* $M = 40$ *(optimal least squares fit). Moreover, we show the empirically best linearized model.*

optimize (q, α). We also optimize over all of the parameters (q, α, M). This is the "Fit (opt)" curve. We note that the *asymptotic* α, which we define as

$$\alpha^\infty := \lim_{t\to\infty} \alpha\varphi(t)/t = qM^{q-1},$$

is roughly the same for both of the choices, and generally the curves are close to each other. As α^∞ describes the behavior of the model on edges, and for total variation denoising $\alpha^\infty = \alpha$, we find α^∞ to be a parameter that should indeed stay roughly constant between models with different q and M. It turns out, however, that α^∞ as obtained by the simple least squares histogram fit is in practice bad; it gives a far too high regularization, i.e., a too narrow distribution. The problem is that the simple least squares fit on the logarithm overemphasizes the tail of the distribution, on which we have, moreover, very few statistics due to the discrete nature of the data. This yields a far too high α^∞; i.e., the slope of the linear part is too steep in the figures. Developing a reliable way to obtain the model from the data is beyond the scope of the present paper, although it is definitely an interesting subject for future studies. This is why we have also included the curve "Fit (emp)," which is based on an empirical choice of (α^∞, M, q) from our numerical experiments in section 9. There we keep α^∞ fixed as we vary M and q. We will also incorporate the noise level σ^2 into α. It turns out that for the empirically good distribution, q is close to the values found by histogram fitting above, but α^∞ is very different.

9. Numerical reconstructions.

Next we provide a numerical solver for the following $\mathrm{TV}_\mathrm{c}^\varphi$ model, possibly including the η terms. We note that our solver is only one option, and not the focus of the present paper, which is on modelling and analysis. We therefore do not provide an extensive analysis and comparison to alternative approaches of the solver itself. Compared to first-order splitting approaches, recently analyzed extensively using the Kurdyka–Łojasiewicz property [5, 35], it can be said, however, that our solver can be proved to have theoretically faster local superlinear convergence [27, 28].

In the finite-dimensional setting, we aim to solve

$$\min_u f(u) := \sum_{k\in\Omega_h} \Bigg(\alpha\varphi(|\nabla u(k)|) + \frac{1}{2}|z(k) - u(k)|^2 + \eta_0 \sum_{l=1}^{N} \left(\sqrt{1+|\nabla u(k)|^2} - \sqrt{1+|\nabla_{\epsilon_l} u(k)|^2} \right) \Bigg). \tag{9.1}$$

Here φ is given by (8.1), and α, η_0 are manually chosen to balance the weights of the respective terms. For an image u of resolution n_1-by-n_2, we set $h := \sqrt{1/n_1 n_2}$, $\Omega_d := [0,1]^2 \cap (h\mathbb{Z}^2)$ and discretize the gradient by forward differences, i.e.,

$$\nabla u(k) := ((u(k+e_1) - u(k))/h,\ (u(k+e_2) - u(k))/h) \quad \text{for all} \quad k \in \Omega_d,$$

with homogeneous Dirichlet boundary condition. Each $\nabla_{\epsilon_l} u := \rho_{\epsilon_l} * \nabla u$ is defined through the convolution with a prescribed smoothing kernel ρ_{ϵ_l} ($\epsilon_l > 0$). Here ρ_{ϵ_l} is specified as a two-dimensional Gaussian distribution of standard deviation ϵ_l centered at the origin.

To cope with the kink of the nonsmooth φ term at zero, we introduce a Huber-type local smoothing [27, 28] by replacing φ in (9.1) with a continuously differentiable function φ_γ with locally Lipschitz derivative. More specifically, let $0 < \gamma \ll M$ be the smoothing parameter and $\varphi_\gamma : [0,\infty) \to [0,\infty)$ be defined by

$$\varphi_\gamma(t) = \begin{cases} \varphi(t) - (\varphi(\gamma) - \frac{1}{2}\varphi'(\gamma)\gamma) & \text{if } t \geq \gamma, \\ \frac{\varphi'(\gamma)}{\gamma^2} t^2 & \text{if } 0 \leq t < \gamma. \end{cases}$$

Thus, the resulting Huberized $\mathrm{TV}^{\varphi_\gamma}$ model appears as

$$\begin{aligned} \text{(9.2)} \quad \min_u f_\gamma(u) := \sum_{k\in\Omega_d} \Bigg(& \alpha\varphi_\gamma(|\nabla u(k)|) + \frac{1}{2}|z(k) - u(k)|^2 \\ & + \eta_0 \sum_{l=1}^N \left(\sqrt{1 + |\nabla u(k)|^2} - \sqrt{1 + |\nabla_{\epsilon_l} u(k)|^2} \right) \Bigg). \end{aligned}$$

For this problem, the first-order optimality condition reads as

$$\text{(9.3)} \quad \begin{cases} \alpha\nabla^\top p + u - z + \eta_0 \displaystyle\sum_{l=1}^N \left(\nabla^\top \left(\frac{\nabla u}{\sqrt{1+|\nabla u|^2}} \right) - \nabla_{\epsilon_l}^\top \left(\frac{\nabla_{\epsilon_l} u}{\sqrt{1+|\nabla_{\epsilon_l} u|^2}} \right) \right) = 0, \\ \psi(\max(|\nabla u|, \gamma))p = \nabla u, \end{cases}$$

where $p \in \left(\mathbb{R}^{|\Omega_d|}\right)^2$ is an auxiliary variable and $\psi : (0,\infty) \to (0,\infty)$ is defined by $\psi(t) := t/\varphi'(t)$. Note that ψ is locally Lipschitz and monotonically increasing, and in the following we shall denote by $\partial\psi$ a subdifferential of ψ. We remark that the consistency of the Huberized stationary points induced by (9.3) toward the stationary points of the original model (9.1) was investigated in the previous works [27, 28]. Moreover, the system (9.3) is not differentiable in the classical sense. Therefore, in the following we present a generalized Newton-type solver for computing a stationary point satisfying (9.3).

Given the current iterate (u^i, p^i), our solver relies on the following regularized linear system arising from differentiating (9.3) (and further straightforward manipulations; see [27, 28]):

$$(H^i + \beta^i R^i)\delta u^i = -g^i,$$

with

$$\begin{aligned} m^i &:= \max(|\nabla u|, \gamma), \\ \chi^i(k) &:= \begin{cases} 1 & \text{if } |\nabla u(k)| \geq \gamma, \\ 0 & \text{if } |\nabla u(k)| < \gamma, \end{cases} \\ H^i &:= I + \alpha\nabla^\top \frac{1}{\psi(m^i)} \left(I - \frac{\chi^i \partial\psi(m^i)}{2m^i} \left((p^i)(\nabla u^i)^\top + (\nabla u^i)(p^i)^\top \right) \right) \nabla \\ &\quad + \eta_0 \sum_{l=1}^N \left(\nabla^\top \frac{1}{\sqrt{1+|\nabla u^i|^2}} \left(I - \frac{(\nabla u^i)(\nabla u^i)^\top}{1+|\nabla u^i|^2} \right) \nabla \right. \end{aligned}$$

$$-\nabla_{\epsilon_l}^\top \frac{1}{\sqrt{1+|\nabla u^i|^2}}\left(I-\frac{(\nabla u^i)(\nabla u^i)^\top}{1+|\nabla u^i|^2}\right)\nabla_{\epsilon_l}\Bigg),$$

$$R^i := \varepsilon I + \alpha\nabla^\top \frac{\chi^i \partial\psi(m^i)}{2\psi(m^i)m^i}\left((p^i)(\nabla u^i)^\top + (\nabla u^i)(p^i)^\top\right)\nabla + \eta_0 \sum_{l=1}^N \nabla_{\epsilon_l}^\top \frac{1}{\sqrt{1+|\nabla u^i|^2}}\left(I-\frac{(\nabla u^i)(\nabla u^i)^\top}{1+|\nabla u^i|^2}\right)\nabla_{\epsilon_l},$$

$$g^i := \nabla f_\gamma(u^i) = \alpha\nabla^\top\left(\frac{\nabla u^i}{\psi(m^i)}\right) + u^i - z + \eta_0\sum_{l=1}^N\left(\nabla^\top\left(\frac{\nabla u^i}{\sqrt{1+|\nabla u^i|^2}}\right) - \nabla_{\epsilon_l}^\top\left(\frac{\nabla_{\epsilon_l} u^i}{\sqrt{1+|\nabla_{\epsilon_l} u^i|^2}}\right)\right). \tag{9.4}$$

