SIAM J. IMAGING SCIENCES
Vol. 8, No. 2, pp. 1090–1122

© 2015 Society for Industrial and Applied Mathematics

Structure Tensor Total Variation*

Stamatios Lefkimmiatis†, Anastasios Roussos‡, Petros Maragos§, and Michael Unser¶

Abstract. We introduce a novel generic energy functional that we employ to solve inverse imaging problems within a variational framework. The proposed regularization family, termed as structure tensor total variation (STV), penalizes the eigenvalues of the structure tensor and is suitable for both grayscale and vector-valued images. It generalizes several existing variational penalties, including the total variation seminorm and vectorial extensions of it. Meanwhile, thanks to the structure tensor's ability to capture first-order information around a local neighborhood, the STV functionals can provide more robust measures of image variation. Further, we prove that the STV regularizers are convex while they also satisfy several invariance properties w.r.t. image transformations. These properties qualify them as ideal candidates for imaging applications. In addition, for the discrete version of the STV functionals we derive an equivalent definition that is based on the patch-based Jacobian operator, a novel linear operator which extends the Jacobian matrix. This alternative definition allow us to derive a dual problem formulation. The duality of the problem paves the way for employing robust tools from convex optimization and enables us to design an efficient and parallelizable optimization algorithm. Finally, we present extensive experiments on various inverse imaging problems, where we compare our regularizers with other competing regularization approaches. Our results are shown to be systematically superior, both quantitatively and visually.

Key words. structure tensor, patch-based Jacobian, image reconstruction, convex optimization, total variation, inverse problems

AMS subject classifications. 65F22, 65K10, 90C06, 90C25, 49M29

DOI. 10.1137/14098154X

1. Introduction.

Inverse problems are ubiquitous in science and engineering and they have been an important topic of interest for many years. Nowadays, with the advent of modern imaging techniques, inverse problems appear increasingly in a host of imaging applications ranging from microscopy and medical imaging to remote sensing and astronomical imaging [6]. They also appear in a plethora of computer vision applications, including motion estimation, image registration, stereo, and dense three-dimensional (3D) reconstruction [42]. These problems pertain to estimating unknown features of interest from partial or indirect measurements

*Received by the editors August 8, 2014; accepted for publication (in revised form) February 5, 2015; published electronically May 7, 2015. The first and second authors have a joint first authorship.
http://www.siam.org/journals/siims/8-2/98154.html

†Department of Mathematics, University of California, Los Angeles, Los Angeles, CA 90095 (stamatis@math.ucla.edu). The research of this author was supported by the Swiss National Science Foundation (SNF) under grant P300P2_151325.

‡Department of Computing, Imperial College London, London NW1 2FD, UK (troussos@cs.ucl.ac.uk). The research of this author was supported by the European Research Council Starting Grant 204871-HUMANIS.

§School of ECE, National Technical University of Athens (maragos@cs.ntua.gr). The research of this author was supported in part by the grant COGNIMUSE under the ARISTEIA Action co-funded by the European Social Fund and Greek National Resources.

¶School of Engineering, Swiss Federal Institute of Technology, Lausanne (michael.unser@epfl.ch). The research of this author was supported in part by the European Research Council under grant ERC-2010-AdG 267439-FUN-SP.

of image attributes. In practice, inverse imaging problems are typically ill-posed. This is to say that the equations relating the object of interest to the measurements are not enough by themselves to uniquely characterize the solution; a model of known properties of the underlying image is necessary for the recovered information to be physically and/or statistically meaningful.

A common strategy for dealing with ill-posed inverse problems is the variational approach. The main element of this framework is the regularization functional, whose role is to favor solutions that exhibit desirable properties. The proper selection of a regularizer is of paramount importance since it significantly influences the reconstruction quality. During the past years, this has triggered an increasing research interest on the design of regularizers able to encode important properties of natural images.

One of the most popular regularizers for imaging applications is the total variation (TV) seminorm [47]. TV is a convex functional which allows the use of powerful optimization techniques. Usually it is intended for grayscale images but it has also been extended to apply to vector-valued images; see, for instance, [7, 30, 48, 55, 58]. The main reason for its success is its ability to reconstruct images with sharp, well-preserved edges. However, it has the drawbacks of oversmoothing homogenous regions and creating staircase artifacts [16].

To overcome the limitations of TV, several regularization functionals have been proposed in the literature. Among them there are extensions that adapt the penalization of image variation, by locally varying the regularization weight [22, 29, 32] or by incorporating an anisotropicity that is steered by the local structures of the image [5, 49, 53]. However, these methods extract the local information either from the input image in a preprocessing step or as an additional unknown function of the optimization problem, not directly depending on the underlying image. On the contrary, the so-called *anisotropic TV* (ATV) [33, 38] adapts the penalization of image variation by introducing a "diffusion" tensor that depends on the structure tensor of the unknown image itself. Nevertheless, in this case the adaptivity on image structures is heuristically designed, similarly to the design of the coherence-enhancing diffusion [57]. Another related regularizer is the Beltrami functional [52] of vector-valued images, which corresponds to the area of the generalized graph of the image embedded in a high dimensional space, using the induced metric. This framework has been recently generalized and unified with the Mumford–Shah functional [51], while in [46] an extension of the Beltrami framework that uses image patches has been introduced. A different regularization approach that also aims to eliminate the staircase effect of TV, involves functionals that can encode higher-order information. These regularizers promote piecewise-linear instead of piecewise-constant reconstructions and are either combined with TV as in [13, 41] or used in a standalone fashion as the total generalized variation (TGV) [10, 50] and the Hessian Schatten-norm regularizers [37]. Finally, there also exist nonlocal variational techniques that exploit the nonlocal self-similarity property of natural images. This is made possible by employing functionals that allow nonlocal interactions between image points [24, 28]. Among the most efficient regularizer of this category is the nonlocal TV functional (NLTV) [28]. Discrete-domain extensions of NLTV for vector-valued images are studied in [17, 18].

1.1. Contributions. In this work, we propose a family of convex regularizers that penalize the image variation at every point of the domain by taking into account the information in a local neighborhood. This is accomplished by adopting a penalty that depends on the

eigenvalues of the structure tensor of the underlying image. In detail, our key contributions are the following:

1. The introduction of a family of energy functionals that generalize TV and vectorial extensions of it. Our regularizers preserve all the favorable properties of TV, such as convexity and invariance w.r.t. image transformations. Moreover, they provide a more robust and richer measure of image variation by exploiting additional information from the neighborhood of every point. We use these functionals in a variational framework to obtain regularized solutions of ill-posed inverse imaging problems.
2. The proof that our regularization criteria, which involve the eigenvalues of the nonlinear structure tensor operator, are convex and invariant to image transformations.
3. The concept of the patch-based Jacobian as a novel neighborhood-aware extension of the discrete Jacobian operator. We employ this operator to obtain an equivalent definition for the discrete version of our proposed energy functionals. This further allows us to derive a dual problem formulation.
4. The development of a practical first-order algorithm for evaluating the proximal map of the introduced regularizers. Our optimization algorithm copes with the nonsmoothness of the energy functionals and leads to efficient numerical solutions.
5. An extensive quantitative experimental evaluation of our method on several inverse problems for different degradation conditions. The comparisons are performed on a large dataset of both graylevel and color images, where we compare the proposed method with TV and vectorial extensions of it, as well as with the state-of-the-art TGV [10]. The results show that in all the studied problems, our method systematically outperforms all other evaluated techniques.

We note that in the conference papers [36, 46], which contain preliminary parts of this work, we established a connection between the minimization of a class of structure tensor-based regularizers, which also includes a relaxation of the proposed family of regularizers, and a novel generic type of nonlinear anisotropic diffusion. This type of diffusion is driven by a diffusion tensor that depends on the solution itself and is spatially regularized (in the sense that it involves convolutions with a kernel, as, e.g., in [55, 57]). This connection, in conjunction with the efficient optimization strategy that we present here, implies that we can employ powerful convex optimization techniques for implementing this type of anisotropic diffusion schemes. We intend to present in depth the relations of the proposed regularizers and anisotropic diffusion in a separate article.

An important extension of the present work with respect to the conference paper in [36] is that we introduce a continuous-domain formulation of the problem, which is guided by fundamental invariance properties (translation and rotation invariance, 1-homogeneity), and further provide a proof of convexity. These new results are presented in Theorem 2.1. Additionally, we establish a bound for the Lipschitz constant of the dual objective function that arises in the evaluation of the proximal operator of our functionals (Proposition 4.2). This allows us to use a fixed step-size in our gradient-based optimization algorithm. Further, we provide extended comparisons on several inverse imaging problems, including two new applications, namely, image reconstruction from sparse Fourier measurements and image magnification. Finally, for the deblurring problem that we study, we consider an improved forward model than the one we used in [36]. This formulation avoids the need of making any specific assumption about

the image boundaries and holds the regularizer responsible for reconstructing the image in a way that best explains the observation.

This paper is organized as follows: in section 2, we introduce the proposed family of regularizers and motivate their use. In section 3, we focus on the discrete domain and introduce the patch-based Jacobian operator. In section 4, we design our optimization algorithm, while in section 5 we assess the reconstruction performance of our approach with extensive experiments on several linear inverse imaging problems. Conclusions are drawn in section 6, while all the mathematical proofs are provided in Appendix A.

2. Regularization for inverse problems.

2.1. Problem formulation and variational recovery.

The relationship between the image of interest and the measurements is given by the physical forward model that provides the basis of the acquisition process. For many imaging applications, the acquisition is well-described by a linear process, which can be mathematically formulated as

$$\boldsymbol{v}(\boldsymbol{x}) \sim \mathcal{N}(\mathcal{A}\boldsymbol{u}(\boldsymbol{x})). \tag{2.1}$$

In this case, $\boldsymbol{u}(\boldsymbol{x}) = [u_1(\boldsymbol{x}) \dots u_M(\boldsymbol{x})] : \mathbb{R}^2 \mapsto \mathbb{R}^M$ represents the generic vector-valued image with M channels that we wish to reconstruct and $\mathcal{A}$ is a linear operator that provides a mapping from the space of underlying images to the space of measurements. The symbol $\mathcal{N}$ represents the measurement noise which accounts for all possible types of errors during the acquisition, including stochastic noise.

The recovery of $\boldsymbol{u}$ from the measurements $\boldsymbol{v}$ in (2.1) belongs to the category of linear inverse problems. For most cases of practical interest, the operator $\mathcal{A}$ is either ill-conditioned or singular. This kind of *ill-posedness* [56] is dealt with within a variational framework, where the reconstruction of $\boldsymbol{u}$ is cast as a minimization problem of an objective function of the form

$$\mathcal{E}(\boldsymbol{u}) = \varphi(\mathcal{A}\boldsymbol{u}) + \tau\psi(\boldsymbol{u}). \tag{2.2}$$

Such a cost function consists of (a) the *data fidelity* term $\varphi(\mathcal{A}\boldsymbol{u})$, which measures how well a candidate solution explains the observed data, and (b) the *regularizer* $\psi(\boldsymbol{u})$ which encodes any available prior information about the underlying image. The exact form of the data fidelity term depends on the assumed noise model perturbing the measurements. From a Bayesian point of view, the overall reconstruction approach corresponds to either a penalized maximum likelihood or a maximum a posteriori estimation problem [26].

2.2. TV regularization.

As stated earlier, TV [47] applies to grayscale images $u(M = 1)$ and for smooth images is defined as

$$\mathrm{TV}(u) = \int_{\mathbb{R}^2} \|\nabla u\|_2 \, \mathrm{d}\boldsymbol{x}. \tag{2.3}$$

Apart from the summation over the domain, another key ingredient of TV is the gradient magnitude $\|\nabla u\|_2$ of the underlying image. Their combination leads to an L_1 weighting that does not overpenalize high variations of u. Therefore, the reconstructed images feature sharp edges, which are visually appealing. Due to this favorable property, TV has been

extended to vector-valued images $\boldsymbol{u}$. A popular extension that retains TV's properties and at the same time introduces a coupling between the different channels of $\boldsymbol{u}$ is the *vectorial TV* (VTV) [7, 48]. VTV has a definition similar to (2.3), where the gradient magnitude is replaced by its straight-forward vectorial extension $\|\nabla \boldsymbol{u}\|_2 = \{\sum_{i=1}^{M} \|\nabla u_i\|_2^2\}^{\frac{1}{2}}$.

A known drawback of TV is that it favors piecewise-constant solutions and, thus, it can create strong staircase artifacts in smooth regions of the image. Furthermore, a basic shortcoming of both TV and VTV is that the gradient magnitude, employed to penalize the image variation at every point $\boldsymbol{x}$, is too simple as an image descriptor; it relies only on $\boldsymbol{x}$ without taking into account the available information from its neighborhood. In fact, most of the existing extensions of TV [30, 33, 55, 58] as well as related regularizers [52] share the same drawback: they integrate a penalty of image variation that is completely localized.

To overcome these limitations, we next adopt more sophisticated descriptors of image variations that take into account information from the neighborhood of every point. Consequently, the resulting regularizers are able to provide a richer and more robust measure of image regularity.

2.3. Directional derivatives and the structure tensor. Our goal is to develop neighborhood-aware measures of vectorial image variations that will be incorporated in a novel variational framework. To do so, we first revisit the theory[1] behind the structure tensor [35, 57]. Hereafter, we assume that the vector-valued image $\boldsymbol{u}$ belongs to the Sobolev space $W^{1,2}(\mathbb{R}^2, \mathbb{R}^M)$. Let $\boldsymbol{n}$ be an arbitrary two-dimensional (2D) direction ($\|\boldsymbol{n}\|_2 = 1$). The *vectorial directional derivative* of the vector-valued image $\boldsymbol{u}$ in the direction $\boldsymbol{n}$ and at any specific image point $\boldsymbol{x}$ is given by [21]

$$\frac{\partial \boldsymbol{u}}{\partial \boldsymbol{n}}(\boldsymbol{x}) = (J\boldsymbol{u}(\boldsymbol{x}))\boldsymbol{n}\,, \tag{2.4}$$

where $J\boldsymbol{u}$ is the *Jacobian matrix* of $\boldsymbol{u}$ defined as

$$J\boldsymbol{u}(\boldsymbol{x}) = \begin{bmatrix} \nabla u_1(\boldsymbol{x}) & \dots & \nabla u_M(\boldsymbol{x}) \end{bmatrix}^T . \tag{2.5}$$

The magnitude of the directional derivative $\|\partial \boldsymbol{u}/\partial \boldsymbol{n}\|_2$ yields a measure of the amount of change of the image $\boldsymbol{u}$ at the direction $\boldsymbol{n}$ for any point $\boldsymbol{x}$. This measure is computed by concentrating completely at the point $\boldsymbol{x}$. In order to be more robust and also capture the behavior of the image $\boldsymbol{u}$ in the neighborhood of $\boldsymbol{x}$, we consider instead the weighted *root mean square* (RMS) of $\|\partial \boldsymbol{u}/\partial \boldsymbol{n}\|_2$, which we call (local) *directional variation*:

$$\mathrm{RMS}_K \left\{\|\partial \boldsymbol{u}/\partial \boldsymbol{n}\|_2\right\} = \sqrt{K * \|\partial \boldsymbol{u}/\partial \boldsymbol{n}\|_2^2} = \sqrt{\boldsymbol{n}^T (S_k \boldsymbol{u})\, \boldsymbol{n}}. \tag{2.6}$$

In the above equation $*$ denotes the convolution operation, $K(\boldsymbol{x})$ is a nonnegative, rotationally symmetric convolution kernel, $K(\boldsymbol{x}) = K(|\boldsymbol{x}|)$, that performs the weighted averaging (e.g., a 2D Gaussian) and $S_K \boldsymbol{u}$ is the so-called *structure tensor* [27] of the image $\boldsymbol{u}$ at point $\boldsymbol{x}$ defined as

$$S_K \boldsymbol{u}\,(\boldsymbol{x}) = K * \left[J\boldsymbol{u}^T J\boldsymbol{u}\right](\boldsymbol{x})\,. \tag{2.7}$$

[1] Instead of the usual focus on the eigenvalues of the structure tensor [35, 57], our study focuses on their square roots.

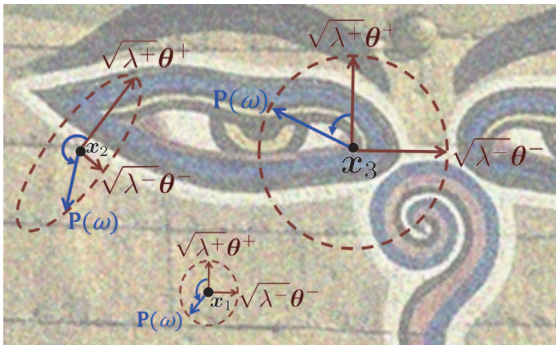


Figure 1. *Structure tensor of an image at sample points $\boldsymbol{x}_1$,$\boldsymbol{x}_2$,$\boldsymbol{x}_3$. The structure tensor is visualized as an ellipse and its unit eigenvectors $\boldsymbol{\theta}^+, \boldsymbol{\theta}^-$ and rooted eigevalues $\sqrt{\lambda^+}$, $\sqrt{\lambda^+}$ are also depicted. For any arbitrary direction $\boldsymbol{n}$, which is characterized by its angle ω with the eigenvector $\boldsymbol{\theta}^+$, there is a corresponding point $\boldsymbol{P}(\omega)$ on the ellipse. The distance of this point from the ellipse center yields the directional variation $V(\omega)$.*

Note that in the above definition, similarly to [55], we do not consider any presmoothing of the image before computing its Jacobian. The reason is that the single convolution with K seems sufficient for the needs of image regularization while it also avoids the introduction of additional complexity in the computations.

