

The Natural Vectorial Total Variation Which Arises from Geometric Measure Theory*

Bastian Goldluecke[†], Evgeny Strelakovsky[†], and Daniel Cremers[†]

Abstract. Several ways to generalize scalar total variation to vector-valued functions have been proposed in the past. In this paper, we give a detailed analysis of a variant we denote by TV_J , which has not been previously explored as a regularizer. The contributions of the manuscript are twofold: on the theoretical side, we show that TV_J can be derived from the generalized Jacobians from geometric measure theory. Thus, within the context of this theory, TV_J is the most natural form of a vectorial total variation. As an important feature, we derive how TV_J can be written as the support functional of a convex set in \mathcal{L}^2 . This property allows us to employ fast and stable minimization algorithms to solve inverse problems. The analysis also shows that in contrast to other total variation regularizers for color images, the proposed one penalizes across a common edge direction for all channels, which is a major theoretical advantage. Our practical contribution consist of an extensive experimental section, where we compare the performance of a number of provable convergent algorithms for inverse problems with our proposed regularizer. In particular, we show in experiments for denoising, deblurring, superresolution, and inpainting that its use leads to a significantly better restoration of color images, both visually and quantitatively. Source code for all algorithms employed in the experiments is provided online.

Key words. algorithms, duality, vectorial total variation regularization, color image restoration

AMS subject classifications. 68U10, 49M29, 65K10

DOI. 10.1137/110823766

1. Introduction. Regularity is of central importance in computer vision. Many problems, like denoising, deblurring, superresolution, and inpainting, are ill-posed and require the choice of a good prior in order to arrive at sensible solutions. This prior often takes the form of a regularization term for an energy functional which is to be minimized. For optimization purposes, it is important that the regularizer is convex, since only then can one hope to reliably determine global optima of respective energies within reasonable time. Furthermore, since natural images are often piecewise smooth, popular regularizers tend to penalize the function's gradient. In this context, the *total variation (TV)* of a function has emerged as a very successful regularizer for a wide range of applications: It is convex yet discontinuity preserving in the sense that it assigns the same cost to sharp and smooth transitions. Another favorable property of the TV is that application to the binary indicator function of a set gives rise to its boundary length. This geometric interpretation makes it well suited for various shape optimization problems like image segmentation [18, 15] or multiple view reconstruction [27].

*Received by the editors February 8, 2011; accepted for publication (in revised form) December 14, 2011; published electronically April 12, 2012.

<http://www.siam.org/journals/siims/5-2/82376.html>

[†]Fakultät für Mathematik, TU München, 85748 Garching, Germany (bastian.goldluecke@in.tum.de, evgeny.strelakovsky@in.tum.de, cremers@in.tum.de).

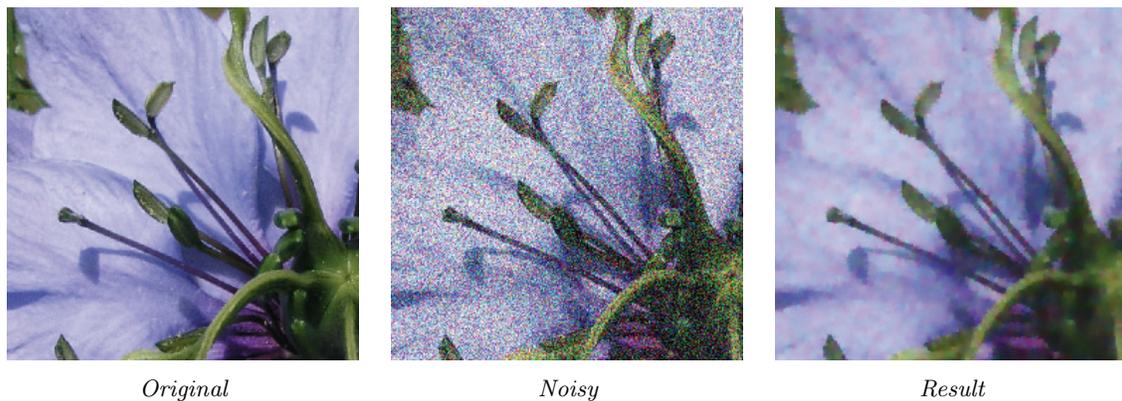


Figure 1. For inverse problems like denoising, inpainting, or superresolution, TV is among the most powerful regularizers. We propose a novel generalization of TV to vector-valued images which naturally arises in the context of geometric measure theory.

While most existing work involving TV focuses on scalar valued functions, the generalization to vector-valued (color or multichannel) images remains an important challenge. In this paper, we will see that this generalization is by no means straightforward, but that in the context of geometric measure theory there is a natural choice (see Figure 1). Our aim is to demonstrate that it is possible to perform an efficient minimization for this variant and that the choice has advantages when solving practical problems.