Here H^i represents a (modified) generalized Hessian matrix of f_γ at u^i, while R^i, with an arbitrarily fixed $0 < \varepsilon \ll \alpha$, serves as a structural Hessian regularization. Note that H^i may not be positive definite at the iterate u^i. For this reason, the regularization weight β^i is automatically tuned by a trust-region–based mechanism; see steps 8–20 in Algorithm 1. Further, whenever $\beta^i = 1$, the regularized Hessian, i.e., $H^i + \beta^i R^i$, is positive definite. Consequently, our β^i-update scheme guarantees δu^i to be a descent direction for f_γ at u^i, and thus the overall iterative scheme can be globalized by, e.g., the Wolfe–Powell line search [19]; see step 21 in Algorithm 1. Moreover, following the algorithm development in [27, 28], one can show that β^i asymptotically vanishes as u^i approaches a stationary point where a certain type of second-order sufficient optimality condition is satisfied. Thus, local superlinear convergence can be attained. The overall algorithm is detailed in Algorithm 1 below. The following parameters associated with Algorithm 1 are specified throughout our experiments: $c = 1$, $\underline{\rho} = 0.25$, $\bar\rho = 0.75$, $\underline{\kappa} = 0.5$, $\bar\kappa = 2$, $\epsilon_d = 10^{-10}$, $\tau_1 = 0.1$, $\tau_2 = 0.9$, $\gamma = 0.001$, $\epsilon_r = 0.01$. Algorithm 1 is terminated once $\|\nabla f_\gamma(u^i)\|/\|\nabla f_\gamma(u^0)\|$ drops below 10^{-7}.

We report in Figures 4–6 and Table 1 the results of denoising our three test images using this algorithm with rather high artificial noise levels. We have added Gaussian noise of standard deviation $\sigma = 30$ to all test images. We report both the conventional peak-signal-to-noise ratio (PSNR) as well as the structural similarity measure (SSIM) of [48]. The latter better quantifies the visual quality of images by essentially computing the PSNR in local windows and combining the results in a nonlinear fashion. The range of the SSIM is $[0, 1]$, the higher the better.

In our computations, we keep α^∞ and q fixed and vary M (by altering α as necessary). For $M = \infty$, i.e., the original TV^q model, we simply take α as our chosen fixed α^∞. This is because $\varphi^\infty = 0$, so the real asymptotic α for the model is always zero. It is quite remarkable that in our results fine features of the images are always retained very well although higher M tends to increase the staircasing effect (not applicable to $M = \infty$). At the optimal choice of M by PSNR or SSIM, more noise can be seen to be removed than by TV ($M = 0$). Generally, we can say that adding the cut-off M improves the results compared to the earlier TV^q model without cut-off ($M = \infty$). Whether the results are better than conventional total variation denoising is open to debate. By PSNR and SSIM the results tend to favor the TV$^\varphi$ model.

Visually oscillatory effects of noise are better removed, but at the same time the staircasing is accentuated. The best result is in the eye of the beholder.

Algorithm 1. Superlinearly convergent Newton-type method for $\mathrm{TV}^{\varphi_\gamma}$ denoising.

Require: $c > 0$, $0 < \underline{\rho} \le \bar{\rho} < 1$, $0 < \underline{\kappa} < 1 < \bar{\kappa}$, $\epsilon_d > 0$, $0 < \tau_1 < 1/2$, $\tau_1 < \tau_2 < 1$, $0 < \gamma \ll 1$, $0 < \epsilon_r < 1$.

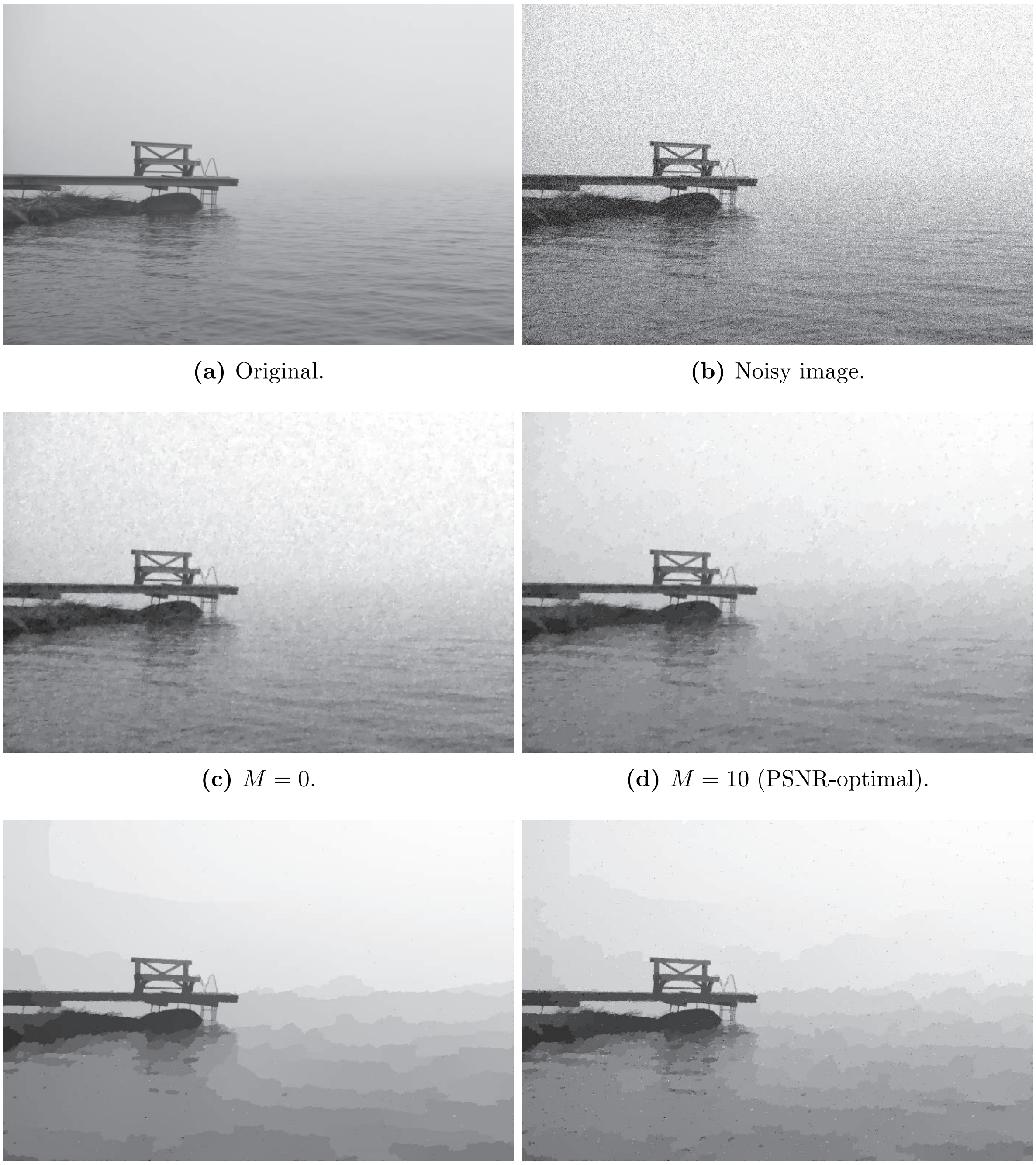
1: Initialize the iterate (u^0, p^0), the regularization weight $\beta^0 \ge 0$, and the trust-region radius $r^0 > 0$. Set $i := 0$.
2: **repeat**
3: Generate H^i, R^i, and g^i at the current iterate (u^i, p^i).
4: Solve the linear system $(H^i + \beta^i R^i)\delta u^i = -g^i$ (inexactly) for δu^i by the conjugate gradient (CG) method up to the residual tolerance ϵ_r, or detect the nonpositive definiteness of $H^i + \beta^i R^i$ during the CG iterations.
5: **if** $H^i + \beta^i R^i$ is not positive definite **or** $-((g^i)^\top \delta u^i)/(\|g^i\|\|\delta u^i\|) < \epsilon_d$ **then**
6: Set $\beta^i := 1$, and return to step 4.
7: **end if**
8: **if** $\beta^i = 1$ **and** $(\delta u^i)^\top R^i \delta u^i > (r^i)^2$ **then**
9: Set $r^i := \sqrt{(\delta u^i)^\top R^i \delta u^i}$, $\beta^{i+1} := 1$, and go to step 13.
10: **else**
11: Set $\beta^{i+1} := \max(\min(\beta^i + c^{-1}((\delta u^i)^\top R^i \delta u^i - (r^i)^2), 1), 0)$.
12: **end if**
13: Evaluate $\rho^i := \left(f_\gamma(u^i + \delta u^i) - f_\gamma(u^i)\right) / \left((g^i)^\top \delta u^i + (\delta u^i)^\top H^i \delta u^i / 2\right)$.
14: **if** $\rho^i < \underline{\rho}$ **then**
15: Set $r^{i+1} := \underline{\kappa} r^i$.
16: **else if** $\rho^i > \bar{\rho}$ **then**
17: Set $r^{i+1} := \bar{\kappa} r^i$.
18: **else**
19: Set $r^{i+1} := r^i$.
20: **end if**
21: Determine the step size a^i along the search direction δu^i such that $u^{i+1} = u^i + a^i \delta u^i$ satisfies the following Wolfe–Powell conditions:

$$\begin{cases} f_\gamma(u^{i+1}) \le f_\gamma(u^i) + \tau_1 a^i (g^i)^\top \delta u^i, \\ \nabla f_\gamma(u^{i+1})^\top \delta u^i \ge \tau_2 (g^i)^\top \delta u^i. \end{cases}$$

22: Generate the next iterate:

$$\begin{aligned} u^{i+1} &:= u^i + a^i \delta u^i, \\ p^{i+1} &:= \frac{1}{\psi(m^i)} \left(\nabla u^i + \nabla \delta u^i - \frac{\chi^i \partial \psi(m^i)}{2m^i} \left((p^i)(\nabla u^i)^\top + (\nabla u^i)(p^i)^\top \right) \nabla \delta u^i \right). \end{aligned}$$

23: Set $i := i + 1$.
24: **until** the stopping criterion is fulfilled.

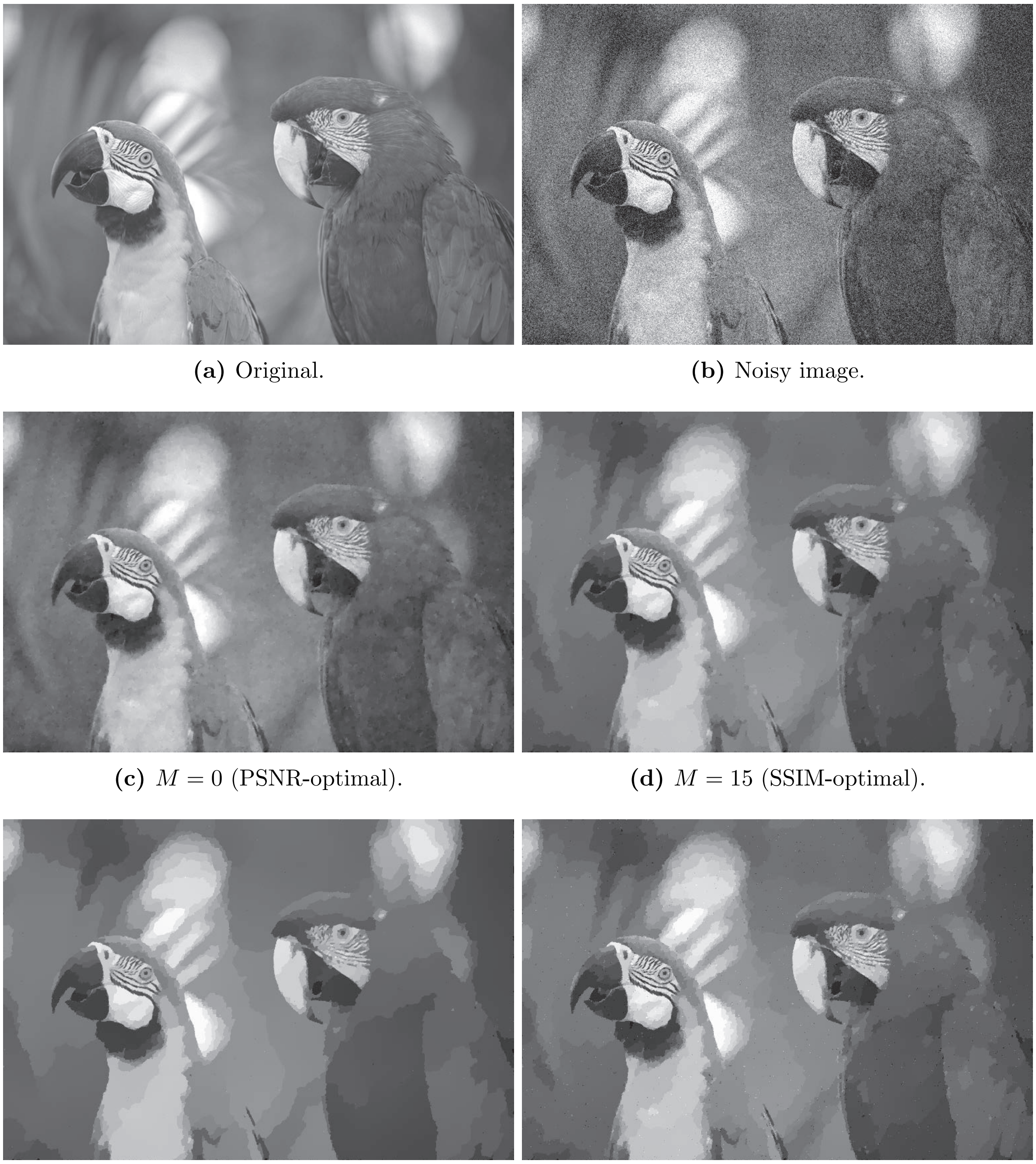


(a) Original. **(b)** Noisy image.

(c) $M = 0$. **(d)** $M = 10$ (PSNR-optimal).

(e) $M = 40$ (SSIM-optimal). **(f)** $M = \infty$.

Figure 4. *Pier photo: Denoising results with noise level* $\sigma = 30$ *(Gaussian) for varying cut-off* M*, fixed* $q = 0.4$*, and fixed* $\alpha^\infty = 0.0207$.