Let $\lambda^+ = \lambda^+(S_K\boldsymbol{u}(\boldsymbol{x}))$, $\lambda^- = \lambda^-(S_K\boldsymbol{u}(\boldsymbol{x}))$ be the *eigenvalues* of $S_K\boldsymbol{u}(\boldsymbol{x})$, ordered so that $\lambda^+ \geq \lambda^-$, and $\boldsymbol{\theta}^+, \boldsymbol{\theta}^-$ are the corresponding unit *eigenvectors*. Also, let $\omega \in (-\pi, \pi]$ denote the angle between the direction vector $\boldsymbol{n}$ and the eigenvector $\boldsymbol{\theta}^+$. Using the eigendecomposition of $S_K(\boldsymbol{u})$, the directional variation (2.6) can be expressed as a function of the angle ω:

$$V(\omega) \triangleq \mathrm{RMS}_K\left\{\|\partial\boldsymbol{u}/\partial\boldsymbol{n}\|_2\right\} = \sqrt{\lambda^+\cos^2\omega + \lambda^-\sin^2\omega}\,. \tag{2.8}$$

Now, let us consider the ellipse given by the parametric equation

$$\boldsymbol{P}(\omega) = \cos\omega\sqrt{\lambda^+}\boldsymbol{\theta}^+ + \sin\omega\sqrt{\lambda^-}\boldsymbol{\theta}^-\,, \omega \in [0, 2\pi). \tag{2.9}$$

The major and minor radius of this ellipse are $\sqrt{\lambda^+}$ and $\sqrt{\lambda^-}$, respectively, whereas the direction of the major and minor axis are given by $\boldsymbol{\theta}^+$ and $\boldsymbol{\theta}^-$, respectively. Therefore, this type of ellipse offers a way to visualize the structure tensor $S_K\boldsymbol{u}$ as shown in Figure 1. More importantly, we observe that $V(\omega) = \|\boldsymbol{P}(\omega)\|_2$. This implies that the directional variation $V(\omega)$ can be interpreted as the distance of any point $\boldsymbol{P}(\omega)$ from the center of the ellipse.

From (2.8), we observe that the maximum of the directional variation $V(\omega)$ is $\sqrt{\lambda^+}$ and is achieved for $\omega = 0, \pi$, whereas the minimum of $V(\omega)$ is $\sqrt{\lambda^-}$ and is achieved for $\omega = \pi \pm \pi/2$. In conclusion, the structure tensor measures the geometry of image structures in the neighborhood of each point: its eigenvectors $\boldsymbol{\theta}^+$ and $\boldsymbol{\theta}^-$ describe the orientation of maximum and minimum vectorial variation of $\boldsymbol{u}$ and the square roots of its eigenvalues, $\sqrt{\lambda^+}$ and $\sqrt{\lambda^-}$, describe measures of these variations.

The eigenvalues of the structure tensor offer a rich and discriminative description of the local geometry of the image. When both λ^+ and λ^- are relatively small (e.g., point $\boldsymbol{x}_1$ in Figure 1) there are small vectorial variations around the current point, indicating that the

region is homogeneous. When λ^+ is large and λ^- is small (e.g., point $\boldsymbol{x}_2$ in Figure 1) there are strong variations, but only in a dominant direction. Therefore, the current point is located close to an image edge. When both λ^+ and λ^- are large (e.g., point $\boldsymbol{x}_3$ in Figure 1) there are high variations in all directions, which implies that the current point is close to an image corner.

2.4. The structure tensor TV functional. To design regularizers that integrate a scalar penalty of the local image variation at every image point, we need to consider measures that provide a synopsis of the function $V(\omega)$. Such measures, which are also computed from the eigenvalues of the structure tensor, are the following:

Case 1. RMS value of $V(\omega)$: $((2\pi)^{-1}\int_0^{2\pi} V^2(\omega)\mathrm{d}\omega)^{1/2} = \sqrt{\lambda^+ + \lambda^-}/\sqrt{2}$, as it can be easily verified using equation (2.8).

Case 2. Maximum value of $V(\omega)$: the maximum of $V(\omega)$ w.r.t. ω is $\sqrt{\lambda^+}$.

Case 3. Midrange of $V(\omega)$: this is defined as the average of minimum and maximum values of $V(\omega)$, therefore, it is given by $(\sqrt{\lambda^+} + \sqrt{\lambda^-})/2$.

If for every image point $\boldsymbol{x}$ we define the 2D vector

$$\sqrt{\boldsymbol{\lambda}} = \sqrt{\boldsymbol{\lambda}(S_K\boldsymbol{u}(\boldsymbol{x}))} = \left(\sqrt{\lambda^+(S_K\boldsymbol{u}(\boldsymbol{x}))}, \sqrt{\lambda^-(S_K\boldsymbol{u}(\boldsymbol{x}))}\right),$$

we observe that the above measures of image variation correspond (up to appropriate scale factors) to the following norms of $\sqrt{\boldsymbol{\lambda}}$: $\left\|\sqrt{\boldsymbol{\lambda}}\right\|_2$ (*Case* 1), $\left\|\sqrt{\boldsymbol{\lambda}}\right\|_\infty$ (*Case* 2), $\left\|\sqrt{\boldsymbol{\lambda}}\right\|_1$ (*Case* 3).

This connection further motivates us to consider the more general case of the ℓ_p norm $\left\|\sqrt{\boldsymbol{\lambda}}\right\|_p$ $(p \geq 1)$ as a measure of image variation. These norms measure the image variation more coherently and robustly than the gradient magnitude used in TV, as they take into account the variations in its neighborhood. At the same time, they incorporate richer information, since they depend both on the maximum and minimum of the directional variation. In this way, their response is in general better adapted to the image geometry.

The fact that the ℓ_p norms of $\sqrt{\boldsymbol{\lambda}}$ are able to measure the local image variation suggests generalizing the TV (2.3) via replacing $\|\nabla u\|_2$ by $\left\|\sqrt{\boldsymbol{\lambda}}\right\|_p$. This leads us to define the following novel class of regularizers, with $\boldsymbol{u} \in W^{1,2}(\mathbb{R}^2, \mathbb{R}^M)$ and $p \geq 1$:

$$\mathrm{STV}_p(\boldsymbol{u}) = \int_{\mathbb{R}^2} \left\|\left(\sqrt{\lambda^+}, \sqrt{\lambda^-}\right)\right\|_p \mathrm{d}\boldsymbol{x}. \tag{2.10}$$

We call the resulting family of regularizers *structure tensor TV* (STV), due to its dependence on the structure tensor of the image and to the fact that it generalizes TV.

Next, we provide a result about the properties that the proposed family of regularizers satisfies. Specifically, the following theorem, whose proof is given in Appendix A.2, verifies that all the favorable properties of TV are indeed preserved by STV.

Theorem 2.1. *The STV functional,* $\mathrm{STV}_p(\boldsymbol{u})$*, is translation and rotation invariant,* 1*-homogeneous, and convex.*

Connections to TV-based functionals. Several existing variational methods emerge as special cases of the proposed regularizers when $K(\boldsymbol{x})$ is chosen to be the Dirac delta $\delta(\boldsymbol{x})$ (degenerated case where no convolution takes place at the computation of the structure tensor). Specifically,

if we consider grayscale images ($M = 1$) and $K(\boldsymbol{x}) = \delta(\boldsymbol{x})$, then all the regularizers of the type (2.10) are equivalent to TV. The reason is that in this case, $\lambda^+ = \|\nabla u\|_2^2$ and λ^- is always 0, therefore $\left\|\sqrt{\boldsymbol{\lambda}}\right\|_p = \|\nabla u\|_2$ for any $p \geq 1$. Furthermore, if we consider vector-valued images ($M > 1$) and $K(\boldsymbol{x}) = \delta(\boldsymbol{x})$, then the choice of $p = 2$ corresponds to the VTV, while the choice of $p = \infty$ corresponds to the regularizer of [30], which the authors call *natural vectorial TV* (TV_J). Finally, by choosing $K(\boldsymbol{x}) = \delta(\boldsymbol{x})$ and $p = 1$ we obtain a novel vectorial extension of TV that we initially introduced in [36]. This regularizer penalizes the nuclear norm of the Jacobian. We denote this regularizer as $\mathrm{STV}_1(\mathrm{loc})$, since it corresponds to a completely localized version of STV_1, where the corresponding structure tensor is formed by only combining the information from the different image channels without any spatial smoothing.

3. Discrete STV. In practice, for all inverse problems of interest we need to deal with discrete data. Therefore, from now on we solely focus on the discrete problem formulation of (2.1) and its manipulation. In this case, $\boldsymbol{A}$ will denote the system matrix that models the spatial response of the imaging device, while $\boldsymbol{v}$ and $\boldsymbol{u}$ will denote the vectorized versions of the observed image and the image to be estimated, respectively.

3.1. Notation and definitions. Hereafter, lower-case bold-faced letters are used for vectors while upper-case bold-faced letters are reserved for matrices and multidimensional arrays. The set of unitary matrices is denoted as $\mathbb{U}^N = \left\{\boldsymbol{X} \in \mathbb{C}^{N\times N} : \boldsymbol{X}^{-1} = \boldsymbol{X}^H\right\}$, where $\mathbb{C}$ is the set of complex numbers and $(\cdot)^H$ is the Hermitian transpose. Further, the set of positive semidefinite diagonal matrices is denoted as $\mathbb{D}^{N_1\times N_2} = \{\boldsymbol{X} \in \mathbb{R}_+^{N_1\times N_2} : \boldsymbol{X}(i,j) = 0\ \forall\, i \neq j\}$.

We assume that the discretized vector-valued image $\boldsymbol{u} = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_M)$ is defined on a rectangular grid with unary steps and contains N pixels. Each channel m of $\boldsymbol{u}(m = 1, \ldots, M)$ is rasterized in the vector $\boldsymbol{u}_m \in \mathbb{R}^N$ and by combining all the image channels, we have that $\boldsymbol{u} \in \mathbb{R}^{NM}$. We use the index $n = 1, \ldots, N$ to refer to a specific pixel of the grid and we denote by $\boldsymbol{x}_n$ its coordinates. Further, the convolution kernel K is also discretized and truncated in order to have compact support $\mathcal{P} = \{-L_K, \ldots, L_K\}^2$, where L_K is a nonnegative integer much smaller than the dimensions of the discrete image. We denote by $[\boldsymbol{S}_K\boldsymbol{u}]_n$ the discrete structure tensor of $\boldsymbol{u}$ evaluated at pixel n, which is defined by adopting a discrete gradient in (2.5) and discrete convolution in (2.7). To properly handle values outside the image boundaries in the definition of this convolution, we assume a type of image extension beyond the boundaries such as mirroring, periodic extension, or zero padding. In our implementation we have used the mirroring boundary extension but the framework that we describe next can easily accommodate all the boundary extensions. Now, we can formally define the discrete STV as follows:

$$\mathrm{STV}_p(\boldsymbol{u}) = \sum_{n=1}^{N} \left\| \left(\sqrt{\lambda_n^+}, \sqrt{\lambda_n^-} \right) \right\|_p , \tag{3.1}$$

where $p \geq 1$ and λ_n^+, λ_n^- denote the two nonnegative eigenvalues of the discrete structure tensor $[\boldsymbol{S}_K\boldsymbol{u}]_n$ at pixel n.

Finally, the following definition will be of help in what follows.

Definition 3.1 (Schatten norms). *Let $\boldsymbol{X} \in \mathbb{C}^{N_1\times N_2}$ be a matrix with the singular-value decomposition (SVD) $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^H$, where $\boldsymbol{U} \in \mathbb{U}^{N_1}$ and $\boldsymbol{V} \in \mathbb{U}^{N_2}$ consist of the singular vectors*

of $\boldsymbol{X}$*, and* $\boldsymbol{\Sigma} \in \mathbb{D}^{N_1 \times N_2}$ *consists of the singular values of* $\boldsymbol{X}$*. The Schatten norm of order* p *(*$\mathcal{S}_p$ *norm) of* $\boldsymbol{X}$*, is defined as*

$$\|\boldsymbol{X}\|_{\mathcal{S}_p} = \left(\sum_{n=1}^{\min(N_1,N_2)} \boldsymbol{\sigma}_n^p \right)^{\frac{1}{p}}, \tag{3.2}$$

where $p \geq 1$*, and* $\boldsymbol{\sigma}_n$ *is the* n*th singular value of* $\boldsymbol{X}$*, which corresponds to the* (n,n) *entry of* $\boldsymbol{\Sigma}$*.*

In words, the Schatten matrix norm of order p corresponds to the ℓ_p norm of the vector that contains the singular values of the matrix. Three of the most popular matrix norms of this family are the $\mathcal{S}_1$ (nuclear), $\mathcal{S}_2$ (Frobenius), and $\mathcal{S}_\infty$ (spectral) norms, corresponding to the ℓ_1, ℓ_2, and ℓ_∞ norms of the matrix singular values, respectively.

3.2. Patch-based Jacobian operator. As described earlier, the proposed regularizers depend on the eigenvalues of the structure tensor. Their current form is challenging to work with, mainly because of the nonlinearity of the operator and the presence of the convolution kernel K. These two features pose significant difficulties in the efficient minimization of our functionals.

To overcome these difficulties, in this section we introduce an alternative formulation of the proposed regularizers. This equivalent formulation depends on a novel generalization of the Jacobian of an image based on local weighted patches. This new operator, which we call *patch-based Jacobian*, contains weighted shifted versions of the Jacobian of $\boldsymbol{u}$, whose weights are determined by the smoothing kernel K. As we will show next, the alternative definition of STV provides more intuition and paves the way for employing tools from convex optimization, which in their turn facilitate the development of an efficient optimization strategy.

We define the *patch-based Jacobian* of an image $\boldsymbol{u}$ as the linear mapping $\boldsymbol{J}_K : \mathbb{R}^{NM} \mapsto \mathcal{X}$, where $\mathcal{X} \triangleq \mathbb{R}^{N \times (LM) \times 2}$ and $L = (2L_K + 1)^2$. This definition implies that if we apply the patch-based Jacobian on the nth pixel of $\boldsymbol{u}$, then the result is a matrix of size $(LM \times 2)$, which we denote by $[\boldsymbol{J}_K \boldsymbol{u}]_n$. This matrix is constructed by (1) evaluating the discrete versions of the $M \times 2$ Jacobian matrices of $\boldsymbol{u}$ for all the pixels $\{\boldsymbol{x}_n - \boldsymbol{s} : \boldsymbol{s} \in \mathcal{P}\}$ in the $\mathcal{P}$-neighborhood of the pixel $\boldsymbol{x}_n$, (2) weighting these matrices with the window function $\boldsymbol{\omega}[\boldsymbol{s}] = \sqrt{K[\boldsymbol{s}]}$, and (3) stacking all of them vertically to form the final matrix. Formally, the patch-based Jacobian can be defined as

$$[\boldsymbol{J}_K \boldsymbol{u}]_n = \left([\tilde{\nabla} \boldsymbol{u}_1]_n, \ldots, [\tilde{\nabla} \boldsymbol{u}_M]_n \right)^T, \tag{3.3}$$

where

$$[\tilde{\nabla} \boldsymbol{u}_m]_n = \left([T_{\boldsymbol{s}_1,\boldsymbol{\omega}} \circ \nabla \boldsymbol{u}_m]_n, \ldots, [T_{\boldsymbol{s}_L,\boldsymbol{\omega}} \circ \nabla \boldsymbol{u}_m]_n \right), \tag{3.4}$$

∇ is the discrete gradient, $(\cdot)^T$ is the transpose operator, $\circ$ denotes the composition of operators, the shift vectors $\boldsymbol{s}_l$ $(l = 1, \ldots, L)$ are the elements of the lattice $\mathcal{P}$, and $T_{\boldsymbol{s}_l,\boldsymbol{\omega}}$ is a weighted translation operator. The latter takes into account the mirror boundary conditions and is defined as

$$[T_{\boldsymbol{s}_l,\boldsymbol{\omega}} \circ \nabla \boldsymbol{u}_m]_n = \boldsymbol{\omega}[\boldsymbol{s}_l] \nabla \boldsymbol{u}_m [\boldsymbol{x}_n - \boldsymbol{s}_l]. \tag{3.5}$$

Next, we equip the space $\mathcal{X}$, which is the target space of $\boldsymbol{J}_K$, with the inner product $\langle \cdot\,,\,\cdot \rangle_{\mathcal{X}}$ and norm $\|\cdot\|_{\mathcal{X}}$. To define them, let $\boldsymbol{X}, \boldsymbol{Y} \in \mathcal{X}$, with $\boldsymbol{X}_n, \boldsymbol{Y}_n \in \mathbb{R}^{(LM)\times 2}\, \forall n = 1, 2, \ldots, N$. Then, we have

$$\langle \boldsymbol{X}\,,\,\boldsymbol{Y} \rangle_{\mathcal{X}} = \sum_{n=1}^{N} \operatorname{trace}\left(\boldsymbol{Y}_n^T \boldsymbol{X}_n\right) \tag{3.6}$$

and

$$\|\boldsymbol{X}\|_{\mathcal{X}} = \sqrt{\langle \boldsymbol{X}\,,\,\boldsymbol{X} \rangle_{\mathcal{X}}} = \left(\sum_{n=1}^{N} \|\boldsymbol{X}_n\|_F^2\right)^{\frac{1}{2}}, \tag{3.7}$$

where $\operatorname{trace}(\cdot)$ is the trace operator of a matrix and $\|\cdot\|_F$ is the Frobenius ($\mathcal{S}_2$) norm. For the Euclidean space $\mathbb{R}^{NM}$ we use the standard inner product $\langle \cdot\,,\,\cdot \rangle_2$ and norm $\|\cdot\|_2$.

Since the patch-based Jacobian is a bounded linear operator, we can further define its adjoint, which performs a reverse mapping from the space $\mathcal{X}$ to $\mathbb{R}^{NM}$. Indeed, the adjoint of the patch-based Jacobian is the discrete linear operator $\boldsymbol{J}_K^* : \mathcal{X} \mapsto \mathbb{R}^{NM}$ such that

$$\langle \boldsymbol{Y}\,,\,\boldsymbol{J}_K \boldsymbol{u} \rangle_{\mathcal{X}} = \langle \boldsymbol{J}_K^* \boldsymbol{Y}\,,\,\boldsymbol{u} \rangle_2 \tag{3.8}$$

should hold for every $\boldsymbol{Y} \in \mathcal{X}$. The following proposition expresses $\boldsymbol{J}_K^*$ in a more suitable form that facilitates its numerical computation. The proof is provided in Appendix A.3.

Proposition 3.2. *The adjoint operator $\boldsymbol{J}_K^*$ of the patch-based Jacobian is given by*

$$[\boldsymbol{J}_K^* \boldsymbol{Y}]_k = \sum_{l=1}^{L} \left[-\mathbf{div}\left(T_{\boldsymbol{s}_l, \boldsymbol{\omega}}^* \circ \boldsymbol{Y}^{((m-1)L+l,:)} \right) \right]_n, \tag{3.9}$$

where $k = (m-1)\,N + n$ with $1 \le n \le N$ and $1 \le m \le M$, $\mathbf{div}$ is the discrete divergence,[2] *$T_{\boldsymbol{s}_l, \boldsymbol{\omega}}^*$ is the adjoint of the weighted translation operator, and $\boldsymbol{Y}_n^{(i,:)}$ is a two-element vector extracted from the ith row of the nth matrix component, $\boldsymbol{Y}_n \in \mathbb{R}^{(LM)\times 2}$, of $\boldsymbol{Y}$.*

3.3. Equivalent formulation of the discrete STV. Having introduced the discrete patch-based Jacobian, we can now employ it to express the discrete structure tensor in a novel way. This is done in the next proposition.

Proposition 3.3. *The discrete structure tensor of $\boldsymbol{u}$ evaluated at the pixel location n can be written in terms of the patch-based Jacobian as*

$$[\boldsymbol{S}_K \boldsymbol{u}]_n = [\boldsymbol{J}_K \boldsymbol{u}]_n^T\, [\boldsymbol{J}_K \boldsymbol{u}]_n\,. \tag{3.10}$$

The proof of Proposition 3.3 can be derived by a direct computation of the right-hand side (r.h.s.) of (3.10) and, thus, is omitted. However, it is worth mentioning that by exploiting

[2] The exact formula for the discrete divergence depends on the discretization scheme one uses for the gradient. In our implementation we consider the discretization that uses forward differences, as in [12]. In this case, the adjoint operator is the discrete divergence that is defined using backward differences (see [12] for details).

the special structure of the patch-based Jacobian we manage to reproduce the convolution operation that takes place in the formulation of the structure tensor.

Now, from (3.10) we have that the 2×2 matrix $[\boldsymbol{S}_K \boldsymbol{u}]_n$ can be decomposed with the help of the $LM \times 2$ matrix $[\boldsymbol{J}_K \boldsymbol{u}]_n$. Therefore, the singular values of $[\boldsymbol{J}_K \boldsymbol{u}]_n$ are equal to $\sqrt{\lambda_n^+}$ and $\sqrt{\lambda_n^-}$. This connection highlights that by using the $\mathcal{S}_p$ matrix norms given in (3.2), we can equivalently write the discrete version of the proposed regularizers (3.1) as

$$\mathrm{STV}_p(\boldsymbol{u}) = \sum_{n=1}^{N} \|[\boldsymbol{J}_K \boldsymbol{u}]_n\|_{\mathcal{S}_p} \tag{3.11}$$

with $p \geq 1$. This equivalent formulation of $\mathrm{STV}_p(\boldsymbol{u})$ in (3.11) provides more intuition about the fact that the proposed regularizers are effective generalizations of TV. In particular, $[\boldsymbol{J}_K \boldsymbol{u}]_n$ encodes the vectorial variation of the image $\boldsymbol{u}$ in the vicinity of the pixel n. Therefore, the Schatten norms of this matrix provide different measures of the local variation of $\boldsymbol{u}$, by taking into account its neighborhood in a weighted manner. In addition, a significant advantage that we gain from the above result, is that the expression (3.1), which involves the eigenvalues of the nonlinear structure tensor, has been transformed into the expression (3.11). The latter expression is much easier to handle in an optimization setting while it also allows us to derive the following result.

Proposition 3.4. *The discrete $STV_p(\mathbf{u})$ regularizer defined in (3.1) is convex w.r.t.* $\mathbf{u}$ $\forall p \geq 1$.

The validity of the above proposition stems from the fact that the discrete STV can be formed as the composition of a mixed vector/matrix norm (ℓ_1-$\mathcal{S}_p$) and the linear operator $\boldsymbol{J}_K$. Therefore, our discrete regularizers retain the important property of convexity that we established with Theorem 2.1 for their continuous counterparts.

4. Discrete STV minimization.

4.1. Proximal map of STV_p . Having introduced the discrete version of the STV functionals, the next step is to develop a scheme for their efficient minimization. In this section, we focus on the evaluation of the proximal map associated with them. The proximal map or the *Moreau proximity operator* of a function ψ is denoted as $\mathrm{prox}_\psi(\cdot) : \mathbb{R}^N \mapsto \mathbb{R}^N$ and is formally defined as [19]

$$\mathrm{prox}_\psi(\mathbf{z}) = \arg\min_{\boldsymbol{u}} \frac{1}{2} \|\boldsymbol{u} - \mathbf{z}\|_2^2 + \psi(\boldsymbol{u}) \,. \tag{4.1}$$

If ψ is a proper, closed, convex function, then a minimizer of (4.1) exists and the solution is unique. For certain choices of the function ψ the solution of (4.1) can be expressed in closed form. However, in several other cases the solution cannot be obtained explicitly, but instead an iterative approach must be followed. This is also the case for our proposed regularizers.

In our case, $\psi(\boldsymbol{u}) = \tau \mathrm{STV}_p(\boldsymbol{u}) + \iota_{\mathcal{C}}$ and the evaluation of the proximal map associated with the STV regularizers corresponds to finding a numerical solution to the following problem, for any $p \geq 1$:

$$\arg\min_{\boldsymbol{u}} \frac{1}{2} \|\boldsymbol{u} - \mathbf{z}\|_2^2 + \tau \mathrm{STV}_p(\boldsymbol{u}) + \iota_{\mathcal{C}}(\boldsymbol{u}) \,. \tag{4.2}$$

In the above formulation, $\mathcal{C}$ is a convex set that represents additional constraints on the solution, such as nonnegativity or box constraints, and $\iota_{\mathcal{C}}$ is its indicator function: $\iota_{\mathcal{C}}(\boldsymbol{u})$ takes the value 0 for $\boldsymbol{u} \in \mathcal{C}$ and ∞ otherwise. The unconstrained minimization case simply corresponds to choosing $\mathcal{C} = \mathbb{R}^N$. Note that from a signal processing standpoint the solution of (4.2) can also be interpreted as a denoising problem under independently and identially distributed (i.i.d.) Gaussian noise, where the STV functionals are used to regularize the solution.

The reason for focusing on the solution of (4.2) is that the proximal map of the STV penalties can serve as the building block for several optimization algorithms, including the fast iterative shrinkage-thresholding algorithm (FISTA) [4], the alternating direction method of multipliers [1, 8, 25], the split-Bregman algorithm [31], and the primal-dual algorithms introduced in [14, 43]. In their turn, these methods are commonly employed to solve a variety of inverse problems in imaging with different data fidelity terms. Therefore, by being able to accurately and efficiently obtain the solution of (4.2), one can also deal with more general inverse problems than merely Gaussian denoising.

Next, we propose a first-order method for solving the problem under study. To proceed, we write our regularizers in the compact form $\mathrm{STV}_p(\boldsymbol{u}) = \|\boldsymbol{J}_K \boldsymbol{u}\|_{1,p}$. Here, $\|\cdot\|_{1,p}$ corresponds to the mixed ℓ_1-$\mathcal{S}_p$ norm which, for an argument $\boldsymbol{X} = \left[\boldsymbol{X}_1^T, \ldots, \boldsymbol{X}_N^T\right]^T \in \mathcal{X}$, is defined as

$$\|\boldsymbol{X}\|_{1,p} = \sum_{n=1}^{N} \|\boldsymbol{X}_n\|_{\mathcal{S}_p}. \tag{4.3}$$

At this point special attention must be paid to the nonsmoothness of the regularizers. The fact that the proposed STV functionals are not differentiable everywhere implies that the minimizer of (4.2) cannot be obtained simply by employing a gradient-based scheme. To circumvent this difficulty, we follow an alternative strategy which involves the derivation of a dual formulation for our problem. To do so, we make use of Lemma 4.1 that follows.

Lemma 4.1 (see [37]). *Let $p \geq 1$, and let q be the conjugate exponent of p, i.e., $\frac{1}{p} + \frac{1}{q} = 1$. Then, the mixed vector-matrix norm $\|\cdot\|_{\infty,q}$ is dual to the mixed vector-matrix norm $\|\cdot\|_{1,p}$.*

Using Lemma 4.1 and the fact that the dual of the dual norm is the original norm [45], we equivalently write (4.3) as

$$\|\boldsymbol{X}\|_{1,p} = \max_{\boldsymbol{\Omega} \in \mathcal{B}_{\infty,q}} \langle \boldsymbol{\Omega}, \boldsymbol{X} \rangle_{\mathcal{X}}, \tag{4.4}$$

where $\mathcal{B}_{\infty,q}$ denotes the ℓ_∞-$\mathcal{S}_q$ unit-norm ball, defined as

$$\mathcal{B}_{\infty,q} = \left\{ \boldsymbol{\Omega} \in \mathcal{X} : \|\boldsymbol{\Omega}_n\|_{\mathcal{S}_q} \leq 1, \forall n = 1, \ldots, N \right\}. \tag{4.5}$$

From the definition of the convex set $\mathcal{B}_{\infty,q}$, it turns out that the orthogonal projection of $\boldsymbol{\Omega}$ can be performed by projecting separately each submatrix $\boldsymbol{\Omega}_n$ on a $\mathcal{S}_q$ unit-norm ball ($\mathcal{B}_{\mathcal{S}_q}$). This property will be used later for designing our first-order minimization method.

Now, combining (3.8) and (4.4) we rewrite (4.2) as

$$\hat{\boldsymbol{u}} = \arg\min_{\boldsymbol{u} \in \mathcal{C}} \frac{1}{2} \|\boldsymbol{u} - \boldsymbol{z}\|_2^2 + \tau \max_{\boldsymbol{\Omega} \in \mathcal{B}_{\infty,q}} \langle \boldsymbol{J}_K^* \boldsymbol{\Omega}, \boldsymbol{u} \rangle_2. \tag{4.6}$$

This formulation naturally leads us to the minimax problem

$$\min_{\boldsymbol{u}\in\mathcal{C}} \max_{\boldsymbol{\Omega}\in\mathcal{B}_{\infty,q}} \mathcal{L}\left(\boldsymbol{u},\boldsymbol{\Omega}\right), \tag{4.7}$$

where $\mathcal{L}\left(\boldsymbol{u},\boldsymbol{\Omega}\right) = \frac{1}{2}\left\|\boldsymbol{u}-\boldsymbol{z}\right\|_2^2 + \tau\langle \boldsymbol{J}_K^*\boldsymbol{\Omega}\,,\,\boldsymbol{u}\rangle_2$.

Since $\mathcal{L}$ is strictly convex w.r.t. $\boldsymbol{u}$ and concave w.r.t. $\boldsymbol{\Omega}$, we have the guarantee that a saddle value of $\mathcal{L}$ is attained [45] and, thus, the order of the minimum and the maximum in (4.7) does not affect the solution. This means that there exists a common saddle point $(\hat{\boldsymbol{u}},\hat{\boldsymbol{\Omega}})$ when the minimum and the maximum are interchanged, i.e.,

$$\min_{\boldsymbol{u}\in\mathcal{C}} \max_{\boldsymbol{\Omega}\in\mathcal{B}_{\infty,q}} \mathcal{L}\left(\boldsymbol{u},\boldsymbol{\Omega}\right) = \mathcal{L}\left(\hat{\boldsymbol{u}},\hat{\boldsymbol{\Omega}}\right) = \max_{\boldsymbol{\Omega}\in\mathcal{B}_{\infty,q}} \min_{\boldsymbol{u}\in\mathcal{C}} \mathcal{L}\left(\boldsymbol{u},\boldsymbol{\Omega}\right). \tag{4.8}$$

According to the above, it is now possible to define two optimization problems, the primal and the dual one. This can be accomplished by identifying the primal and dual objective functions, respectively. In particular, the left-hand side (l.h.s.) of (4.8) corresponds to the minimization of the primal objective function

$$p\left(\boldsymbol{u}\right) = \max_{\boldsymbol{\Omega}\in\mathcal{B}_{\infty,q}} \mathcal{L}\left(\boldsymbol{u},\boldsymbol{\Omega}\right) = \frac{1}{2}\left\|\boldsymbol{u}-\boldsymbol{z}\right\|_2^2 + \tau\left\|\boldsymbol{J}_K\boldsymbol{u}\right\|_{1,p}, \tag{4.9}$$

which matches the problem of interest in (4.2), and the r.h.s. corresponds to the maximization of the dual objective function

$$d\left(\boldsymbol{\Omega}\right) = \min_{\boldsymbol{u}\in\mathcal{C}} \mathcal{L}\left(\boldsymbol{u},\boldsymbol{\Omega}\right) = \frac{1}{2}\left\|\boldsymbol{w}-\Pi_C\left(\boldsymbol{w}\right)\right\|_2^2 + \frac{1}{2}\left(\left\|\mathbf{z}\right\|_2^2 - \left\|\boldsymbol{w}\right\|_2^2\right), \tag{4.10}$$

where $\boldsymbol{w} = \mathbf{z} - \tau\boldsymbol{J}_K^*\boldsymbol{\Omega}$ and Π_C is the orthogonal projection operator on the convex set $\mathcal{C}$. In addition, if we consider the r.h.s. of (4.8), then we can find the minimizer $\hat{\boldsymbol{u}}$ of the primal objective from the maximizer $\hat{\boldsymbol{\Omega}}$ of the dual objective, through the following closed-form relation

$$\hat{\boldsymbol{u}} = \Pi_C\left(\mathbf{z} - \tau\boldsymbol{J}_K^*\hat{\boldsymbol{\Omega}}\right). \tag{4.11}$$

The advantage of this last formulation is that the solution to our problem boils down to computing the maximizer of $d\left(\boldsymbol{\Omega}\right)$. As opposed to $p\left(\boldsymbol{u}\right)$, this objective function is smooth and has a well-defined gradient. Thus, we can exploit gradient information to derive the solution. This would not have been an option if we had considered the original formulation of the problem in (4.2). Next, we provide the details of our approach.

4.2. Solving the dual problem. From the r.h.s. of (4.8) and from (4.10) it follows that $\hat{\boldsymbol{\Omega}}$ can be derived as the solution of the constrained optimization problem

$$\hat{\boldsymbol{\Omega}} = \arg\max_{\boldsymbol{\Omega}\in\mathcal{B}_{\infty,q}} \frac{1}{2}\left\|\boldsymbol{w}-\Pi_C\left(\boldsymbol{w}\right)\right\|_2^2 + \frac{1}{2}\left(\left\|\boldsymbol{z}\right\|_2^2 - \left\|\boldsymbol{w}\right\|_2^2\right). \tag{4.12}$$

Since (4.12) does not admit a closed-form solution ($\boldsymbol{J}_K$ does not have an inverse operator) we have to resort to a numerical iterative scheme. In this work, we choose to pursue a first-order

projected gradient approach. To this end, the essential tools that we need are the computation of the gradient and the orthogonal projection on the convex set $\mathcal{B}_{\infty,q}$, defined in (4.5).