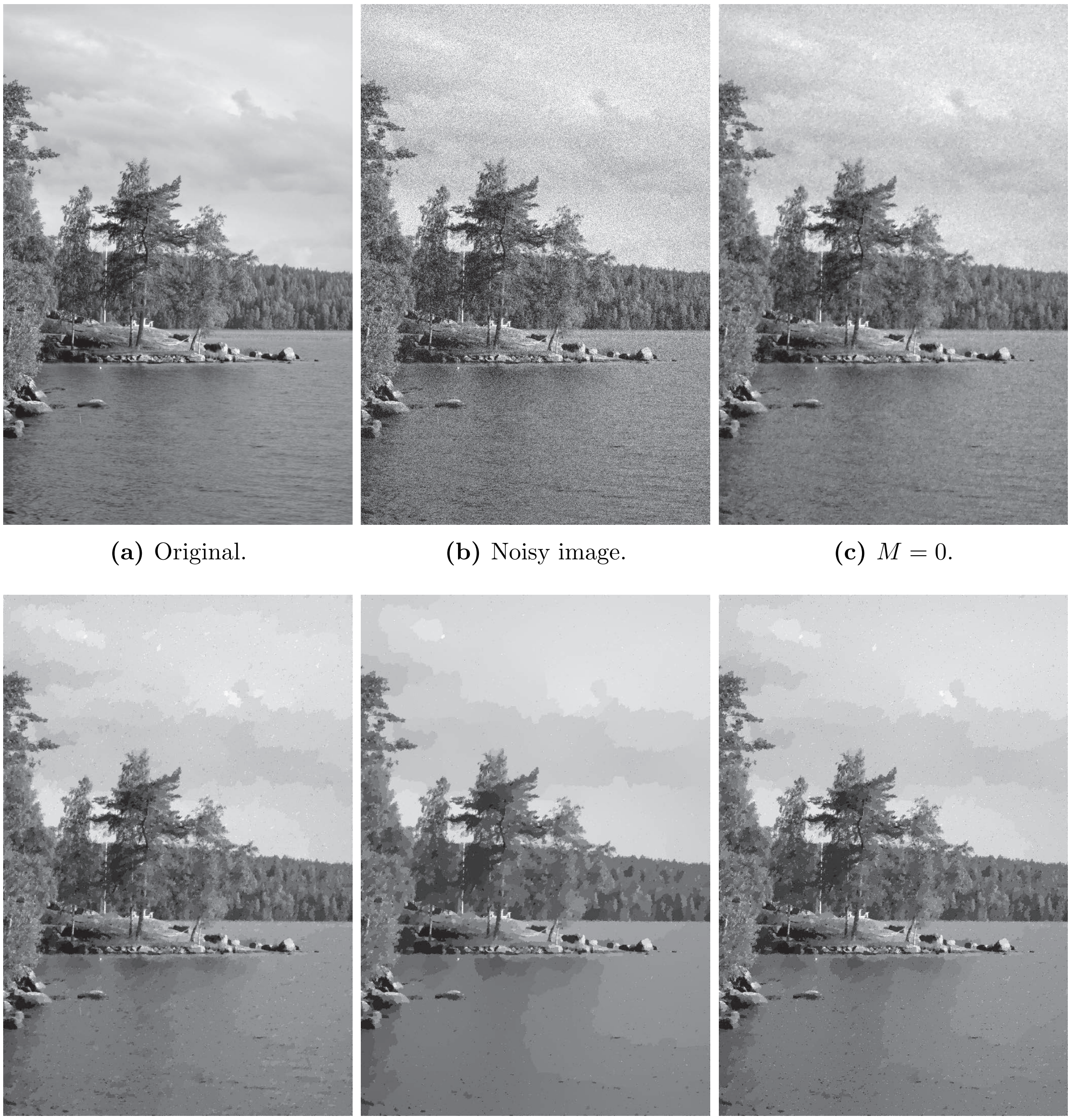


(a) Original. **(b)** Noisy image.

(c) $M = 0$ (PSNR-optimal). **(d)** $M = 15$ (SSIM-optimal).

(e) $M = 40$. **(f)** $M = \infty$.

Figure 5. *Parrot photo: Denoising results with noise level* $\sigma = 30$ *(Gaussian) for varying cut-off* M*, fixed* $q = 0.5$*, and fixed* $\alpha^\infty = 0.0253$.



(a) Original. **(b)** Noisy image. **(c)** $M = 0$.

(d) $M = 20$ (PSNR-optimal). **(e)** $M = 40$ (SSIM-optimal). **(f)** $M = \infty$.

Figure 6. *Summer photo: Denoising results with noise level* $\sigma = 60$ *(Gaussian) for varying cut-off* M*, fixed* $q = 0.3$*, and fixed* $\alpha^\infty = 0.00430$.

Table 1

Denoising results of all three test photos. The noise level σ, exponent q, and asymptotic regularization α^∞ are fixed. The cut-off point M varies. We report both the PSNR and the SSIM, with the optimal values underlined.

Parrot photo: $\sigma = 30$, $q = 0.5$, $\alpha^\infty = 0.0253$								
M	0	10	15	30	40	50	60	∞
PSNR	<u>30.3432</u>	29.9156	29.5876	28.8098	28.3655	28.0398	27.7424	28.6829
SSIM	0.7914	0.7919	<u>0.7922</u>	0.7906	0.7887	0.7854	0.7823	0.7552
Pier photo: $\sigma = 30$, $q = 0.4$, $\alpha^\infty = 0.0207$								
M	0	10	20	30	40	50	60	∞
PSNR	29.0019	<u>29.3959</u>	29.3472	29.0918	28.8988	28.679	28.5327	28.5836
SSIM	0.6737	0.7191	0.7477	0.7556	<u>0.7619</u>	0.7608	0.7593	0.7297
Summer photo: $\sigma = 30$, $q = 0.4$, $\alpha^\infty = 0.00430$								
M	0	10	20	30	40	50	60	∞
PSNR	26.0919	26.2750	<u>26.2851</u>	26.0643	25.7443	25.3627	25.0449	25.2755
SSIM	0.589	0.6175	0.6489	0.6641	<u>0.6644</u>	0.6615	0.6543	0.6175

Table 2

Effect of the η term on the parrot photo. Only the first term η_1 of the sum is included, with varying convolution width ϵ_1 and weight η_0. The noise level $\sigma = 30$ (Gaussian), cut-off $M = 10$, exponent $q = 0.5$, and asymptotic regularization $\alpha^\infty = 0.0253$ are fixed. Optimal SSIM and PSNR are underlined.

	$\epsilon_1 = 0.5$		$\epsilon_1 = 1$		$\epsilon_1 = 2$	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
$\eta_0 = 0.1\alpha$	<u>29.9433</u>	0.8036	29.8023	<u>0.8091</u>	29.6197	0.8085
$\eta_0 = \alpha$	29.2700	0.8072	28.2237	0.8014	27.5659	0.7970
$\eta_0 = 10\alpha$	26.8069	0.7939	25.7874	0.7921	24.9003	0.7886

We also tested on the parrot photo the effect of the multiscale regularization term η by including the first term η_1 of the sum for varying weights of η_0 and convolution kernel widths ϵ_1. The results are presented in Figure 7 and Table 2. Clearly large η_0 has a deteriorating effect on both PSNR and SSIM, whereas the effect of the choice of ϵ_1 is less severe. Visually, large η_0 creates an almost artistic quantization and feature-filtering effect. The latter is also controlled by ϵ_1: Large ϵ_1 tends to remove large features. A particular feature to notice is the eye of the parrot on the right in Figure 7(a) versus 7(b). It has disappeared altogether in the latter.

10. Conclusion.

We have studied difficulties with nonconvex total variation models in the function space setting. We have demonstrated that the model (1.2) continues to do what it is proposed to do in the discrete setting—to promote piecewise constant solutions—for most, but not all, energies φ employed in the literature. Naive forms of the model (1.1), proposed to model real gradient distributions in images, however, have much more severe difficulties. We have shown that the model can be remedied if we replace the topology of weak* convergence by that of area-strict convergence. In order to do this, we have to add additional multiscale regularization in terms of the functional η into the model, and have to linearize the energy φ for large gradients. The latter is needed to make the model BV-coercive and to have any kind of penalization for jumps. We have demonstrated through numerical experiments and simple statistics that this model, in fact, better matches reality than the simple energies $\varphi(t) = t^q$.

(a) $\eta_0 = 7.071\text{E}^{-03}$, $\epsilon_1 = 1.0$.

(b) $\eta_0 = 7.071\text{E}^{-03}$, $\epsilon_1 = 2.0$.

(c) $\eta_0 = 7.071\text{E}^{-04}$, $\epsilon_1 = 1.0$.

(d) $\eta_0 = 7.071\text{E}^{-04}$, $\epsilon_1 = 2.0$.

(e) $\eta_0 = 7.071\text{E}^{-05}$, $\epsilon_1 = 1.0$.

(f) $\eta_0 = 7.071\text{E}^{-05}$, $\epsilon_1 = 2.0$.

Figure 7. *Effect of the η term on the parrot photo. Only the first term η_1 of the sum is included, with varying convolution width ϵ_1 and weight η_0. The noise level $\sigma = 30$ (Gaussian), cut-off $M = 10$, exponent $q = 0.5$, and asymptotic regularization $\alpha^{\infty} = 0.0253$ are fixed.*

Our purely theoretical starting point has therefore led to improved practical models. The η functional, however, remains a "theoretical artifact." It has its own regularization effect that, naturally, does not distort the results too much for small parameters (though it does so for large parameters). As we show in Proposition 6.7, it can be ignored in discretizations when not passing to the function space limit.