First, we proceed by deriving the gradient. To do so, we use that the gradient of a function $h(\boldsymbol{u}) = \|\boldsymbol{u} - \Pi_C(\boldsymbol{u})\|_2^2$ is well-defined and is equal to $\nabla h(\boldsymbol{u}) = 2(\boldsymbol{u} - \Pi_C(\boldsymbol{u}))$ [3, Lemma 4.1]. Based on this, we obtain

$$\nabla d(\boldsymbol{\Omega}) = \tau \boldsymbol{J}_K \Pi_C(\mathbf{z} - \tau \boldsymbol{J}_K^* \boldsymbol{\Omega}). \tag{4.13}$$

Having the gradient of the dual objective at hand, we also need an efficient way to perform the projection. We accomplish this by projecting independently each of the N components $\boldsymbol{\Omega}_n$ of $\boldsymbol{\Omega}$ on the set $\mathcal{B}_{\mathcal{S}_q} = \{\boldsymbol{X} \in \mathbb{R}^{LM\times 2} : \|\boldsymbol{X}\|_{\mathcal{S}_q} \leq 1\}$.

Now, let $\boldsymbol{X} \in \mathbb{R}^{N_1\times N_2}$ with an SVD $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$ and $\boldsymbol{\Sigma} = \operatorname{diag}(\sigma_1, \ldots, \sigma_N)$ with $N = \min(N_1, N_2)$. According to [37, Proposition 1], the projection of $\boldsymbol{X}$ onto the unit-norm $\mathcal{S}_q$ ball is computed as

$$\Pi_{\mathcal{S}_q}(\boldsymbol{X}) = \boldsymbol{U}\boldsymbol{\Sigma}_q\boldsymbol{V}^T, \tag{4.14}$$

where $\boldsymbol{\Sigma}_q = \operatorname{diag}(\boldsymbol{\sigma}_q)$ and the $\boldsymbol{\sigma}_q$ are the projected singular values of $\boldsymbol{\Sigma}$ onto the ℓ_q unit-norm ball $\mathcal{B}_q = \{\boldsymbol{\sigma} \in \mathbb{R}_+^N : \|\boldsymbol{\sigma}\|_q \leq 1\}$. We note that the projection in (4.14) requires knowledge of the singular vectors and singular values of $\boldsymbol{X}$. In our case $N_2 = 2 < N_1$, and we compute the projection in an efficient way as described next. First, we observe that the matrix $\boldsymbol{X}^T\boldsymbol{X}$ is $N_2 \times N_2$ symmetric with an eigenvalue decomposition $\boldsymbol{V}\boldsymbol{\Sigma}^2\boldsymbol{V}^T$. Therefore, for $N_2 = 2$ both $\boldsymbol{V}$ and $\boldsymbol{\Sigma}$ can be computed in closed form. Moreover, if $\boldsymbol{\Sigma}^+$ is the pseudoinverse matrix of $\boldsymbol{\Sigma}$, defined as $\boldsymbol{\Sigma}^+ = \operatorname{diag}\left(\sigma_1^{-1}, \ldots, \sigma_k^{-1}, 0, \ldots, 0\right)$, with σ_k the smallest nonzero singular value, then $\boldsymbol{U} = \boldsymbol{X}\boldsymbol{V}\boldsymbol{\Sigma}^+$. Using this result we write (4.14) as

$$\Pi_{\mathcal{S}_q}(\boldsymbol{X}) = \boldsymbol{X}\boldsymbol{V}\boldsymbol{\Sigma}^+\boldsymbol{\Sigma}_q\boldsymbol{V}^T, \tag{4.15}$$

which actually avoids the computation of $\boldsymbol{U}$.

Thanks to (4.15) the projection of $\boldsymbol{X}$ reduces to the projection of the singular values. In our case, where $N_2 = 2$, for $q = 1, 2, \infty$ these projections can be computed analytically. Indeed, for $q = 1$ the projection corresponds to the soft-thresholding operation $S_\gamma(\boldsymbol{\sigma}) = \max(\boldsymbol{\sigma} - \gamma, 0)$ [23], where the max operator is applied componentwise and γ is given by

$$\gamma = \begin{cases} 0 & \text{if } \sigma_1 \leq 1 - \sigma_2, \\ \frac{\sigma_1+\sigma_2-1}{2} & \text{if } 1 - \sigma_2 < \sigma_1 \leq 1 + \sigma_2, \\ \sigma_1 - 1 & \text{if } \sigma_1 > 1 + \sigma_2 \end{cases} \tag{4.16}$$

with the singular values sorted in decreasing order, i.e., $\sigma_1 \geq \sigma_2$. For $q = 2$ the projection of $\boldsymbol{X}$ is computed at a reduced computational cost since it does not require the knowledge of the singular values and vectors. In particular, we have $\Pi_{\mathcal{S}_2}(\boldsymbol{X}) = \boldsymbol{X}/\max(1, \|\boldsymbol{X}\|_F)$. Finally, for $q = \infty$ the projection is performed by setting to one the singular values that exceed this value and leaving the rest untouched. For other values of q, we can employ the ℓ_q norm projection algorithm described in [54], but at an increased computational cost. This projection method is based on an iterative proximity algorithm for ℓ_q norms [39].

Algorithm 1 : Proximal algorithm for the STV functionals.

Input: $\boldsymbol{z}$, $\tau > 0$, $p \geq 1$, $\Pi_{\mathcal{C}}$.
Initialization: $\boldsymbol{\Psi}_1 = \boldsymbol{\Omega}_0 = \mathbf{0} \in \mathcal{X}$, $t_1 = 1$.
while *stopping criterion is not satisfied* **do**
  $\boldsymbol{\Omega}_n \leftarrow \Pi_{\mathcal{B}_{\infty,q}} \left(\boldsymbol{\Psi}_n + \frac{1}{8\sqrt{2}\tau} \boldsymbol{J}_K \Pi_{\mathcal{C}} \left(\boldsymbol{z} - \tau \boldsymbol{J}_K^* \boldsymbol{\Psi}_n \right) \right)$;
  $t_{n+1} \leftarrow \frac{1+\sqrt{1+4t_n^2}}{2}$;
  $\boldsymbol{\Psi}_{n+1} \leftarrow \boldsymbol{\Omega}_n + \left(\frac{t_n - 1}{t_{n+1}} \right) (\boldsymbol{\Omega}_n - \boldsymbol{\Omega}_{n-1})$;
  $n \leftarrow n + 1$;
end
return $\Pi_{\mathcal{C}} \left(\boldsymbol{z} - \tau \boldsymbol{J}_K^* \boldsymbol{\Omega}_{n-1} \right)$;

4.3. Numerical algorithm. With all the necessary tools at our disposal, we can now invoke a gradient-based scheme for deriving the solution of the dual problem. In this work, we employ Nesterov's iterative method [40] for smooth functions. This scheme exhibits convergence rates of one order higher than the standard gradient-ascent method. To ensure convergence of the algorithm, we further need to choose an appropriate step size. Since our dual objective is smooth with Lipschitz continuous gradient, we use a constant step size equal to the inverse of the Lipschitz constant of $\nabla d\left(\boldsymbol{\Omega}\right)$. We derive an upper bound of this quantity in the following proposition. The proof can be found in Appendix A.4.

Proposition 4.2. *Let $L\left(d\right)$ denote the Lipschitz constant of the dual objective function d defined in* (4.10)*. Then,*

$$L\left(d\right) \leq 8\sqrt{2}\,\tau^2\,. \tag{4.17}$$

To conclude, in Algorithm 1 we provide a detailed description of all the steps involved in the proposed algorithm for computing the proximal associated with the STV functionals.

4.4. Solution of general inverse problems. So far, we have focussed on computing the proximal map of the STV functionals that solves a simple denoising problem.This basic tool, however, can also be deployed for solving more general inverse problems as we discuss next for the case of a quadratic data term. In this case, the solution is derived as

$$\hat{\boldsymbol{u}} = \arg\min_{\boldsymbol{u} \in \mathcal{C}} \frac{1}{2} \left\| \boldsymbol{v} - \boldsymbol{A}\boldsymbol{u} \right\|_2^2 + \tau \mathrm{STV}_p\left(\boldsymbol{u}\right) \;\; \forall p \geq 1. \tag{4.18}$$

We note that it is possible, by using appropriate optimization tools, for one to also deal with nonquadratic data terms. The description of such techniques, however, exceeds the scope of this paper.

An efficient way to cope with the presence of the operator $\boldsymbol{A}$ in (4.18) is to employ the monotone FISTA (MFISTA) [3] algorithm, which exhibits state-of-the-art convergence rates. The basic idea behind MFISTA is to obtain the minimizer in (4.18) via the successive minimization of a sequence of surrogate functions that upper bound the initial objective function and are easier to minimize. In this context, the solution of (4.18) boils down to iteratively computing the proximal of STV (4.2) for different inputs, which are formed as a combination of solutions from the past iterations of the algorithm. Our main reason for preferring MFISTA

Figure 2. *Thumbnails of the color versions of the* 30 *images used in the experiments. These images are of size of* 481×321 *pixels and are a subset of the Berkeley BSDS*500 *image dataset.*

over other alternative approaches is that it converges quickly to the solution while it does not require any parameter tuning.

5. Applications and experiments.

5.1. Experimental setting. To validate the effectiveness of the proposed STV regularizers we compare them with other related methods on several inverse imaging problems. In particular, we consider the problems of image restoration (denoising and deblurring), image magnification, and image reconstruction from a limited number of Fourier measurements. For the first three problems we present results on both grayscale and color versions of the images shown in Figure 2. For the last problem, our experiments are conducted on the brain phantom images shown in Figure 9. In all cases, the image intensities are normalized so that they lie in the range [0, 1]. Based on this prior information, the image reconstruction obtained by any of the penalty functions under comparison is constrained to lie in the convex set $\mathcal{C} = \left\{\boldsymbol{u} \in \mathbb{R}^N : 0 \leq \boldsymbol{u}_n \leq 1,\ \forall n = 1 \ldots, N\right\}$.

For all the problems under study, we compare our grayscale results against the ones obtained by using TV and the TGV [10]. In our comparisons we consider the second-order extension of TGV, TGV^2, which is the most commonly used in practice and is defined as

$$\text{TGV}^2_\alpha(\boldsymbol{u}) = \min_{\boldsymbol{v} \in \mathbb{R}^{N \times 2}} \|\nabla \boldsymbol{u} - \boldsymbol{v}\|_2 + \alpha \|\mathcal{E}\boldsymbol{v}\|_F . \tag{5.1}$$

Here, the operator $\mathcal{E}$ denotes the symmetrized gradient, $\mathcal{E}\boldsymbol{v} = 0.5(\nabla \boldsymbol{v} + \nabla \boldsymbol{v}^T)$. For the weight α of TGV^2 that balances the first- and second-order terms, we use the value $\alpha = 2$. According to the authors of TGV, this value is suitable for most applications and does not need to be tuned. For the color image reconstruction problems, we compare our results against two vectorial extensions of TV, namely VTV [7, 11, 48] and TV_J [30], and the vectorial extension of TGV^2 [9].

As far as the proposed family of STV regularizers is concerned, we report the results we obtained by employing the STV of $p = 1, 2$, respectively. For the convolution kernel K, used in the definition of the structure tensor, we choose it to be a Gaussian with standard deviation $\sigma = 0.5$ pixels, truncated in a 3×3 pixel window. From our experience this choice provides a good balance between reconstruction quality and computational complexity and it is suitable for a wide range of applications. Furthermore, especially for the case of applying STV_1 on color images, we also consider the kernel K to be a discrete Dirac delta. This results in the localized version of STV_1 denoted as STV_1(loc).

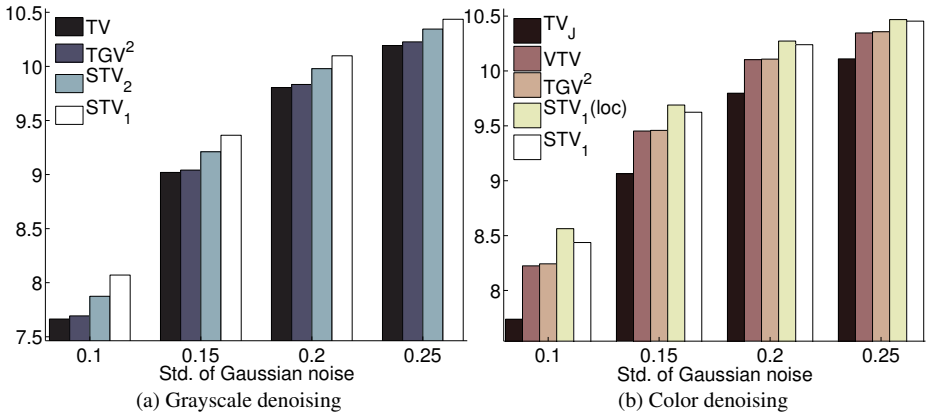


Figure 3. *Image denoising comparisons among different regularizers for four noise levels. The performance is measured in terms of the average SNR improvement (in dB) w.r.t. the noisy input, over all* 30 *images of Figure* 2.

For the objective function minimization we employ the MFISTA framework [3] for all methods but TGV. For TGV, we have experimentally observed that this approach does not work well. Therefore, we employ instead the method proposed by its authors in [9]. Finally, for the sake of consistency among comparisons, all the reported results for each regularizer and each degradation condition are derived using the regularization parameter τ that leads to the best reconstruction performance in terms of peak signal to noise ratio (PSNR).

5.2. Image restoration. For the *image denoising* problem we consider additive i.i.d. Gaussian noise and four different noise levels that correspond to a standard deviation of $\sigma_w = \{0.1, 0.15, 0.2, 0.25\}$, respectively. Regarding the stopping criterion for the optimization algorithm of all methods but TGV, this is set to either reaching a relative normed difference of 10^{-5} between two successive estimates, or a maximum of 100 iterations for solving the dual problem, as described in Algorithm 1. For TGV^2 where we employ the algorithm of [9] we use a maximum of 200 iterations.

In Figure 3 we report the average performance over all tested images of all the methods under comparison. The performance of the regularizers is measured in terms of the average PSNR improvement (ISNR) w.r.t. the noisy input. From these results we observe that, in the grayscale experiments, TV is the worst performing method for all the noise levels. TGV yields an improvement over TV, which reflects the fact that it favors piecewise linear rather than piecewise constant solutions and, thus, it avoids the staircasing artifacts of TV. Both versions of our regularizers, STV_2 and STV_1, systematically outperform TV and TGV^2. This can be attributed to the increased robustness that the neighborhood information offers. Our best performing method is STV_1 which involves the nuclear norm of the patch-based Jacobian. In the color denoising results of Figure 3, we see that TV_J results in the worst overall performance. This can be explained by the fact that this method penalizes only the maximum singular value of the Jacobian, and thus, it completely ignores the information provided by the minimum singular value. VTV yields better measures than TV_J, since it explores additional

Figure 4. *Image denoising examples: close-ups of inputs and optimum results.* Top row*: grayscale denoising of an input with noise level* $\sigma_w = 0.15$*. The ground truth corresponds to the grayscale version of the image in row* 1*, column* 4 *of Figure* 2. Bottom row*: color denoising of an input with noise level* $\sigma_w = 0.2$*. The ground truth corresponds to the image in row* 3*, column* 7 *of Figure* 2*. For each result, the individualized regularization parameter* τ *and corresponding optimum PSNR are reported.*

information about the image variations by imposing an ℓ_2 penalty on the singular values of the Jacobian. Therefore, VTV transfers the characteristics of TV to the vectorial case more effectively than TV_J. Similarly to the improvement of TGV^2 over TV in the grayscale case, TGV^2 yields an improvement over VTV in the color experiments. Again, both of the tested versions of the proposed regularizers, STV_1 and $\mathrm{STV}_1(\mathrm{loc})$, result in the best error measures for all noise levels. In this case, the completely localized version $\mathrm{STV}_1(\mathrm{loc})$ performs better than STV_1, especially in the case of relatively low noise levels. A possible explanation for this outcome is that, in the case of denoising, the coupling of the information from the different channels is more important than the additional spatial information obtained from the neighborhood of each pixel.

For a visual assessment of the reconstruction performance, in Figure 4 we provide representative examples of grayscale and color denoising. In the grayscale example, we observe that the TV's result suffers from strong block artifacts. TGV^2 manages to reduce these artifacts but cannot completely avoid them while it also excessively blurs some edges. STV_2 and STV_1 succeed in removing the noise and at the same time reconstructing the image in a visually more appealing way than TV and TGV^2. Comparing all methods we see that STV_1 yields the most effective reconstruction of image structures; compare, e.g., the reconstructions of the numbers in the athlete's budge, the edges and interior of the athlete's left hand. In the color example, we observe that TV_J creates spurious one-dimensional (1D) segments perpendicular to the dominant edges of the image. Staircasing artifacts are also present in the VTV reconstruction. TGV^2 yields a smoother result, but some of the reconstructed structures are blurred; see, e.g., the edges of the bench back. Finally, STV_1 leads to the best result by avoiding most of the artifacts of the other methods. It sufficiently removes the noise and, in parallel, preserves the sharpness and shape of image edges in the most effective way.

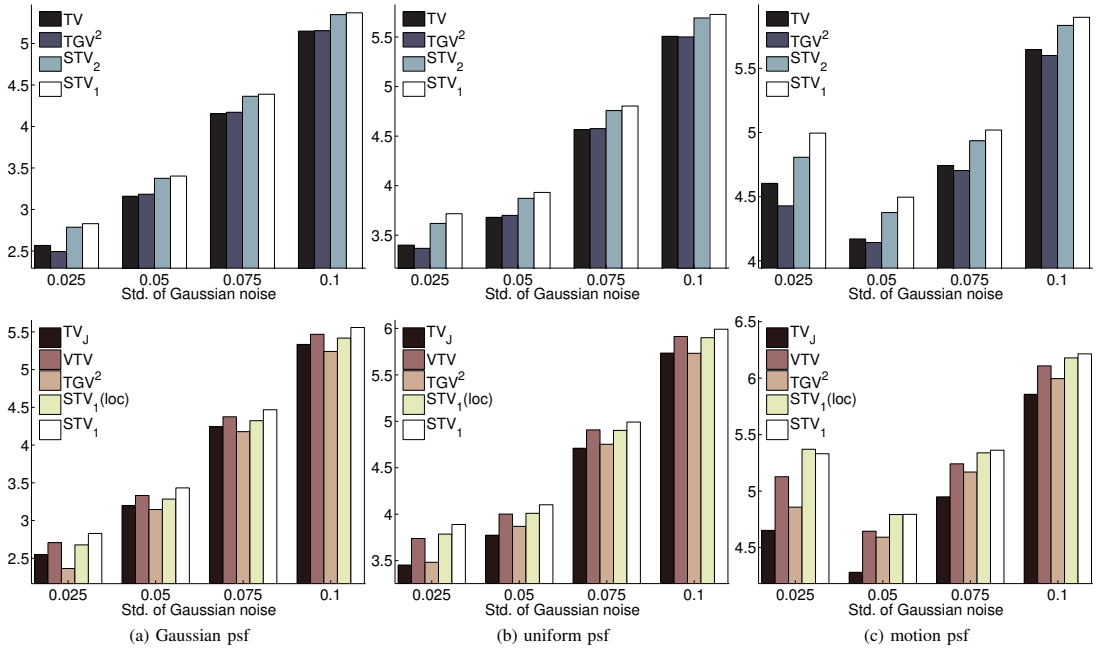
For the *image deblurring* problem the performance of the methods under comparison is assessed for various blurring kernels and different noise levels. In particular, in our set-up we

employ three *point spread functions* (PSFs) to generate blurred versions of the test images. We use a Gaussian PSF of standard deviation $\sigma_b = 4$ and support of 13×13 pixels, a 9×9 uniform (moving-average) PSF, and a real motion blurring kernel[3] which has a support of 19×19 pixels. Additional distortion on the measurements is introduced by adding i.i.d. Gaussian noise of four different noise levels, corresponding to a standard deviation of $\sigma_w = \{0.025, 0.05, 0.075, 0.1\}$. Further, to test the deblurring performance of all the studied methods under more realistic conditions, we do not assume the common circulant convolution model, but instead we adopt the approach followed in [44]. Specifically, in the forward model in (2.1) the degradation operator is expressed as $\boldsymbol{A} = \boldsymbol{M}\boldsymbol{H}$, where $\boldsymbol{H}$ is a circulant convolution operator and $\boldsymbol{M}$ is a mask that keeps only the valid part of the convolution and truncates the circular wraparound at the boundaries of the measurements. The motivation here is that this observation model allows us to avoid making any assumptions about the image boundaries. Instead, we hold the deblurring algorithm responsible for reconstructing the image in a way that bests explains the data. As far as the details of the optimization task are concerned, the stopping criterion is set to the same relative normed difference as previously or a maximum of 100 MFISTA iterations with 10 internal iterations for the proximal evaluation. For TGV^2, since the algorithm of [9] exhibits a slower convergence than MFISTA, we use a maximum of 2000 iterations.

Similarly to the image denoising case, in Figure 5 we report the average performance of all the tested methods. Once more we verify that for grayscale image deblurring the STV regularizers lead to the best results for all considered degradation conditions. When we consider color images, this is still valid with the STV_1 regularizer achieving systematically the best ISNR scores. It also worths mentioning that in this case, unlikely to color image denoising, $\text{STV}_1(\text{loc})$ does not perform better than its patch-based counterpart. This can be attributed to the fact that image deblurring is a more ill-conditioned problem and, thus, the use of a smoothing kernel plays a more important role in the reconstruction. Beyond the ISNR comparisons, the effectiveness of the proposed methods can also be visually appreciated by inspecting the representative deblurring examples provided in Figure 6.

5.3. Image magnification. Image magnification is an inverse problem that is closely related to image deblurring. In particular, the system matrix is expressed as $\boldsymbol{A} = \boldsymbol{S}\boldsymbol{G}$, where $\boldsymbol{S}$ is a masking operator performing the subsampling, while $\boldsymbol{G}$ is an antialiasing smoothing operator applied to the underlying image, prior to subsampling. In our experiments, we consider square subsamplings of the type $d \times d$, where d is the zoom factor and we use two different values for the zoom factor, $d = 3$ and $d = 5$ pixels. Also, we consider the antialiasing operator to be a Gaussian kernel with a standard deviation $0.35d$, defined relatively to the zoom factor. Further, we introduce Gaussian noise in the measurements of a standard deviation $\sigma_w = 0.01$. In Figure 7 we provide measurements of the overall performance of all the methods, both for grayscale and color images, while in Figure 8 we present two magnification examples where we juxtapose the results obtained by using different regularization approaches. Similarly to the other two inverse problems we considered so far, once more we verify that STV_1 not only improves the SNR but also leads to reconstructions of enhanced visual quality.

[3]The blurring kernel was obtained from http://www.wisdom.weizmann.ac.il/~levina/papers/LevinEtalCVPR09Data.rar.



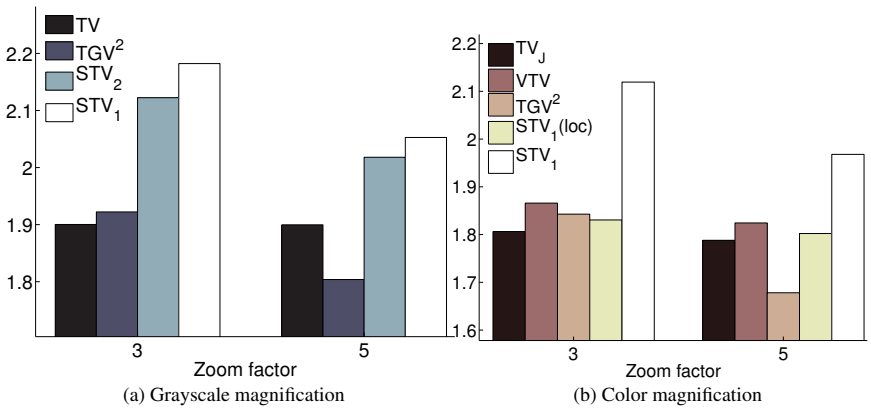
(a) Gaussian psf (b) uniform psf (c) motion psf

Figure 5. *Image deblurring comparisons among different regularizers for three PSFs and four noise levels. The performance is measured in terms of the average PSNR improvement (in dB) w.r.t. the degraded input, over all* 30 *images of Figure* 2*. The bars on the top row refer to grayscale results while on the bottom row to color results.*

(a) Input (PSNR=22.53) (b) TV (τ=4.32·10^{-3}, PSNR=27.85) (c) TGV^2 (τ=4.59·10^{-3}, PSNR=27.53) (d) STV$_2$ (τ=3.62·10^{-3}, PSNR=28.19) (e) STV$_1$ (τ=2.80·10^{-3}, PSNR=28.52)

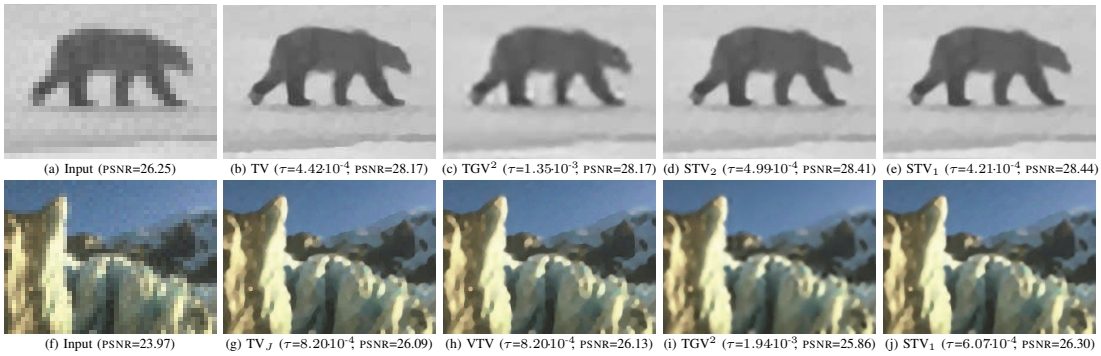
(f) Input (PSNR=24.77) (g) TV$_J$ (τ=4.29·10^{-3}, PSNR=28.94) (h) VTV (τ=3.37·10^{-3}, PSNR=29.25) (i) TGV^2 (τ=3.00·10^{-3}, PSNR=28.92) (j) STV$_1$ (τ=2.07·10^{-3}, PSNR=29.44)

Figure 6. *Image deblurring examples: close-ups of inputs and optimum results.* Top row*: grayscale deblurring of an input with motion blurring and noise level* $\sigma_w = 0.025$*. The ground truth corresponds to the grayscale version of the image in row* 3*, column* 4 *of Figure* 2*.* Bottom row*: color deblurring of an input with uniform blurring and noise level* $\sigma_w = 0.025$*. The ground truth corresponds to the image in row* 3*, column* 6 *of Figure* 2*. For each result, the individualized regularization parameter* τ *and corresponding optimum PSNR are reported.*



(a) Grayscale magnification (b) Color magnification

Figure 7. *Image magnification comparisons among different regularizers for two zoom factors. The performance is measured in terms of the average SNR improvement (in dB) w.r.t. the input, over all* 30 *images of Figure* 2.



(a) Input (PSNR=26.25) (b) TV (τ=4.42·10^{-4}, PSNR=28.17) (c) TGV^2 (τ=1.35·10^{-3}, PSNR=28.17) (d) STV$_2$ (τ=4.99·10^{-4}, PSNR=28.41) (e) STV$_1$ (τ=4.21·10^{-4}, PSNR=28.44)

(f) Input (PSNR=23.97) (g) TV$_J$ (τ=8.20·10^{-4}, PSNR=26.09) (h) VTV (τ=8.20·10^{-4}, PSNR=26.13) (i) TGV^2 (τ=1.94·10^{-3}, PSNR=25.86) (j) STV$_1$ (τ=6.07·10^{-4}, PSNR=26.30)

Figure 8. *Image magnification examples: close-ups of inputs and optimum results. Inputs are enlarged by simple zero order hold.* Top row*: grayscale magnification of an input with a zoom factor* $d = 5$*. The ground truth corresponds to the grayscale version of the image in row* 1*, column* 1 *of Figure* 2. Bottom row*: color magnification of an input with a zoom factor* $d = 5$*. The ground truth corresponds to the image in row* 1*, column* 3 *of Figure* 2*. For each result, the individualized regularization parameter* τ *and corresponding optimum PSNR are reported.*

5.4. Reconstruction from sparse Fourier measurements. In this section we consider the problem of image reconstruction from a limited number of Fourier measurements. In this setting, the forward operator of (2.1) corresponds to $\boldsymbol{A} = \boldsymbol{M}\boldsymbol{F}$, where $\boldsymbol{F}$ denotes the Fourier transform and $\boldsymbol{M}$ is a masking operator that retains only a subset of the available Fourier coefficients. The adopted forward model is closely related to the one of magnetic resonance imaging acquisition, where the k-space of the image is undersampled.

The reported experiments are conducted on the images shown in Figure 9. These are extracted slices from a 3D brain phantom that was obtained from the BrainWeb database (http://www.bic.mni.mcgill.ca/brainweb/). To create the measured data we use two different Fourier sampling patterns, namely, a radial mask with 32, 48, and 64 radial lines and a

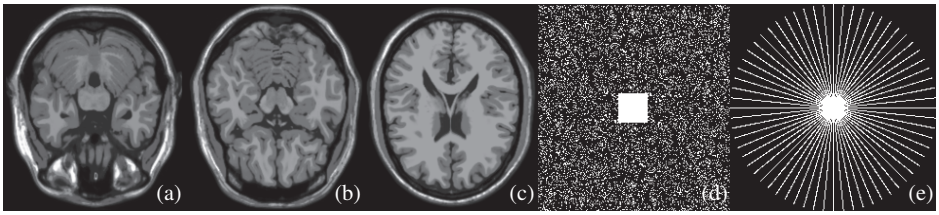


Figure 9. *Sparse Fourier image reconstruction setup.* (a)–(c) *Slices of a 3D brain phantom,* (d) *Poisson disk with* 20% *sampling density (white pixel values indicate sample locations),* (e) *radial sampling mask with* 32 *lines.*

Table 1

ISNR Comparisons on sparse Fourier image reconstruction for several sampling patterns and noise conditions.

	Sampling	Poisson disk – 20%				Poisson disk – 30%				Poisson disk – 40%				Radial – 32 lines				Radial – 48 lines				Radial – 64 lines			
	SNR	10dB	20dB	30dB	∞dB	10dB	20dB	30dB	∞dB	10dB	20dB	30dB	∞dB	10dB	20dB	30dB	∞dB	10dB	20dB	30dB	∞dB	10dB	20dB	30dB	∞dB
Slice 1	TV	2.14	1.92	2.65	2.86	3.27	2.89	3.97	4.62	4.44	3.90	5.79	7.43	3.00	4.15	5.42	5.72	3.30	3.99	6.27	7.23	3.95	3.34	5.91	7.97
	TGV^2	2.74	2.35	2.95	3.22	3.76	3.20	4.23	4.83	4.79	4.16	5.83	7.03	3.48	4.68	5.89	6.25	3.71	4.47	6.77	7.74	4.36	3.82	6.51	8.34
	STV_1	**2.99**	**2.68**	**3.29**	**3.73**	**4.14**	**3.74**	**4.74**	**5.67**	**5.34**	**4.84**	**6.65**	**8.82**	**4.10**	**5.43**	**6.90**	**7.50**	**4.48**	**5.55**	**8.21**	**10.04**	**5.15**	**5.01**	**8.27**	**11.88**
Slice 2	TV	2.22	1.84	2.51	2.67	3.33	2.66	3.77	4.28	4.28	3.29	5.23	6.43	2.58	3.48	5.02	4.99	3.07	3.25	5.22	6.16	3.83	2.66	5.03	6.94
	TGV^2	2.74	2.30	2.90	3.10	3.67	2.92	3.95	4.49	4.55	3.47	5.05	6.13	3.06	3.86	4.98	5.32	3.42	3.71	5.76	6.60	4.20	3.10	5.56	7.39
	STV_1	**2.99**	**2.52**	**3.23**	**3.53**	**4.07**	**3.35**	**4.40**	**5.31**	**5.05**	**4.08**	**5.76**	**7.66**	**3.60**	**4.62**	**6.12**	**6.64**	**4.18**	**4.71**	**7.19**	**9.06**	**4.97**	**4.29**	**7.35**	**10.88**
Slice 3	TV	2.48	2.46	3.48	3.91	3.75	3.63	5.38	6.59	4.90	4.47	6.83	9.65	3.45	5.01	6.66	7.20	3.67	3.91	6.51	8.11	4.47	3.49	6.20	9.00
	TGV^2	3.29	3.16	3.92	4.31	4.41	3.94	5.37	6.39	5.40	4.58	6.59	8.55	3.98	5.49	7.12	7.66	4.20	4.41	7.01	8.49	4.92	3.85	6.72	9.34
	STV_1	**3.43**	**3.38**	**4.26**	**4.92**	**4.71**	**4.55**	**6.18**	**7.87**	**5.88**	**5.44**	**7.74**	**11.34**	**4.56**	**6.38**	**8.31**	**9.45**	**4.86**	**5.49**	**8.56**	**11.60**	**5.65**	**5.19**	**8.61**	**13.84**

Poisson disk that retains 20%, 30%, and 40% of the Fourier coefficients. As an additional degradation factor we consider the presence of complex Gaussian noise in the Fourier domain of four different levels. These correspond to a signal-to-noise-ratio (SNR) of $\{10, 20, 30, \infty\}$ dB, respectively. Note that the last SNR value indicates the absence of noise in the measurements. In this case, for all regularizers under comparison, we do not fine tune the regularization parameter but instead we set it to $\tau = 10^{-6}$. The value of τ is chosen to be small so as to ensure that the reconstructed image in the sampled locations will retain the same frequency content as that of the measurements. Since the very small value of τ can affect the convergence speed of the minimization process, we adopt a simple "continuation" strategy that significantly speeds up the convergence. The idea is that we start with a large value for τ and then during the iterations of the minimization algorithm we gradually decrease it until we reach the target value. We note that for this problem instead of 100 we run 200 MFISTA iterations. The rest of the optimization set-up remains the same.