A data statement of the Engineering and Physical Sciences Research Council. This is primarily a theoretical mathematics paper, and any data used mainly serves as a demonstration of mathematically proven results. Moreover, photographs that are for all intents and purposes statistically comparable to those used for the final experiments can easily be produced with a digital camera or downloaded from the Internet; in particular, the parrot test photo is available in the Kodak Lossless True Color Image Suite.[1] For the computations, we directly applied software developed in an earlier research program. This was funded by various non-UK agencies, whose rules govern the code.

Appendix A. Vectorial η functional. We now study a condition ensuring the convergence of the total variation $|\mu^i|(\Omega)$ subject to the weak* convergence of the measures $\mu^i \in \mathcal{M}(\Omega;\mathbb{R}^m)$ $(i = 0, 1, 2, \ldots)$. Improving a result first presented in [44, 45], we show in Theorem A.4 that if $\{f_\ell\}_{\ell=0}^\infty$ is a normalized nested sequence of functions as in Definition A.1, then it suffices to bound

$$\text{(A.1)}\quad \eta(\mu^i) := \sum_{\ell=0}^{\infty} \eta_\ell(\mu^i), \quad \text{where} \quad \eta_\ell(\mu^i) := \int |\mu^i|(\tau_x f_\ell) - |\mu^i(\tau_x f_\ell)|\, dx \quad (\mu \in \mathcal{M}(\Omega;\mathbb{R}^N)).$$

Here we employ the notation $\tau_x f(y) := f(y - x)$. Also, we write $|\mu^i(\tau_x f_\ell)| := \|\mu^i(\tau_x f_\ell)\|_2$.

Definition A.1. *Let $f_\ell : \mathbb{R}^m \to \mathbb{R}$ $(\ell = 0, 1, 2, \ldots)$ be bounded Borel functions with compact support that are continuous in $\mathbb{R}^m \setminus S_{f_\ell}$; i.e. the approximate discontinuity set equals the discontinuity set. Also let $\{\nu_\ell\}_{\ell=0}^\infty$ be a sequence in $\mathcal{M}(\mathbb{R}^m)$ with $|\nu_\ell|(\mathbb{R}^m) = 1$. The sequence $\{(f_\ell, \nu_\ell)\}_{\ell=0}^\infty$ is then said to form a* nested sequence of functions *if $f_\ell(x) = \int f_{\ell+1}(x-y)\, d\nu_\ell(y)$ (a.e.). The sequence is said to be* normalized *if $f_\ell \geq 0$ and $\int f_\ell\, dx = 1$.*

Example A.1. Let ρ be the standard convolution mollifier such that

$$\rho(x) := \begin{cases} \exp(-1/(1 - \|x\|^2)), & \|x\| < 1, \\ 0, & \|x\| \geq 1, \end{cases}$$

and define $\rho_\epsilon(x) := \epsilon^{-m}\rho(x/\epsilon)$. Since $\rho_{\epsilon+\delta} = \rho_\epsilon * \rho_\delta$, where $*$ denotes the convolution operation, we deduce that $f_\ell := \rho_{2^{-\ell}}$ and $\nu_\ell = \rho_{2^{-\ell-1}}$ form a normalized nested sequence.

We require the following basic lemma for our vectorial case.

Lemma A.2. *Let $\nu \in \mathcal{M}(\Omega)$ be a positive Radon measure, and let $g \in L^1(\Omega;\mathbb{R}^N)$. Then*

$$\left\| \int_\Omega g(x)\, d\nu(x) \right\|_2 \leq \int_\Omega \|g(x)\|_2\, d\nu(x).$$

[1] At the time of this writing, the image suite was available at http://r0k.us/graphics/kodak/.

Proof. For any $x \in \Omega$, we write $g(x) = \theta(x)v(x)$ with $v(x) = (v_1(x), \ldots, v_N(x))$, $0 \le \theta(x)$, and $\|v(x)\|_2 = 1$. Then we define $\lambda := \theta\nu$. Now

$$\left\| \int_\Omega g(x)\, d\nu(x) \right\|_2^2 = \sum_{j=1}^N \left(\int_\Omega v_j(x)\, d\lambda(x) \right)^2 = \sum_{j=1}^N \lambda(\Omega)^2 \left(\frac{1}{\lambda(\Omega)} \int_\Omega v_j(x)\, d\lambda(x) \right)^2$$
$$\le \sum_{j=1}^N \lambda(\Omega) \int_\Omega |v_j(x)|^2\, d\lambda(x) = \lambda(\Omega)^2.$$

Here we have used Jensen's inequality. From this we conclude that

$$\lambda(\Omega) = \int_\Omega \theta(x)\, d\nu(x) = \int_\Omega \|g(x)\|_2\, d\nu(x),$$

proving the claim. ■

With the help of the above lemma, in [45] Theorem A.4 below was proved exactly as in the case of scalar-valued measures ($N = 1$). Our proof here, however, is slightly different. We base it on the following more general lemma on partial sums, which we also need for the proof of Proposition 6.7.

Lemma A.3. *Let $\Omega \subset \mathbb{R}^m$ be an open and bounded set, and let $\{(f_\ell, \nu_\ell)\}_{\ell=0}^\infty$ be a normalized nested sequence of functions. Let $\{K_i\}_{i=1}^\infty \subset \mathbb{N}^+$ satisfy $\lim_{i\to\infty} K_i = \infty$. Suppose $\{\mu^i\}_{i=0}^\infty \subset \mathcal{M}(\Omega; \mathbb{R}^N)$ weakly* converges to $\mu \in \mathcal{M}(\Omega; \mathbb{R}^N)$ with*

$$\sup_i |\mu^i|(\Omega) + \sum_{\ell=1}^{K_i} \eta_\ell(\mu^i) < \infty \tag{A.2}$$

and

$$\eta(\mu) < \infty. \tag{A.3}$$

Then

$$\eta_\ell(\mu) \le \liminf_{i\to\infty} \eta_\ell(\mu^i) \quad (\ell = 0, 1, 2, \ldots). \tag{A.4}$$

If also $|\mu^i| \overset{*}{\rightharpoonup} \lambda$ in $\mathcal{M}(\Omega)$, then $\lambda = |\mu|$. Moreover, provided that the weak convergences hold in $\mathcal{M}(\mathbb{R}^m; \mathbb{R}^N)$ (resp., $\mathcal{M}(\mathbb{R}^m)$), then*

$$\eta_\ell(\mu) = \lim_{i\to\infty} \eta_\ell(\mu^i) \quad (\ell = 0, 1, 2, \ldots). \tag{A.5}$$

Proof. Let us suppose first that $\mu^i \overset{*}{\rightharpoonup} \mu$ and $|\mu^i| \overset{*}{\rightharpoonup} \lambda$ weakly* in $\mathcal{M}(\mathbb{R}^m; \mathbb{R}^N)$ (resp., $\mathcal{M}(\mathbb{R}^m)$) rather than just within Ω. We denote by E_f the discontinuity set of f, while S_f stands for the approximate discontinuity set. Fubini's theorem and the fact that S_f is an $\mathcal{L}^m$-negligible Borel set imply that $\int \lambda(S_{\tau_x f_\ell})\, dx = 0$. This and the nonnegativity of λ show that $\lambda(S_{\tau_x f_\ell}) = 0$ for a.e. $x \in \mathbb{R}^m$. Since by assumption $E_f \subset S_f$, it follows that $\lambda(E_{\tau_x f_\ell}) = 0$, so that (see, e.g., [3, Proposition 1.62]) $\mu^i(\tau_x f_\ell) \to \mu(\tau_x f_\ell)$ for a.e. $x \in \mathbb{R}^m$.