In Table 1 we report the reconstruction results we obtained using TV, TGV^2, and our proposed STV_1 regularizer. The quality of the reconstructions is measured in terms of increase in SNR w.r.t. the reconstructed image obtained by projecting the Fourier measurements back to the image domain. From these results we clearly observe that our proposed regularizer consistently outperforms TV and TGV^2 for all sampling strategies and noise levels. Further, we note that in the cases where the noise level is relatively low and the Fourier measurements are not very sparse, then the difference in SNR improvement is substantial. Indeed, on several occasions the SNR difference between STV_1 and the other two competing regularizers reaches

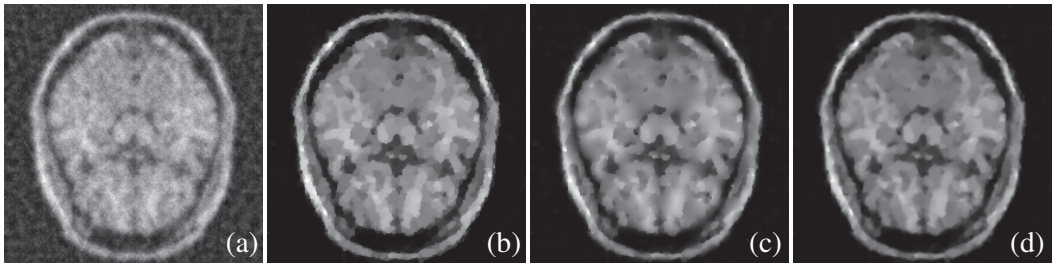


Figure 10. *Reconstruction of the brain image (Figure 9(b)) from Fourier data sampled with 32 radial lines and 10-dB SNR.* (a) *Back-projected image* ($PSNR = 23.30$), (b) *TV reconstruction* ($\tau = 3.84 \cdot 10^{-2}, PSNR = 25.88$), (c) *$TGV^2$ reconstruction* ($\tau = 4.01 \cdot 10^{-2}, PSNR = 26.36$), *and* (d) *$STV_1$ reconstruction* ($\tau = 2.62 \cdot 10^{-2}, PSNR = 26.90$).

more than 3 dB. In Figure 10 we show a representative reconstruction example for a radial sampling with 32 lines and a SNR of 10 dB. From this example, we can verify the superiority of our reconstruction over the ones of TV and TGV^2 both quantitatively and visually. Our STV_1 regularizer manages to better preserve the image edges than TV and at the same time avoid intensity smearing artifacts such as those appearing in the case of TGV^2 regularization.

5.5. Implementation details and computational runtimes. All methods were implemented in unoptimized MATLAB code that only makes use of the CPU.[4] Using the average runtime of TV as baseline, the relative average runtimes of the rest of the tested methods in the conducted grayscale experiments were as follows: 0.96 for TGV^2, 3.42 for STV_2, and 3.54 for STV_1. Regarding the color experiments and using the average runtime of VTV as baseline, the relative average runtimes of the tested methods were 1.11 for TV_J, 1.14 for TGV^2, 1.09 for STV_1(loc), and 4.66 for STV_1. Note that all runtimes were computed on a computer equipped with Intel Xeon Processor E5-1620 (Quad Core, 3.60-GHz Turbo, 10 MB).

We observe that TV, VTV, TGV^2, TV_J, and the proposed STV_1(loc) regularizers demonstrate a similar runtime performance. The proposed STV_2 and STV_1 functionals are around 3.5–4 times slower in the current implementation, which is due to the presence of a convolution kernel. We believe that this additional computational effort is definitely worth spending in several applications, given the significant improvement in the reconstructions that we observed in the reported experiments. Furthermore, it is worth mentioning that the runtimes of the proposed algorithm could be significantly improved by exploiting the fact that it is highly parallelizable and admits fast GPU implementations.

6. Conclusions. In this work we have introduced a novel family of convex energy functionals that extend and generalize the total variation seminorm and vectorial extensions of it. The key feature of our functionals is that they incorporate information from the vicinity of every point in the image domain by penalizing the eigenvalues of the structure tensor at this point. Therefore, they provide a richer and more robust measure of image variation which translates to improved reconstruction performance. This consistent behavior has been verified

[4]The source code for STV regularization is publicly available from the authors' websites.

by extensive comparisons we performed against competing methods on several inverse imaging problems.

Since our regularizers can be used instead of TV or its vectorial extensions in any energy functional, we believe that there is a broad spectrum of imaging applications where they can be proven very useful. Another interesting research direction is related to the open theoretical issue regarding the continuous version of the STV regularizers. In this article, we have defined STV by assuming that the underlying image lies in the Sobolev space $W^{1,2}\left(\mathbb{R}^2, \mathbb{R}^M\right)$. Since TV, which can be viewed as a special case of STV, is defined for images that belong to the bounded variation (BV) space [15], an open theoretical question is whether STV can also be defined for functions of a BV space or a space containing BV.

Appendix A. Mathematical Proofs.

A.1. Compact Operators. In this section we review basic definitions and results from the theory of compact operators. The concept of a compact operator will be essential for the proof of Theorem 2.1.

Definition A.1 (compact operators). *An operator $\boldsymbol{A} : \mathcal{H}_1 \to \mathcal{H}_2$, where $\mathcal{H}_1$ and $\mathcal{H}_2$ are two separable Hilbert spaces is* compact *if and only if it can be specified as*

$$f \mapsto \boldsymbol{A}f = \sum_{n=1}^{\infty} \sigma_n \, u_n \langle \tilde{u}_n, f \rangle_{\mathcal{H}_1} \tag{A.1}$$

for any $f \in \mathcal{H}_1$. In (A.1) $\tilde{u}_1, \tilde{u}_2, \ldots$ and $u_1, u_2, \ldots$ are orthonormal elements (functions) of $\mathcal{H}_1$ and $\mathcal{H}_2$, respectively, and $\sigma_1, \sigma_2, \ldots$ is a sequence of positive numbers with limit zero, called the singular values *of the operator. The same property applies to the adjoint operator $\boldsymbol{A}^* : \mathcal{H}_2 \to \mathcal{H}_1$ with the role of u_n and $\tilde{u}_n$ being interchanged.*

While all finite-dimensional linear operators (i.e., matrices) are compact, the property is more restrictive in infinite dimensions. However, the property is robust in the sense that it is conserved through subsequent linear transformation; i.e., if $\boldsymbol{A} : \mathcal{H}_1 \to \mathcal{H}_2$ is compact, then $\boldsymbol{B}\boldsymbol{A}$ is compact as well for any bounded operator $\boldsymbol{B} : \mathcal{H}_2 \to \mathcal{H}_1$ (not necessarily compact).

In order to stay closer to the standard (finite-dimensional) matrix-vector notation, we shall not distinguish between the operator $\boldsymbol{A}$ and its kernel, which is an element of the Hilbert space $\mathcal{H}_2 \otimes \mathcal{H}_1$, and simply write

$$\boldsymbol{A} = \sum_{n=1}^{\infty} \sigma_n(\boldsymbol{A}) \, u_n \otimes \tilde{u}_n \tag{A.2}$$

with the understanding that $u_n \in \mathcal{H}_2$ is in the "column" space of the operator, while $\tilde{u}_n$ is in the "row" space $\mathcal{H}_1$ and is acting as a linear functional (inner product) on the input.

Proposition A.2 (Hölder's inequality for the trace of operators). *Let $\boldsymbol{A}, \boldsymbol{B} : \mathcal{H}_1 \to \mathcal{H}_2$ be two compact operators. Then, for any $1 \leq p, q \leq +\infty$ with $\frac{1}{p} + \frac{1}{q} = 1$,*

$$|\mathrm{trace}(\boldsymbol{A}\boldsymbol{B}^*)| \leq \|\boldsymbol{\sigma}(\boldsymbol{A})\|_{\ell_p} \, \|\boldsymbol{\sigma}(\boldsymbol{B})\|_{\ell_q} \,, \tag{A.3}$$

where $\boldsymbol{\sigma}(\boldsymbol{A})$ and $\boldsymbol{\sigma}(\boldsymbol{B})$ are the vectors formed from the singular values of $\boldsymbol{A}$ and $\boldsymbol{B}$.

Proof. By direct calculation, we can express the trace$(\boldsymbol{A}\boldsymbol{B}^*)$ in terms of the singular values of $\boldsymbol{A}$ and $\boldsymbol{B}$ as

$$\text{trace}(\boldsymbol{A}\boldsymbol{B}^*) = \sum_{m=1}^{\infty}\sum_{n=1}^{\infty}\sigma_n(\boldsymbol{A})\sigma_m(\boldsymbol{B})\,\langle u_n, v_m\rangle_{\mathcal{H}_2}\langle \tilde{u}_n, \tilde{v}_m\rangle_{\mathcal{H}_1}, \tag{A.4}$$

where $u_n, v_n \in \mathcal{H}_2$ (resp., $\tilde{u}_n\,\tilde{v}_n \in \mathcal{H}_1$) are the left-sided (resp., right-sided) eigenfunctions associated with the singular decomposition of $\boldsymbol{A}$ and $\boldsymbol{B}$, respectively.

Using the Cauchy–Schwarz inequality, we first show that

$$\begin{aligned}\sum_{n=1}^{\infty}|\langle u_n, v_m\rangle_{\mathcal{H}_2}\,\langle \tilde{u}_n, \tilde{v}_m\rangle_{\mathcal{H}_1}| &\le \left(\sum_{n=1}^{\infty}|\langle u_n, v_m\rangle_{\mathcal{H}_2}|^2\right)^{1/2}\left(\sum_{n=1}^{\infty}|\langle \tilde{u}_n, \tilde{v}_m\rangle_{\mathcal{H}_1}|^2\right)^{1/2}\\ &\le \|v_m\|_{\mathcal{H}_2}\,\|\tilde{v}_m\|_{\mathcal{H}_1} = 1,\end{aligned} \tag{A.5}$$

where the latter norm estimate follows from Bessel's inequality based on the property that $\{u_n\}$ and $\{\tilde{u}_n\}$ are orthonormal sets of $\mathcal{H}_2$ and $\mathcal{H}_1$, respectively. There is also the dual counterpart of (A.5) where the summation is performed over m rather than n. Next, starting from the explicit trace formula (A.4), we apply the weighted Hölder inequality [20] for double sums, which gives

$$\begin{aligned}&|\text{trace}(\boldsymbol{A}\boldsymbol{B}^*)|\\ &\le \sum_{m=1}^{\infty}\sum_{n=1}^{\infty}\sigma_n(\boldsymbol{A})\sigma_m(\boldsymbol{B})\;|\langle u_n, v_m\rangle_{\mathcal{H}_2}\,\langle \tilde{u}_n, \tilde{v}_m\rangle_{\mathcal{H}_1}|\\ &\le \left(\sum_{n=1}^{\infty}\sum_{m=1}^{\infty}\sigma_n^p(\boldsymbol{A})\,|\langle u_n, v_m\rangle_{\mathcal{H}_2}\,\langle \tilde{u}_n, \tilde{v}_m\rangle_{\mathcal{H}_1}|\right)^{1/p}\left(\sum_{n=1}^{\infty}\sum_{m=1}^{\infty}\sigma_m^q(\boldsymbol{B})\,|\langle u_n, v_m\rangle_{\mathcal{H}_2}\,\langle \tilde{u}_n, \tilde{v}_m\rangle_{\mathcal{H}_1}|\right)^{1/q}\\ &= \left(\sum_{n=1}^{\infty}\sigma_n^p(\boldsymbol{A})\sum_{m=1}^{\infty}|\langle u_n, v_m\rangle_{\mathcal{H}_2}\,\langle \tilde{u}_n, \tilde{v}_m\rangle_{\mathcal{H}_1}|\right)^{1/p}\left(\sum_{m=1}^{\infty}\sigma_m^q(\boldsymbol{B})\sum_{n=1}^{\infty}|\langle u_n, v_m\rangle_{\mathcal{H}_2}\,\langle \tilde{u}_n, \tilde{v}_m\rangle_{\mathcal{H}_1}|\right)^{1/q}\\ &\le \left(\sum_{n=1}^{\infty}\sigma_n^p(\boldsymbol{A})\,\|u_n\|_{\mathcal{H}_2}\|\tilde{u}_n\|_{\mathcal{H}_1}\right)^{1/p}\left(\sum_{m=1}^{\infty}\sigma_m^q(\boldsymbol{B})\,\|v_m\|_{\mathcal{H}_2}\|\tilde{v}_m\|_{\mathcal{H}_1}\right)^{1/q},\end{aligned}$$

where the availability of (A.5) is essential for reducing the double sum to a single one. ∎

A.2. Proof of Theorem 2.1. While in (2.10) we define the STV functional for a 2D image; next we provide a proof of Theorem 2.1 for the more general case where $\boldsymbol{u}$ is defined on $\mathbb{R}^d$. To show the translation and rotation invariance property of the STV functional, we consider the transformation of the image coordinates $T(\boldsymbol{x}) = \boldsymbol{R}\boldsymbol{x} + \boldsymbol{\alpha}$, where $\boldsymbol{R}$ is a rotation matrix and $\boldsymbol{\alpha}$ represents a translation. Applying the chain rule to the Jacobian matrix, we get

$$J\{\boldsymbol{u}\circ T\}(\boldsymbol{x}) = J\boldsymbol{u}(T(\boldsymbol{x}))\,\boldsymbol{R}, \tag{A.6}$$

where $\circ$ indicates the composition operation. Now, if we denote $\boldsymbol{S} = S_K\{\boldsymbol{u}\circ T\}(\boldsymbol{x})$ we have

$$\begin{aligned}
\boldsymbol{S} &= \boldsymbol{R}^T\left(\int_{\mathbb{R}^d} K(\boldsymbol{y})\, J\mathbf{u}^T(T(\boldsymbol{x}-\boldsymbol{y}))\, J\mathbf{u}(T(\boldsymbol{x}-\boldsymbol{y}))\,\mathrm{d}\boldsymbol{y}\right)\boldsymbol{R}\\
&= \boldsymbol{R}^T\left(\int_{\mathbb{R}^d} K(\boldsymbol{y})\, J\mathbf{u}^T(T(\boldsymbol{x})-\boldsymbol{R}\boldsymbol{y})\, J\mathbf{u}(T(\boldsymbol{x})-\boldsymbol{R}\boldsymbol{y})\,\mathrm{d}\boldsymbol{y}\right)\boldsymbol{R}\\
&\overset{\boldsymbol{y}'=\boldsymbol{R}\boldsymbol{y}}{=} \boldsymbol{R}^T\left(\int_{\mathbb{R}^d} K(\boldsymbol{R}^{-1}\boldsymbol{y}')\, J\mathbf{u}^T(T(\boldsymbol{x})-\boldsymbol{y}')\, J\mathbf{u}(T(\boldsymbol{x})-\boldsymbol{y}')\,\mathrm{d}\boldsymbol{y}'\right)\boldsymbol{R}\\
&= \boldsymbol{R}^T\left(\int_{\mathbb{R}^d} K(\boldsymbol{y}')\, J\mathbf{u}^T(T(\boldsymbol{x})-\boldsymbol{y}')\, J\mathbf{u}(T(\boldsymbol{x})-\boldsymbol{y}')\,\mathrm{d}\boldsymbol{y}'\right)\boldsymbol{R}\\
&= \boldsymbol{R}^T S_K\{\boldsymbol{u}\}(T(\boldsymbol{x}))\,\boldsymbol{R}. \qquad \text{(A.7)}
\end{aligned}$$

We note that in order to derive the above result we have considered the convolution kernel K to be isotropic, i.e., $K(\boldsymbol{x}) = K(|\boldsymbol{x}|)$. From (A.7) and since the eigenvalues of the structure tensor in the r.h.s. are not affected by the presence of the rotation matrix $\boldsymbol{R}$ and its inverse $\boldsymbol{R}^T$, we have that $\lambda^{\pm}(\boldsymbol{x}) = \lambda^{\pm}(T(\boldsymbol{x}))$. Using this result and having in mind that in order to evaluate the STV_p functional in (2.10) we need to integrate over the whole image domain, it is now clear that the energy will be preserved despite the transformation of the coordinate system. In other words, it holds that $\mathrm{STV}_p(\boldsymbol{u}\circ T) = \mathrm{STV}_p(\boldsymbol{u})$.

Next, in order to prove the convexity and 1-homogeneity properties, let us first introduce the $d\times d$ matrix $\mathbf{G}_K(\boldsymbol{u})$ defined as

$$\mathbf{G}_K(\boldsymbol{u}) = \int_{\mathbb{R}^d} K(\boldsymbol{y})\, J\boldsymbol{u}^T(\boldsymbol{y})\, J\boldsymbol{u}(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y}, \qquad \text{(A.8)}$$

which corresponds to evaluating the structure tensor (2.7) at the origin of $\boldsymbol{u}$. This matrix is well-defined over the functions that belong to the Sobolev space $W^{1,2}(\mathbb{R}^d,\mathbb{R}^M)$. Next, we show the following result.

Theorem A.3. *Let λ_n denote the nth eigenvalue of the matrix $\boldsymbol{G}_K(\boldsymbol{u})$. Then $\rho(\boldsymbol{u}) = (\sum_{n=1}^{d}\lambda_n^{\frac{p}{2}})^{1/p}$ is a 1-homogenous and convex function of $\boldsymbol{u}$ in the sense that*

$$\rho(a\boldsymbol{u}) = |a|\,\rho(\boldsymbol{u}), \qquad \text{(A.9)}$$

$$\rho(t\boldsymbol{u}_1 + (1-t)\boldsymbol{u}_2) \le t\rho(\boldsymbol{u}_1) + (1-t)\rho(\boldsymbol{u}_2) \qquad \text{(A.10)}$$

$\forall a\in\mathbb{R}$, $t\in[0,1]$, and $\boldsymbol{u},\boldsymbol{u}_1,\boldsymbol{u}_2\in W^{1,2}(\mathbb{R}^d,\mathbb{R}^M)$.

Proof. To prove convexity of $\rho(\boldsymbol{u})$, let us first introduce the $(\infty\times d)$-dimensional matrix defined as $\boldsymbol{A} = \boldsymbol{A}(\boldsymbol{u},\boldsymbol{y}) = [\boldsymbol{f}_1(\boldsymbol{y})\ldots\boldsymbol{f}_d(\boldsymbol{y})]$ for arbitrary $\boldsymbol{y}\in\mathbb{R}^d$, where

$$\boldsymbol{f}_n(\boldsymbol{y}) = \Big(\underbrace{k(\boldsymbol{y})\,u^1_{x_n}(\boldsymbol{y})}_{\boldsymbol{f}_{n,1}(\boldsymbol{y})},\ldots,\underbrace{k(\boldsymbol{y})\,u^M_{x_n}(\boldsymbol{y})}_{\boldsymbol{f}_{n,M}(\boldsymbol{y})}\Big),\ 1\le n\le d,$$

is a vector function $\boldsymbol{f}_n\in\mathcal{H} = L_2(\mathbb{R}^d,\mathbb{R}^M)$, $u^m_{x_n}\ \forall\, 1\le m\le M$ denotes the derivative of the mth component of $\boldsymbol{u}$ w.r.t. the nth dimension, and $k(\boldsymbol{y}) = \sqrt{K(\boldsymbol{y})}$. Note that the last

assignment is valid since the convolution kernel K is assumed to be nonnegative valued. Since the functions $\boldsymbol{f}_n = \boldsymbol{f}_n(\boldsymbol{y})$ belong to the Hilbert space $\mathcal{H}$, the matrix $\boldsymbol{A}$ defines a linear map: $\mathbb{R}^d \to \mathcal{H}$. The Hilbert space $\mathcal{H}$ is associated with the inner product

$$\langle \boldsymbol{f}_n \, , \, \boldsymbol{g}_n \rangle_{\mathcal{H}} = \sum_{m=1}^{M} \int_{\mathbb{R}^d} \boldsymbol{f}_{n,m}(\boldsymbol{y}) \boldsymbol{g}_{n,m}(\boldsymbol{y}) d\boldsymbol{y}.$$

The adjoint of $\boldsymbol{A}$ is the unique operator $\boldsymbol{A}^* : \mathcal{H} \to \mathbb{R}^d$ that satisfies

$$\langle \boldsymbol{f} \, , \, \boldsymbol{A}\boldsymbol{v} \rangle_{\mathcal{H}} = \langle \boldsymbol{A}^* \boldsymbol{f} \, , \, \boldsymbol{v} \rangle \tag{A.11}$$

for every $\boldsymbol{v} \in \mathbb{R}^d$ and every $f \in \mathcal{H}$. Having introduced the adjoint operator of $\boldsymbol{A}$, we are in the position of specifying the Gram matrix of size $d \times d$:

$$\begin{aligned}
\boldsymbol{A}^*\boldsymbol{A} &= \begin{pmatrix} \langle \boldsymbol{f}_1 \, , \, \boldsymbol{f}_1 \rangle_{\mathcal{H}} & \dots & \langle \boldsymbol{f}_1 \, , \, \boldsymbol{f}_d \rangle_{\mathcal{H}} \\ \vdots & \ddots & \vdots \\ \langle \boldsymbol{f}_d \, , \, \boldsymbol{f}_1 \rangle_{\mathcal{H}} & \dots & \langle \boldsymbol{f}_d \, , \, \boldsymbol{f}_d \rangle_{\mathcal{H}} \end{pmatrix} \\
&= \begin{pmatrix} \sum_{m=1}^{M} \int_{\mathbb{R}^d} k^2(\boldsymbol{y}) \left(u_{x_1}^m(\boldsymbol{y})\right)^2 d\boldsymbol{y} & \dots & \sum_{m=1}^{M} \int_{\mathbb{R}^d} k^2(\boldsymbol{y}) \, u_{x_1}^m(\boldsymbol{y}) \, u_{x_d}^m(\boldsymbol{y}) \, d\boldsymbol{y} \\ \vdots & \ddots & \vdots \\ \sum_{m=1}^{M} \int_{\mathbb{R}^d} k^2(\boldsymbol{y}) \, u_{x_d}^m(\boldsymbol{y}) \, u_{x_1}^m(\boldsymbol{y}) \, d\boldsymbol{y} & \dots & \sum_{m=1}^{M} \int_{\mathbb{R}^d} k^2(\boldsymbol{y}) \left(u_{x_d}^m(\boldsymbol{y})\right)^2 d\boldsymbol{y} \end{pmatrix} \\
&= \begin{pmatrix} \int_{\mathbb{R}^d} K(\boldsymbol{y}) \sum_{m=1}^{M} \left(u_{x_1}^m(\boldsymbol{y})\right)^2 d\boldsymbol{y} & \dots & \int_{\mathbb{R}^d} K(\boldsymbol{y}) \sum_{m=1}^{M} u_{x_1}^m(\boldsymbol{y}) \, u_{x_d}^m(\boldsymbol{y}) \, d\boldsymbol{y} \\ \vdots & \ddots & \vdots \\ \int_{\mathbb{R}^d} K(\boldsymbol{y}) \sum_{m=1}^{M} u_{x_d}^m(\boldsymbol{y}) \, u_{x_1}^m(\boldsymbol{y}) \, d\boldsymbol{y} & \dots & \int_{\mathbb{R}^d} K(\boldsymbol{y}) \sum_{m=1}^{M} \left(u_{x_d}^m(\boldsymbol{y})\right)^2 d\boldsymbol{y} \end{pmatrix} \\
&= \boldsymbol{G}_K(\boldsymbol{u}) \, .
\end{aligned} \tag{A.12}$$

From (A.12) it turns out that the $d \times d$ Gram matrix, which provides a means to compute the spectrum of $\boldsymbol{A}$, is equal to the $\boldsymbol{G}_K(\boldsymbol{u})$ matrix defined in (A.8). This relation further implies that the vector consisting of the singular values of the infinite-dimensional but finite rank matrix (or compact operator) $\boldsymbol{A}$, which we will denote as $\boldsymbol{\sigma}(\boldsymbol{A})$, is directly related to the eigenvalues of $\boldsymbol{G}_K(\boldsymbol{u})$. Indeed, we have

$$\boldsymbol{\sigma}(\boldsymbol{A}) = \sqrt{\boldsymbol{\lambda}(\boldsymbol{A}^*\boldsymbol{A})} = \sqrt{\boldsymbol{\lambda}(\boldsymbol{G}_K(\boldsymbol{u}))}, \tag{A.13}$$

where $\boldsymbol{\lambda}(\boldsymbol{A}^*\boldsymbol{A})$ is the vector with the eigenvalues of the matrix $\boldsymbol{A}^*\boldsymbol{A}$. Thus, we conclude that the nth singular value of $\boldsymbol{A}$ can be derived as $\sigma_n = \sqrt{\lambda_n}$ with $1 \leq n \leq d$. This means that we can equivalently write the function $\rho(\boldsymbol{u})$ as

$$\rho(\boldsymbol{u}) = \left(\sum_{n=1}^{d} \sigma_n^p \right)^{1/p} = \|\boldsymbol{\sigma}(\boldsymbol{A})\|_{\ell_p} \, . \tag{A.14}$$

Next, we shall show that for two compact operators $\boldsymbol{A}, \boldsymbol{B} : \mathcal{H}_1 \mapsto \mathcal{H}_2$ and for any $1 \leq p, q \leq +\infty$ with $\frac{1}{p} + \frac{1}{q} = 1$, it holds

$$\|\boldsymbol{\sigma}(\boldsymbol{A})\|_{\ell_p} = \sup_{\|\boldsymbol{\sigma}(\boldsymbol{B})\|_{\ell_q}=1} \operatorname{trace}(\boldsymbol{A}\boldsymbol{B}^*). \tag{A.15}$$

We start by invoking Hölder's inequality for the trace of operators, whose proof is provided in Proposition A.2. According to this inequality we have that

$$|\operatorname{trace}(\boldsymbol{A}\boldsymbol{B}^*)| \leq \|\boldsymbol{\sigma}(\boldsymbol{A})\|_{\ell_p} \|\boldsymbol{\sigma}(\boldsymbol{B})\|_{\ell_q}. \tag{A.16}$$

This, in turn, implies that

$$\sup_{\|\boldsymbol{\sigma}(\boldsymbol{B})\|_{\ell_q}=1} \operatorname{trace}(\boldsymbol{A}\boldsymbol{B}^*) \leq \|\boldsymbol{\sigma}(\boldsymbol{A})\|_{\ell_p}. \tag{A.17}$$

Now, the critical point is to show that for every operator $\boldsymbol{A}$ we can always select another operator $\boldsymbol{B}$ with $\|\boldsymbol{\sigma}(\boldsymbol{B})\|_{\ell_q} = 1$ such that $\operatorname{trace}(\boldsymbol{A}\boldsymbol{B}^*) = \|\boldsymbol{\sigma}(\boldsymbol{A})\|_{\ell_p}$. To show this, we choose

$$\boldsymbol{B} = \sum_{n=1}^{\infty} \sigma_n(\boldsymbol{B}) u_n \otimes \tilde{u}_n \tag{A.18}$$

with

$$\sigma_n(\boldsymbol{B}) = \frac{\sigma_n^{p-1}(\boldsymbol{A})}{\|\boldsymbol{\sigma}(\boldsymbol{A})\|_{\ell_p}^{p-1}}. \tag{A.19}$$

Indeed, since $\boldsymbol{B}$ has the same eigenvectors as $\boldsymbol{A}$, we have that

$$\operatorname{trace}(\boldsymbol{A}\boldsymbol{B}^*) = \frac{\sum_{n=1}^{\infty} \sigma_n^p(\boldsymbol{A})}{\|\boldsymbol{\sigma}(\boldsymbol{A})\|_{\ell_p}^{p-1}} = \|\boldsymbol{\sigma}(\boldsymbol{A})\|_{\ell_p}. \tag{A.20}$$

Moreover, by noting that $p = q(p-1)$, we easily verify that

$$\|\boldsymbol{\sigma}(\boldsymbol{B})\|_{\ell_q}^q = \sum_{n=1}^{\infty} \sigma_n^q(\boldsymbol{B}) = \frac{\sum_{n=1}^{\infty} \sigma_n^{q(p-1)}(\boldsymbol{A})}{\|\boldsymbol{\sigma}(\boldsymbol{A})\|_{\ell_p}^{q(p-1)}} = 1. \tag{A.21}$$

This shows that the supremum in (A.17) achieves the equality, which proves (A.15).

Thanks to the dual definition in (A.15), we can now show that

$$\begin{aligned}
\|\boldsymbol{\sigma}(t\boldsymbol{A}_1 + (1-t)\boldsymbol{A}_2)\|_{\ell_p} &= \sup_{\|\boldsymbol{\sigma}(\boldsymbol{B})\|_{\ell_q}=1} \operatorname{trace}((t\boldsymbol{A}_1 + (1-t)\boldsymbol{A}_2)\boldsymbol{B}^*) \\
&\leq \sup_{\|\boldsymbol{\sigma}(\boldsymbol{B}_1)\|_{\ell_q}=1} \operatorname{trace}(t\boldsymbol{A}_1\boldsymbol{B}_1^*) + \sup_{\|\boldsymbol{\sigma}(\boldsymbol{B}_2)\|_{\ell_q}=1} \operatorname{trace}((1-t)\boldsymbol{A}_2\boldsymbol{B}_2^*) \\
&\leq t \sup_{\|\boldsymbol{\sigma}(\boldsymbol{B}_1)\|_{\ell_q}=1} \operatorname{trace}(\boldsymbol{A}_1\boldsymbol{B}_1^*) + (1-t) \sup_{\|\boldsymbol{\sigma}(\boldsymbol{B}_2)\|_{\ell_q}=1} \operatorname{trace}(\boldsymbol{A}_2\boldsymbol{B}_2^*) \\
&= t\|\boldsymbol{\sigma}(\boldsymbol{A}_1)\|_{\ell_p} + (1-t)\|\boldsymbol{\sigma}(\boldsymbol{A}_2)\|_{\ell_p}.
\end{aligned} \tag{A.22}$$

Next, we recall that the kernel of $\boldsymbol{A}$ is linear in $\boldsymbol{u}$, since it only involves derivatives which are linear operations and a multiplication with a constant kernel. This together with (A.22) proves that the function $\rho(\boldsymbol{u})$ is a convex function of $\boldsymbol{u}$.

Finally, the 1-homogeneity of $\rho(\boldsymbol{u})$ is easily verified by using the definition of ρ, given in (A.14). ∎

Using the results of the above theorem, the convexity and 1-homogeneity properties of the STV_p functional can now be easily verified by first noting that we can write it as

$$\mathrm{STV}_p(\boldsymbol{u}) = \int_{\mathbb{R}^d} \rho(\boldsymbol{u} \circ P_{\boldsymbol{x}})\, \mathrm{d}\boldsymbol{x}, \tag{A.23}$$

where $P_{\boldsymbol{x}}(\boldsymbol{y}) = \boldsymbol{x} - \boldsymbol{y}$ and $\rho(\boldsymbol{u})$ is defined in Theorem A.3. If we apply the chain rule to the Jacobian matrix we get

$$J\{\boldsymbol{u} \circ P_{\boldsymbol{x}}\}(\boldsymbol{y}) = -J\boldsymbol{u}(P_{\boldsymbol{x}}(\boldsymbol{y})). \tag{A.24}$$

Now, for the matrix $\boldsymbol{G}_K(\boldsymbol{u} \circ P_{\boldsymbol{x}})$, which is the main component of the function $\rho(\boldsymbol{u} \circ P_{\boldsymbol{x}})$, we have

$$\begin{aligned} \boldsymbol{G}_K(\boldsymbol{u} \circ P_{\boldsymbol{x}}) &= \int_{\mathbb{R}^d} K(\boldsymbol{y})\, \nabla \mathbf{u}^T(P_{\boldsymbol{x}}(\boldsymbol{y}))\, \nabla \mathbf{u}(P_{\boldsymbol{x}}(\boldsymbol{y}))\, \mathrm{d}\boldsymbol{y} \\ &\overset{\boldsymbol{y}'=P_{\boldsymbol{x}}(\boldsymbol{y})}{=} \int_{\mathbb{R}^d} K(\boldsymbol{x} - \boldsymbol{y}')\, \nabla \mathbf{u}^T(\boldsymbol{y}')\, \nabla \mathbf{u}(\boldsymbol{y}')\, \mathrm{d}\boldsymbol{y}' \\ &= \int_{\mathbb{R}^d} K_{\boldsymbol{x}}(\boldsymbol{y})\, \nabla \mathbf{u}^T(\boldsymbol{y})\, \nabla \mathbf{u}(\boldsymbol{y})\, \mathrm{d}\boldsymbol{y} = \boldsymbol{G}_{K_{\boldsymbol{x}}}(\boldsymbol{u}), \end{aligned} \tag{A.25}$$

where $K_{\boldsymbol{x}}(\boldsymbol{y}) = K(\boldsymbol{x} - \boldsymbol{y}) = K(\boldsymbol{y} - \boldsymbol{x})$. The above result implies that $\rho(\boldsymbol{u} \circ P_{\boldsymbol{x}}) = \rho_{\boldsymbol{x}}(\boldsymbol{u})$, where $\rho_{\boldsymbol{x}}(\boldsymbol{u})$ is defined similarly to $\rho(\boldsymbol{u})$, with the only difference being that its input eigenvalues correspond to those of the matrix $\boldsymbol{G}_{K_{\boldsymbol{x}}}(\boldsymbol{u})$ instead of $\boldsymbol{G}_K(\boldsymbol{u})$. Therefore, it holds that

$$\mathrm{STV}_p(\boldsymbol{u}) = \int_{\mathbb{R}^d} \rho_{\boldsymbol{x}}(\boldsymbol{u})\, \mathrm{d}\boldsymbol{x}. \tag{A.26}$$

Since the kernel $K_{\boldsymbol{x}}$, which is used in $\rho_{\boldsymbol{x}}(\boldsymbol{u})$, is just a shifted version of the initial convolution kernel K and does not depend on $\boldsymbol{u}$, the convexity and homogeneity results of $\rho(\boldsymbol{u})$ are also inherited by $\rho_{\boldsymbol{x}}(\boldsymbol{u})$ $\forall \boldsymbol{x} \in \mathbb{R}^d$. Therefore, since the STV_p functional corresponds to the integral of convex and 1-homogeneous functionals of $\boldsymbol{u}$, it is also a convex and 1-homogeneous function of $\boldsymbol{u}$.