Likewise $|\mu^i|(\tau_x f_\ell) \to \lambda(\tau_x f_\ell)$ for a.e. $x \in \mathbb{R}^m$. Since $\sup_i |\mu^i|(\Omega) < \infty$, and Ω is bounded, an application of the dominated convergence theorem now yields

$$\lim_{i\to\infty} \int |\mu^i(\tau_x f_\ell)|\,dx = \int |\mu(\tau_x f_\ell)|\,dx. \tag{A.6}$$

By the lower semicontinuity of the total variation, recalling that

$$\int |\mu^i|(\tau_x f_\ell)\,dx = |\mu^i|(\mathbb{R}^m),$$

this shows (A.4) under the assumption that the weak* convergences are in $\mathcal{M}(\mathbb{R}^m)$.

Observe then that since $\{(f_\ell, \nu_\ell)\}_{\ell=0}^\infty$ is a nested sequence of functions, $\{\eta_\ell(\mu)\}_{\ell=0}^\infty$ forms a decreasing sequence (for any $\mu \in \mathcal{M}(\Omega)$). Indeed, as $f_\ell(x) = \int f_{\ell+1}(x-y)\,d\nu_\ell(y)$ and $\nu_\ell(\mathbb{R}^m) = 1$ with $\nu_\ell \geq 0$, using Lemma A.2 we have

$$\begin{aligned}\int \|\mu(\tau_x f_\ell)\|_2\,dx &= \int \left\| \int \mu(\tau_{x+y} f_{\ell+1})\,d\nu_\ell(y) \right\|_2 dx \\ &\leq \int\int \|\mu(\tau_{x+y} f_{\ell+1})\|_2\,d\nu_\ell(y)\,dx = \int \|\mu(\tau_x f_{\ell+1})\|_2\,dx.\end{aligned}$$

Referring to (A.1), it follows that

$$\eta_\ell(\mu) \geq \eta_{\ell+1}(\mu). \tag{A.7}$$

To show $\lambda = |\mu|$, that is, $|\mu^i| \overset{*}{\rightharpoonup} |\mu|$, we need only show that $|\mu^i|(\Omega) \to |\mu|(\Omega)$. To see the latter, we choose an arbitrary $\epsilon > 0$ and write

$$|\mu|(\Omega) - |\mu^i|(\Omega) = \eta_\ell(\mu) - \eta_\ell(\mu^i) + \int |\mu(\tau_x f_\ell)| - |\mu^i(\tau_x f_\ell)|\,dx. \tag{A.8}$$

Since $\eta_\ell \geq 0$, using (A.7) in (A.2) and (A.3), we now observe that taking ℓ large enough and i_ℓ such that $K_{i_\ell} \geq \ell$, we have

$$\sup\{\eta_\ell(\mu), \eta_\ell(\mu^{i_\ell}), \eta_\ell(\mu^{i_\ell+1}), \eta_\ell(\mu^{i_\ell+2}), \ldots\} \leq \epsilon.$$

Employing this in (A.8), we deduce for any large enough ℓ and all $i \geq i_\ell$ that

$$\big||\mu|(\Omega) - |\mu^i|(\Omega)\big| \leq 2\epsilon + \left| \int |\mu(\tau_x f_\ell)| - |\mu^i(\tau_x f_\ell)|\,dx \right|.$$

The integral term tends to zero as $i \to \infty$ by (A.6). Therefore

$$\lim_{i\to\infty} \big||\mu^i|(\Omega) - |\mu|(\Omega)\big| \leq 2\epsilon.$$

Since $\epsilon > 0$ was arbitrary, we conclude that $\lambda = |\mu|$. Moreover, (A.5) now follows from (A.6). This concludes the proof of the lemma under the assumption that the weak* convergences are in $\mathcal{M}(\mathbb{R}^m)$.

If this assumption does not hold, we may still switch to a subsequence for which $\mu^{i_k} \overset{*}{\rightharpoonup} \bar\mu$ weakly* in $\mathcal{M}(\mathbb{R}^m;\mathbb{R}^N)$ for some $\bar\mu \in \mathcal{M}(\mathbb{R}^m;\mathbb{R}^N)$. But, since Ω is open, necessarily $\bar\mu \llcorner \Omega = \mu$. Moreover, an application of the triangle inequality gives

$$\eta_\ell(\mu) = \eta_\ell(\bar\mu \llcorner \Omega) \le \eta_\ell(\bar\mu) \le \liminf_{k\to\infty} \eta_\ell(\mu^{i_k}).$$

As this bound holds for every subsequence, we deduce (A.4). Likewise, we have $|\mu^{i_k}| \overset{*}{\rightharpoonup} \bar\lambda$ weakly* in $\mathcal{M}(\mathbb{R}^m)$ for a common subsequence. Again $\bar\lambda \llcorner \Omega = \lambda$. Since by the previous paragraphs $|\bar\mu| = \bar\lambda$, this implies $\lambda = |\mu|$. ∎

Theorem A.4. *Let $\Omega \subset \mathbb{R}^m$ be an open and bounded set, and let $\{(f_\ell, \nu_\ell)\}_{\ell=0}^\infty$ be a normalized nested sequence of functions. Suppose the sequence $\{\mu^i\}_{i=0}^\infty$ in $\mathcal{M}(\Omega;\mathbb{R}^N)$ converges weakly* to a measure $\mu \in \mathcal{M}(\Omega;\mathbb{R}^N)$ and satisfies $\sup_i |\mu^i|(\Omega) + \eta(\mu^i) < \infty$. If also $|\mu^i| \overset{*}{\rightharpoonup} \lambda$, then $\lambda = |\mu|$. Moreover, each of the functionals η and η_ℓ ($\ell = 0, 1, 2, \dots$) is lower semicontinuous with respect to the weak* convergence of $\{\mu^i\}_{i=0}^\infty$.*

Proof. Only lower semicontinuity of η demands a proof; the rest is immediate from Lemma A.3 with $K_i = i$, for instance. Also, lower semicontinuity of η follows simply from Fatou's lemma and (A.4). ∎

Appendix B. Proof of Theorem 7.1.

Here we prove our result on the remedy by resorting to compact operators.

Lemma B.1. *Let Ω, Ω' be open domains, and suppose $K : \mathrm{BV}(\Omega) \to L^1(\Omega';\mathbb{R}^m)$ is a compact linear operator. Let $\varphi : \mathbb{R}^{0,+} \to \mathbb{R}^{0,+}$ be lower semicontinuous. Then*

$$F(u) := \int_{\Omega'} \varphi(\|Ku(x)\|)\,dx$$

is lower semicontinuous with respect to weak convergence in* $\mathrm{BV}(\Omega)$.

Proof. If $\{u^i\}_{i=1}^\infty \subset \mathrm{BV}(\Omega)$ converges weakly* to $u \in \mathrm{BV}(\Omega)$, then it is bounded in $\mathrm{BV}(\Omega)$. Therefore, by the compactness of K, the sequence $\{Ku^i\}_{i=1}^\infty$ has a subsequence, unrelabelled, which converges strongly to some $v \in L^1(\Omega';\mathbb{R}^m)$. By the continuity of K, which follows from compactness, necessarily $v = Ku$. Now [22, Theorem 5.9] shows that

$$F(u) \le \lim_{i\to\infty} F(u^i). \qquad ∎$$

Lemma B.2. *Let $\rho \in C_c^\infty(\mathbb{R}^n)$ and $\Omega \subset \mathbb{R}^n$ be bounded and open. Suppose $\mu^i \to \mu$ weakly* in $\mathcal{M}(\Omega;\mathbb{R}^m)$. Then $\rho * \mu^i \to \rho * \mu$ strongly in $L^\infty(\mathbb{R}^n)$.*

Proof. Let K be a compact set such that $\Omega + \operatorname{supp}\rho \subset K$. We have

$$\|\rho * \mu^i\|_{L^\infty(K;\mathbb{R}^m)} \le \|\rho\|_{L^\infty(\mathbb{R}^n)} |\mu^i|(\Omega)$$

and

$$\|\nabla(\rho * \mu^i)\|_{L^\infty(K;\mathbb{R}^n\times\mathbb{R}^m)} = \|(\nabla\rho) * \mu^i\|_{L^\infty(K;\mathbb{R}^n\times\mathbb{R}^m)} \le \|\nabla\rho\|_{L^\infty(\mathbb{R}^n;\mathbb{R}^n)} |\mu^i|(\Omega).$$

Thus $\{\rho * \mu^i\}_{i=1}^\infty$ is uniformly bounded and equicontinuous. It follows from the Arzelá–Ascoli theorem that $\rho * \mu^i$ converges uniformly (i.e., in $L^\infty(K;\mathbb{R}^m)$) to some $v \in C_c(K;\mathbb{R}^m)$.

Let $\varphi \in L^1(K;\mathbb{R}^m)$. Then by the weak* convergence we have

$$\int_{\mathbb{R}^n} \langle \varphi(x), (\rho * \mu^i)(x)\rangle\, dx = \int_{\mathbb{R}^n} \langle (\varphi * \rho)(x),\, d\mu^i(x)\rangle$$
$$\to \int_{\mathbb{R}^n} \langle (\varphi * \rho)(x),\, d\mu(x)\rangle = \int_{\mathbb{R}^n} \langle \varphi(x), (\rho * \mu)(x)\rangle\, dx$$

as $i \to \infty$, so that $\rho * \mu^i \to \rho * \mu$ weakly in $L^\infty(K;\mathbb{R}^m)$. Then it holds that $\rho * \mu = v$, and the convergence is strong. Because $\operatorname{supp} w \subset K$ for $w = \rho * \mu$ or $w = \rho * \mu^i$, the claim follows. ∎

Proof of Theorem 7.1. Suppose $\{u^i\}_{i=1}^\infty$ is a bounded sequence in $\mathrm{BV}(\Omega)$. We may extract a subsequence, unrelabelled, such that $\{u^i\}_{i=1}^\infty$ is weakly* convergent in $\mathrm{BV}(\Omega)$ to some $u \in \mathrm{BV}(\Omega)$. Then by Lemma B.2,

$$D_\epsilon u^i \to D_\epsilon u \quad \text{strongly in } L^\infty(\Omega;\mathbb{R}^m).$$

Weak* lower semicontinuity now follows from Lemma B.1.

Let then $u \in C^1(\Omega)$. The estimate

$$\limsup_{\epsilon \searrow 0} \mathrm{TV}_c^{\varphi,\epsilon}(u) \le \widetilde{\mathrm{TV}_c^{\varphi}}(u) \quad (u \in C^1(\Omega)) \tag{B.1}$$

follows from the proof of Lemma 4.5. By subadditivity we also have

$$\widetilde{\mathrm{TV}_c^{\varphi}}(u) - \mathrm{TV}_c^{\varphi,\epsilon}(u) \le \int_\Omega \varphi(\|(\rho_\epsilon * \nabla u)(x) - \nabla u(x)\|)\, dx.$$

Writing $g_\epsilon(x) := \|(\rho_\epsilon * \nabla u)(x) - \nabla u(x)\|$, we have $g_\epsilon(x) \to 0$ in $L^1(\mathbb{R}^n)$. We may again proceed as in the proof of Lemma 4.5 to show

$$\limsup_{\epsilon \searrow 0} \big(\widetilde{\mathrm{TV}_c^{\varphi}}(u) - \mathrm{TV}_c^{\varphi,\epsilon}(u)\big) \le 0.$$

This together with (B.1) proves (7.1). ∎

Appendix C. Proof of Theorem 7.2.

We now prove our result on the remedy based on the SBV space.

Proposition C.1. *Let $\bar\eta$ and $p \in (1,\infty)$ be as in (7.2). Suppose $g^i \rightharpoonup g$ weakly in $L^p(\Omega;\mathbb{R}^m)$, and that $\sup_i \bar\eta(g^i) < \infty$. Then $g^i \to g$ strongly in $L^p(\Omega;\mathbb{R}^m)$.*

Proof. Let K be a compact set such that $\Omega + \operatorname{supp}\rho_1 \subset K$, and set $M := \sup_i \bar\eta(g^i)$. We observe that $\bar\eta$ is lower semicontinuous with respect to weak convergence in L^p. Therefore $\eta(g) \le M$. As in the proof of Lemma A.3, we observe that $\bar\eta_\ell \ge \bar\eta_{\ell+1}$, from which it again follows that

$$\bar\eta_\ell(h) \to 0 \quad \text{as } \ell \to \infty \quad \text{uniformly for } h \in \{g, g^1, g^2, g^3, \ldots\}. \tag{C.1}$$

We then observe that as in Lemma B.2, we have

$$\rho_{\epsilon_\ell} * g^i \to \rho_{\epsilon_\ell} * g \quad \text{strongly in } L^\infty(K;\mathbb{R}^m) \tag{C.2}$$

for each $\ell \in \{1, 2, 3, \dots\}$. Thus, it holds that

$$\|g^i\|_{L^p(\Omega;\mathbb{R}^m)} - \|g\|_{L^p(\Omega;\mathbb{R}^m)} \le \eta_\ell(g^i) - \eta_\ell(g) + \|\rho_{\epsilon_\ell} * g^i - \rho_{\epsilon_\ell} * g\|_{L^p(K;\mathbb{R}^m)}.$$

Given $\delta > 0$, thanks to (C.1), we may find ℓ such that

$$\|g^i\|_{L^p(\Omega;\mathbb{R}^m)} - \|g\|_{L^p(\Omega;\mathbb{R}^m)} \le 2\delta + \|\rho_{\epsilon_\ell} * g^i - \rho_{\epsilon_\ell} * g\|_{L^p(K;\mathbb{R}^m)}.$$

With ℓ fixed, we thus get by (C.2) that

$$\limsup_{i\to\infty} \|g^i\|_{L^p(\Omega;\mathbb{R}^m)} - \|g\|_{L^p(\Omega;\mathbb{R}^m)} \le 2\delta.$$

Since $\delta > 0$ was arbitrary, and using weak lower semicontinuity of $\|\cdot\|_{L^p(\Omega;\mathbb{R}^m)}$, we deduce

$$\lim_{i\to\infty} \|g^i\|_{L^p(\Omega;\mathbb{R}^m)} = \|g\|_{L^p(\Omega;\mathbb{R}^m)}.$$

But for $p \in (1, \infty)$, strict convergence implies strong convergence [22]. This concludes the proof. ■

Proof of Theorem 7.2. If $\{u^i\}_{i=1}^\infty$ is weakly* convergent in $\mathrm{BV}(\Omega)$, we may—without loss of generality—assume that $\sup_i F(u^i) < \infty$, for otherwise lower semicontinuity is obvious. Then by the SBV compactness theorem, Theorem 3.5, the convergences (3.1)–(3.4) hold for a subsequence. This also shows weak* convergence if it did not hold originally. Moreover, by the same theorem, $u \mapsto \int_{J_u} \psi(\theta_u(x))\, d\mathcal{H}^{m-1}(x)$ is lower semicontinuous with respect to this convergence. By (7.3) we may further assume $\{\nabla u^i\}_{i=1}^\infty$ is weakly convergent in $L^p(\Omega;\mathbb{R}^m)$. Proposition C.1 now shows strong convergence of $\{\nabla u^i\}_{i=1}^\infty$ in $L^p(\Omega;\mathbb{R}^m)$. The functional F is lower semicontinuous with respect to all of these convergences, which yields lower semicontinuity with respect to weak* convergence in $\mathrm{BV}(\Omega)$. ■

REFERENCES

[1] W. K. Allard, *Total variation regularization for image denoising,* I. *Geometric theory*, SIAM J. Math. Anal., 39 (2007), pp. 1150–1190.

[2] L. Ambrosio, *A compactness theorem for a new class of functions of bounded variation*, Boll. Un. Mat. Ital. B (7), 3 (1989), pp. 857–881.

[3] L. Ambrosio, A. Coscia, and G. Dal Maso, *Fine properties of functions with bounded deformation*, Arch. Rational Mech. Anal., 139 (1997), pp. 201–238.

[4] L. Ambrosio, N. Fusco, and D. Pallara, *Functions of Bounded Variation and Free Discontinuity Problems*, The Clarendon Press, Oxford University Press, New York, 2000.