A.3. Proof of Proposition 3.2. To find the adjoint of the discrete patch-based Jacobian, we exploit the relation of the inner products of the spaces $\mathbb{R}^{NM}$ and $\mathcal{X}$ in (3.8). Using the definition of the inner product in $\mathcal{X}$ (3.6), we can equivalently write (3.8) as

$$\sum_{n=1}^{N} \mathrm{trace}\left([\boldsymbol{J}_K \boldsymbol{u}]_n^T\, \boldsymbol{Y}_n\right) = \sum_{n=1}^{N} \sum_{m=1}^{M} [\boldsymbol{u}_m]_n\, [\boldsymbol{J}_K^* \boldsymbol{Y}]_k, \tag{A.27}$$

where $k = (N-1)\,m + n$. Then, we expand the l.h.s. of (A.27), to obtain

$$\begin{aligned}
&\sum_{n,m,l} \left[T_{\boldsymbol{s}_l,\boldsymbol{\omega}} \circ \boldsymbol{D}_h \boldsymbol{u}_m\right]_n \boldsymbol{Y}_n^{(r,1)} + \left[T_{\boldsymbol{s}_l,\boldsymbol{\omega}} \circ \boldsymbol{D}_v \boldsymbol{u}_m\right]_n \boldsymbol{Y}_n^{(r,2)} \\
&\quad = \sum_{n,m,l} \left[\boldsymbol{u}_m\right]_n \left(\left[\boldsymbol{D}_h^* \circ T_{\boldsymbol{s}_l,\boldsymbol{\omega}}^* \circ \boldsymbol{Y}^{(r,1)}\right]_n + \left[\boldsymbol{D}_v^* \circ T_{\boldsymbol{s}_l,\boldsymbol{\omega}}^* \circ \boldsymbol{Y}^{(r,2)}\right]_n \right) \\
&\quad = \sum_{n,m} \left[\boldsymbol{u}_m\right]_n \left(\sum_{l=1}^{L} \left[-\mathbf{div}\left(T_{\boldsymbol{s}_l,\boldsymbol{\omega}}^* \circ \boldsymbol{Y}^{((m-1)L+l,:)}\right)\right]_n \right), \qquad \text{(A.28)}
\end{aligned}$$

where $r = (m-1)\,L + l$ and $\boldsymbol{D}_h$, $\boldsymbol{D}_v$ are the horizontal and vertical components of the discrete gradient, respectively. Also note that $\boldsymbol{Y}_n^{(i,j)}$ with $1 \le n \le N$, $1 \le i \le LM$, and $1 \le j \le 2$, corresponds to a single element of $\boldsymbol{Y} \in \mathcal{X}$, while $\boldsymbol{Y}_n^{(i,:)}$ is a vector extracted from the ith row of the nth matrix component, $\boldsymbol{Y}_n \in \mathbb{R}^{(LM)\times 2}$, of $\boldsymbol{Y}$. Now, by comparing the r.h.s. of (A.27) to the r.h.s. expansion of (A.28), we can verify that the adjoint of the patch-based Jacobian is indeed computed by the formula provided in Proposition 3.2.

A.4. Proof of Proposition 4.2. For any pair of variables $\boldsymbol{\Omega}, \boldsymbol{\Psi} \in \mathcal{X}$ and

$$d\left(\boldsymbol{\Omega}\right) = \frac{1}{2}\left\|\boldsymbol{w} - \Pi_C\left(\boldsymbol{w}\right)\right\|_2^2 + \frac{1}{2}\left(\left\|\mathbf{z}\right\|_2^2 - \left\|\boldsymbol{w}\right\|_2^2\right),$$

we can show that

$$\begin{aligned}
\left\|\nabla d\left(\boldsymbol{\Omega}\right) - \nabla d\left(\boldsymbol{\Psi}\right)\right\|_{\mathcal{X}} &= \left\|\tau \boldsymbol{J}_K \left(h\left(\boldsymbol{\Omega}\right) - h\left(\boldsymbol{\Psi}\right)\right)\right\|_{\mathcal{X}} \\
&\le \tau \left\|\boldsymbol{J}_K\right\| \left\|h\left(\boldsymbol{\Omega}\right) - h\left(\boldsymbol{\Psi}\right)\right\|_2 \le \tau \left\|\boldsymbol{J}_K\right\| \left\|\tau \boldsymbol{J}_K^* \left(\boldsymbol{\Omega} - \boldsymbol{\Psi}\right)\right\|_2 \\
&\le \tau^2 \left\|\boldsymbol{J}_K\right\|^2 \left\|\boldsymbol{\Omega} - \boldsymbol{\Psi}\right\|_{\mathcal{X}}, \qquad \text{(A.29)}
\end{aligned}$$

where $h\left(\boldsymbol{\Omega}\right) = \Pi_{\mathcal{C}}\left(\boldsymbol{z} - \tau \boldsymbol{J}_K^* \boldsymbol{\Omega}\right)$. This result follows from the relation between the norms defined in the spaces $\mathcal{X}$ and $\mathbb{R}^{NM}$ and the induced operator norm, i.e., $\left\|\boldsymbol{J}_K \boldsymbol{u}\right\|_{\mathcal{X}} \le \left\|\boldsymbol{J}_K\right\| \left\|\boldsymbol{u}\right\|_2$, and the fact that the projection operator onto a convex set $\mathcal{C} \subseteq \mathbb{R}^N$ is firmly nonexpansive [2, Proposition 4.8]. The latter translates to $\left\|\Pi_{\mathcal{C}}\left(\boldsymbol{u}\right) - \Pi_{\mathcal{C}}\left(\boldsymbol{v}\right)\right\|_2 \le \left\|\boldsymbol{u} - \boldsymbol{v}\right\|_2 \; \forall\, \boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^N$.

Now, in order to find an upper bound of the operator norm of the patch-based Jacobian we use that $\left\|\boldsymbol{J}_K\right\|^2 = \left\|\boldsymbol{J}_K^* \boldsymbol{J}_K\right\|$ [34] (general property of bounded linear operators) and denote $\boldsymbol{D} = -\mathbf{div} \circ \sum_l (T_{\boldsymbol{s}_l,\boldsymbol{\omega}}^* \circ T_{\boldsymbol{s}_l,\boldsymbol{\omega}}) \circ \nabla$. Then, we get

$$\left\|\boldsymbol{J}_K^* \boldsymbol{J}_K \boldsymbol{u}\right\|_2 = \left\|\boldsymbol{D}\boldsymbol{u}\right\|_2 \le \left\|\nabla\right\|^2 \left\|T\right\|^2 \left\|\boldsymbol{u}\right\|_2, \qquad \text{(A.30)}$$

where $T \triangleq \sum_{l=1}^{L} (T_{\boldsymbol{s}_l,\boldsymbol{\omega}}^* \circ T_{\boldsymbol{s}_l,\boldsymbol{\omega}})$. From [3] we have that $\left\|\nabla\right\|^2 \le 8$ and we can further show that $\left\|T\right\|^2 \le \sqrt{2}$, if we assume a unit-normalized kernel K and a shift operation with reflexive boundaries. These immediately imply that an upper bound of the Lipschitz constant of $\nabla d\left(\boldsymbol{\Omega}\right)$ will be $L\left(d\right) \le \tau^2 \left\|\boldsymbol{J}_K\right\|^2 \le 8\sqrt{2}\,\tau^2$.

Acknowledgment. The authors would like to thank the anonymous reviewers and the associate editor for their useful comments and suggestions.

REFERENCES

[1] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T Figueiredo, *Fast image recovery using variable splitting and constrained optimization*, IEEE Trans. Image Process., 19 (2010), pp. 2345–2356.

[2] H. H. Bauschke and P. L. Combettes, *Convexity and nonexpansiveness*, in Convex Analysis and Monotone Operator Theory in Hilbert Spaces, J. M. Borwein and P. Borwein, eds., Springer, New York, 2011, pp. 59–74.

[3] A. Beck and M. Teboulle, *Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems*, IEEE Trans. Image Process., 18 (2009), pp. 2419–2434.

[4] A. Beck and M. Teboulle, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM J. Imaging Sci., 2 (2009), pp. 183–202.

[5] B. Benjamin, M. Burger, M. Droske, O. Nemitz, and M. Rumpf, *Cartoon extraction based on anisotropic image classification vision, modeling, and visualization*, in Vision, Modeling, and Visualization, Akademische Verlagsgesellschaft, Berlin, 2006, pp. 293–300.

[6] M. Bertero and P. Boccacci, *Introduction to Inverse Problems in Imaging*, IOP, Bristol, UK, 1998.

[7] P. Blomgren and T. F. Chan, *Color TV: Total variation methods for restoration of vector-valued images*, IEEE Trans. Image Process., 7 (1998), pp. 304–309.

[8] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, *Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers*, Now Publishers, Hanover, MA, 2011.

[9] K. Bredies, *Recovering Piecewise Smooth Multichannel Images by Minimization of Convex Functionals with Total Generalized Variation Penalty*, A. Bruhn et al. eds., in Global Optimization Methods, Lecture Notes in Comput. Sci. 8293, 2014, pp. 44–77.

[10] K. Bredies, K. Kunisch, and T. Pock, *Total generalized variation*, SIAM J. Imaging Sci., 3 (2010), pp. 492–526.

[11] X. Bresson and T. F. Chan, *Fast dual minimization of the vectorial total variation norm and applications to color image processing*, Inverse Probl. Imaging, 2 (2008), pp. 455–484.

[12] A. Chambolle, *An algorithm for total variation minimization and applications*, J. Math. Imaging Vision, 20 (2004), pp. 89–97.

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

[14] A. Chambolle and T. Pock, *A first-order primal-dual algorithm for convex problems with applications to imaging*, J. Math. Imaging Vision, 40 (2011), pp. 120–145.

[15] T. Chan, S. Esedoglu, F. Park, and A. Yip, *Total variation image restoration: Overview and recent developments*, Handbook of Mathematical Models in Computer Vision, Springer, New York, 2005, pp. 17–31.

[16] T. Chan, A. Marquina, and P. Mulet, *High-order total variation-based image restoration*, SIAM J. Sci. Comput., 22 (2000), pp. 503–516.

[17] Q. Cheng, H. Shen, L. Zhang, and P. Li, *Inpainting for remotely sensed images with a multichannel nonlocal total variation model*, IEEE Trans. Geosci. Remote Sensing, 52 (2014), pp. 175–187.

[18] G. Chierchia, N. Pustelnik, B. Pesquet-Popescu, and J.-C Pesquet, *A non-local structure tensor based approach for multicomponent image recovery problems*, IEEE Trans. Image Process., 23 (2014), pp. 5531–5544.

[19] P. L. Combettes and V. R. Wajs, *Signal recovery by proximal forward-backward splitting*, Multiscale Model. Simul., 4 (2005), pp. 1168–1200.

[20] Z. Cvetkovski, *Inequalities: Theorems, Techniques and Selected Problems*, Springer, Heidelberg, 2012.

[21] S. Di Zenzo, *A note on the gradient of a multi-image*, Comp. Vision, Graphics, Image Process., 33 (1986), pp. 116–125.

[22] Y. Dong and M. Hintermüller, *Multi-scale total variation with automated regularization parameter selection for color image restoration*, in Scale Space and Variational Methods in Computer Vision, Springer, Berlin, 2009, pp. 271–281.

[23] D. L. Donoho, *Denoising by soft-thresholding*, IEEE Trans. Inform. Theory, 41 (1995), pp. 613–627.

[24] A. Elmoataz, O. Lezoray, and S. Bougleux, *Nonlocal discrete regularization on weighted graphs: A framework for image and manifold processing*, IEEE Trans. Image Process., 17 (2008), pp. 1047–1060.

[25] E. Esser, *Applications of Lagrangian-Based Alternating Direction Methods and Connections to Split Bregman*, Computational Applied Mathematics report, UCLA, Los Angeles, 2009.
[26] M. A. T. Figueiredo, J. M. Bioucas-Dias, and R. D. Nowak, *Majorization–minimization algorithms for wavelet-based image restoration*, IEEE Trans. Image Process., 16 (2007), pp. 2980–2991.
[27] W. Förstner and E. Gülch, *A fast operator for detection and precise location of distinct points, corners and centres of circular features*, in Proceedings ISPRS Intercommission Conference on Fast Processing of Photogrammetric data, Institut fur Geodäsie und Photogrammetrie, Zurich, 1987, pp. 281–305.
[28] G. Gilboa and S. Osher, *Nonlocal operators with applications to image processing*, Multiscale Model. Simul., 7 (2008), pp. 1005–1028.
[29] G. Gilboa, N. Sochen, and Y. Zeevi, *Variational denoising of partly textured images by spatially varying constraints*, IEEE Trans. Image Process., 15 (2006), pp. 2281–2289.
[30] B. Goldluecke, E. Strekalovskiy, and D. Cremers, *The natural vectorial total variation which arises from geometric measure theory*, SIAM J. Imaging Sci., 5 (2012), pp. 537–563.
[31] T. Goldstein and S. Osher, *The split Bregman method for L_1-regularized problems*, SIAM J. Imaging Sci., 2 (2009), pp. 323–343.
[32] M. Grasmair, *Locally adaptive total variation regularization*, in Scale Space and Variational Methods in Computer Vision, Springer, Berlin, 2009, pp. 331–342.
[33] M. Grasmair and F. Lenzen, *Anisotropic total variation filtering*, Appl. Math. Optim., 62 (2010), pp. 323–339.
[34] V. Hutson, J. S. Pym, and M. J. Cloud, *Applications of Functional Analysis and Operator Theory*, 2nd ed., Elsevier, Amsterdam, 2005.
[35] B. Jähne, *Digital Image Processing*, Springer-Verlag, Berlin, 2002.
[36] S. Lefkimmiatis, A. Roussos, M. Unser, and P. Maragos, *Convex generalizations of total variation based on the structure tensor with applications to inverse problems*, in Scale Space and Variational Methods in Computer Vision, Springer, Berlin, 2013, pp. 48–60.
[37] S. Lefkimmiatis, J. Ward, and M. Unser, *Hessian Schatten-norm regularization for linear inverse problems*, IEEE Trans. Image Process., 22 (2013), pp. 1873–1888.
[38] F. Lenzen, F. Becker, J. Lellmann, S. Petra, and C. Schnörr, *A class of quasi-variational inequalities for adaptive image denoising and decomposition*, Comput. Optim. Appl., 54 (2012), pp. 371–398.
[39] J. Liu and J. Ye, *Efficient ℓ_1/ℓ_q Norm Regularization*, preprint, arXiv:1009.4766, 2010.
[40] Y. Nesterov, *A method for solving a convex programming problem with convergence rates $O(1/k^2)$*, Soviet Math. Dokl., 27 (1983), pp. 372–376.
[41] K. Papafitsoros and C. Schönlieb, *A combined first and second order variational approach for image reconstruction*, J. Math. Imaging Vision, 48 (2014), pp. 308–338.
[42] N. Paragios, Y. Chen, and O. Faugeras, *Handbook of Mathematical Models in Computer Vision*, Springer, New York, 2006.
[43] T. Pock and A. Chambolle, *Diagonal preconditioning for first order primal-dual algorithms in convex optimization*, in IEEE International Conference on Computer Vision (ICCV), IEEE, Piscataway, NJ, 2011, pp. 1762–1769.
[44] S. J. Reeves, *Fast image restoration without boundary artifacts*, IEEE Trans. Image Process., 14 (2005), pp. 1448–1453.
[45] R. T. Rokcafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.
[46] A. Roussos and P. Maragos, *Tensor-based image diffusions derived from generalizations of the total variation and Beltrami functionals*, in Proceedings International Conference on Image Processing, IEEE, Piscataway, NJ, 2010, pp. 4141–4144.
[47] L. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Phys. D, 60 (1992), pp. 259–268.
[48] G. Sapiro, *Color snakes*, Comp. Vision Image Understanding, 68 (1997), pp. 247–253.
[49] S. Setzer, G. Steidl, and T. Teuber, *Restoration of images with rotated shapes*, Numer. Algorithms, 48 (2008), pp. 49–66.
[50] S. Setzer, G. Steidl, and T. Teuber, *Infimal convolution regularizations with discrete ℓ_1-type functionals*, Commun. Math. Sci., 9 (2011), pp. 797–827.
[51] N. Sochen and L. Bar, *The Beltrami-Mumford-Shah functional*, Scale Space and Variational Methods in Computer Vision, Springer, Berlin, 2012, pp. 183–193.

[52] N. Sochen, R. Kimmel, and R. Malladi, *A general framework for low level vision*, IEEE Trans. Image Process., 7 (1998), pp. 310–338.
[53] G. Steidl and T. Teuber, *Anisotropic smoothing using double orientations*, in Scale Space and Variational Methods in Computer Vision, Springer, Berlin, 2009, pp. 477–489.
[54] S. Suvrit, *Fast projections onto $\ell_{1,q}$-norm balls for grouped feature selection*, Machine Learning and Knowledge Discovery in Databases, Vol. 3, Springer, Berlin, 2011, pp. 305–317.
[55] D. Tschumperlé and R. Deriche, *Vector-valued image regularization with PDE's: A common framework for different applications*, IEEE Trans. Pattern Anal. Mach. Intelligence, 27 (2005), pp. 506–517.
[56] C. R. Vogel, *Computational Methods for Inverse Problems*, Front. Appl. Math. 23, SIAM, Philadelphia, 2002.
[57] J. Weickert, *Anisotropic Diffusion in Image Processing*, Teubner, Stuttgart, 1998.
[58] J. Weickert and C. Schnörr, *A theoretical framework for convex regularizers in PDE-based computation of image motion*, Int. J. Comput. Vision, 45 (2001), pp. 245–264.