[5] H. Attouch, J. Bolte, and B. F. Svaiter, *Convergence of descent methods for semi-algebraic and tame problems: Proximal algorithms, forward-backward splitting, and regularized Gauss–Seidel methods*, Math. Program., 137 (2013), pp. 91–129.

[6] H. Attouch, G. Buttazzo, and G. Michaille, *Variational Analysis in Sobolev and BV Spaces: Applications to PDEs and Optimization*, 2nd ed., MOS-SIAM Ser. Optim. 17, SIAM, Philadelphia, 2014.

[7] M. Benning and M. Burger, *Ground states and singular vectors of convex variational regularization methods*, Methods Appl. Anal., 20 (2013), pp. 295–334.

[8] L. Biegler, G. Biros, O. Ghattas, M. Heinkenschloss, D. Keyes, B. Mallick, Y. Marzouk, L. Tenorio, B. van Bloemen Waanders, and K. Willcox, eds., *Large-Scale Inverse Problems and Quantification of Uncertainty*, John Wiley & Sons, Chichester, UK, 2011.

[9] G. Bouchitté and G. Buttazzo, *New lower semicontinuity results for nonconvex functionals defined on measures*, Nonlinear Anal., 15 (1990), pp. 679–692.
[10] K. Bredies, K. Kunisch, and T. Pock, *Total generalized variation*, SIAM J. Imaging Sci., 3 (2010), pp. 492–526.
[11] T. Bui-Thanh, K. Willcox, and O. Ghattas, *Model reduction for large-scale systems with high-dimensional parametric input space*, SIAM J. Sci. Comput., 30 (2008), pp. 3270–3288.
[12] E. Casas, K. Kunisch, and C. Pola, *Regularization by functions of bounded variation and applications to image enhancement*, Appl. Math. Optim., 40 (1999), pp. 229–257.
[13] V. Caselles, A. Chambolle, and M. Novaga, *The discontinuity set of solutions of the TV denoising problem and some extensions*, Multiscale Model. Simul., 6 (2007), pp. 879–894.
[14] T. F. Chan and S. Esedoḡlu, *Aspects of total variation regularized L^1 function approximation*, SIAM J. Appl. Math., 65 (2005), pp. 1817–1837.
[15] X. Chen, *Smoothing methods for nonsmooth, nonconvex minimization*, Math. Program., 134 (2012), pp. 71–99.
[16] X. Chen and W. Zhou, *Smoothing nonlinear conjugate gradient method for image restoration using nonsmooth nonconvex minimization*, SIAM J. Imaging Sci., 3 (2010), pp. 765–790.
[17] J. C. De los Reyes and C.-B. Schönlieb, *Image denoising: Learning the noise model via nonsmooth PDE-constrained optimization*, Inverse Probl. Imaging, 7 (2013), pp. 1183–1214.
[18] S. Delladio, *Lower semicontinuity and continuity of functions of measures with respect to the strict convergence*, Proc. Roy. Soc. Edinburgh Sect. A, 119 (1991), pp. 265–278.
[19] J. E. Dennis, Jr., and R. B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Classics Appl. Math. 16, SIAM, Philadelphia, 1996.
[20] V. Duval, J.-F. Aujol, and Y. Gousseau, *The TVL1 model: A geometric point of view*, Multiscale Model. Simul., 8 (2009), pp. 154–189.
[21] L. C. Evans and R. F. Gariepy, *Measure Theory and Fine Properties of Functions*, CRC Press, Boca Raton, FL, 1992.
[22] I. Fonseca and G. Leoni, *Modern Methods in the Calculus of Variations: L^p Spaces*, Springer, New York, 2007.
[23] S. Geman and D. Geman, *Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images*, IEEE Trans. Pattern Anal. Mach. Intell., 6 (1984), pp. 721–741.
[24] E. Giusti, *Minimal Surfaces and Functions of Bounded Variation*, Monogr. Math. 80, Birkhäuser Verlag, Basel, 1984.
[25] E. Haber, L. Horesh, and L. Tenorio, *Numerical methods for experimental design of large-scale linear ill-posed inverse problems*, Inverse Problems, 24 (2008), 055012.
[26] E. Haber and L. Tenorio, *Learning regularization functionals—supervised training approach*, Inverse Problems, 19 (2003), pp. 611–626.
[27] M. Hintermüller and T. Wu, *Nonconvex TV^q-models in image restoration: Analysis and a trust-region regularization–based superlinearly convergent solver*, SIAM J. Imaging Sci., 6 (2013), pp. 1385–1415.
[28] M. Hintermüller and T. Wu, *A superlinearly convergent R-regularized Newton scheme for variational models with concave sparsity-promoting priors*, Comput. Optim. Appl., 57 (2014), pp. 1–25.
[29] J. Huang and D. Mumford, *Statistics of natural images and models*, in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 1, 1999.
[30] J. Kristensen and F. Rindler, *Relaxation of signed integral functionals in BV*, Calc. Var. Partial Differential Equations, 37 (2010), pp. 29–62.
[31] P.-L. Lions, *The concentration-compactness principle in the calculus of variations. The limit case.* I, Rev. Mat. Iberoamericana, 1 (1985), pp. 145–201.
[32] P Mattila, *Geometry of Sets and Measures in Euclidean Spaces: Fractals and Rectifiability*, Cambridge University Press, Cambridge, UK, 1999.
[33] M. Nikolova, *Minimizers of cost-functions involving nonsmooth data-fidelity terms. Application to the processing of outliers*, SIAM J. Numer. Anal., 40 (2002), pp. 965–994.
[34] M. Nikolova, M. K. Ng, S. Zhang, and W.-K. Ching, *Efficient reconstruction of piecewise constant images using nonsmooth nonconvex minimization*, SIAM J. Imaging Sci., 1 (2008), pp. 2–25.
[35] P. Ochs, Y. Chen, T. Brox, and T. Pock, *iPiano: Inertial Proximal Algorithm for Non-Convex Optimization*, preprint, arXiv:1404.4805v1 [cs.CV], 2014.

[36] P. Ochs, A. Dosovitskiy, T. Brox, and T. Pock, *An iterated* $L1$ *algorithm for non-smooth nonconvex optimization in computer vision*, in Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 1759–1766.
[37] E. Paolini and E. Stepanov, *Optimal transportation networks as flat chains*, Interfaces Free Bound., 8 (2006), pp. 393–436.
[38] J. Rabin and G. Peyré, *Wasserstein regularization of imaging problem*, in Proceedings of the 2011 18th IEEE International Conference on Image Processing, 2011, pp. 1541–1544.
[39] F. Rindler and G. Shaw, *Strictly continuous extensions of functionals with linear growth to the space* BV, Q. J. Math., 66 (2015), pp. 953–978.
[40] L. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Phys. D, 60 (1992), pp. 259–268.
[41] W. Rudin, *Real and Complex Analysis*, McGraw-Hill, New York, Toronto, London, 1966.
[42] P. Swoboda and C. Schnörr, *Convex variational image restoration with histogram priors*, SIAM J. Imaging Sci., 6 (2013), pp. 1719–1735.
[43] T. Valkonen, *Optimal transportation networks and stations*, Interfaces Free Bound., 11 (2009), pp. 569–597.
[44] T. Valkonen, *Transport equation and image interpolation with SBD velocity fields*, J. Math. Pures Appl. (9), 95 (2011), pp. 459–494.
[45] T. Valkonen, *Strong polyhedral approximation of simple jump sets*, Nonlinear Anal., 75 (2012), pp. 3641–3671.
[46] T. Valkonen, *The jump set under geometric regularisation. Part* 2: *Higher-order approaches*, submitted; available online at http://arxiv.org/abs/1407.2334.
[47] T. Valkonen, *The jump set under geometric regularization. Part* 1: *Basic technique and first-order denoising*, SIAM J. Math. Anal., 47 (2015), pp. 2587–2629.
[48] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, *Image quality assessment: From error visibility to structural similarity*, IEEE Trans. Image Process., 13 (2004), pp. 600–612.
[49] B. White, *Rectifiability of flat chains*, Ann. of Math. (2), 150 (1999), pp. 165–184.