1.1. Vectorial TV. For a grayscale image modeled as a differentiable function $u \in \mathcal{C}^1(\Omega, \mathbb{R})$ on a domain $\Omega \subset \mathbb{R}^m$, the scalar total variation $\text{TV}(u)$ is defined as the integral over the Euclidean norm $|\cdot|_2$ of the gradient,

$$(1.1) \quad \text{TV}(u) = \int_{\Omega} |\nabla u|_2 \, dx.$$

The definition can be extended to locally integrable functions $u \in \mathcal{L}^1_{\text{loc}}(\Omega, \mathbb{R})$ using a *dual* formulation: Let \mathbb{E}^m denote the closed unit ball in \mathbb{R}^m ; then

$$(1.2) \quad \text{TV}(u) = \sup_{\xi \in \mathcal{C}^1_c(\Omega, \mathbb{E}^m)} \left\{ \int_{\Omega} u \, \text{div}(\xi) \, dx \right\}.$$

The term “dual” means in this context that TV is the support functional of a convex set, namely, the set

$$(1.3) \quad K_{\text{TV}} := \{ \text{div}(\xi) : \xi \in \mathcal{C}^1_c(\Omega, \mathbb{E}^m) \}$$

with respect to the Hilbert space $\mathcal{L}^2(\Omega)$. This means that it is the convex conjugate of the indicator function of K_{TV} ; in particular it is convex and lower semicontinuous on $\mathcal{L}^2(\Omega)$. Note that the right-hand side of (1.2) makes sense for nondifferentiable u (such as the binary indicator function of a set), as derivatives are taken only of the dual vector fields ξ . For a more in-depth introduction to TV for image processing problems see [14].

The idea of *vectorial* TV is to extend the above definition to vector-valued $\mathbf{u} : \Omega \rightarrow \mathbb{R}^n$, such that in the case $n = 1$ both definitions coincide. Several variants have been proposed, which will be discussed in section 2.

An important criterion for a good regularizer is that efficient and reliable minimization algorithms are available. For the scalar TV, such methods have been developed based on the dual formulation, pioneered by Chan, Golub, and Mulet [19]. The more recent algorithm by Chambolle [12] allows one to handle the nondifferentiability of $|\cdot|_2$ without need to regularize, so one can solve the exact model. In view of this, a useful generalization of scalar to vectorial TV should comprise a similar dual formulation.

1.2. Contributions. In this work, we analyze a variant of the vectorial TV which has not been previously explored as a regularizer. While it already appears in the literature as a special case of the framework introduced by Sapiro and Ringach [35, 37], we show that it also follows as a natural choice from the context of geometric measure theory. The proposed variant of vectorial TV is nondifferentiable with no way to smoothen it readily available; in particular, it cannot be minimized correctly with traditional gradient-descent or diffusion techniques. Our main contribution is therefore to analyze the regularizer from the point of view of convex optimization. Like the grayscale TV, it can be written as the support functional of a convex set in \mathcal{L}^2 , which shows that it is convex and closed. In particular, this characterization leads to efficient minimization algorithms, which we discuss in detail.

We complete the mathematical analysis with an in-depth characterization of the regularizer, proving the existence and uniqueness of solutions for vectorial inverse problems. The analysis also shows that in contrast to other TV regularizers for color images, the proposed one penalizes across a common edge direction for all channels, which is a major theoretical advantage. In experiments, we can show that denoising using the new regularizer thus leads to improved restoration of color edges. Since the regularizer can be used as a substitute for vectorial TV in any energy functional, there is a broad spectrum of further applications; see, e.g., [8, 21, 28, 41]. In the paper, we give further examples for deblurring, superresolution, and inpainting. Code to reproduce all examples is available on our web page.¹

Additional contributions compared to the original conference paper [24] are as follows. First, we derive the characterization of the vectorial TV as a support functional of a convex set explicitly. We show exactly how the convex set looks and show how to efficiently compute the necessary orthogonal projection onto this set. Second, we give a detailed description of several possible algorithms to solve inverse problems based on the novel regularizer and compare their performance in additional experiments.

2. Related work. Approaches to defining TV for vector-valued functions can roughly be divided into two classes. The first class of approaches computes the TV channel by channel and takes a suitable norm of the resulting vector. The second class of approaches emerges when considering the Riemann geometry of the image manifold.

2.1. Channel by channel, l^1 -norm. Probably the most simple and straightforward way to deal with multidimensional TV is to sum up the contributions of the separate channels [3].

¹<http://cvpr.in.tum.de>.

This leads to the definition

$$(2.1) \quad \text{TV}_S(\mathbf{u}) := \sum_{i=1}^n \text{TV}(u_i).$$

The dual formulation follows immediately from (1.2),

$$(2.2) \quad \text{TV}_S(\mathbf{u}) = \sup_{(\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_n) \in K_S} \left\{ \sum_{i=1}^n \int_{\Omega} u_i \operatorname{div}(\boldsymbol{\xi}_i) \, dx \right\}$$

with $K_S = \mathcal{C}_c^1(\Omega, \mathbb{E}^m \times \dots \times \mathbb{E}^m)$.

The dual optimization technique of this method is a straightforward generalization of [12] and is fast, robust, and easy to implement. However, since there is no coupling between channels, there is no preservation of color edges and significant color smearing, as we will see in the experiments. Also, there is no rotational invariance in color space. The norm has, for example, been used in TV- L^1 optic flow models [41], where artifacts are not immediately visible as in color image denoising.

2.2. Channel by channel, Euclidean norm. Blomgren and Chan [9] define multidimensional TV as the Euclidean norm of the vector of channelwise scalar TV,

$$(2.3) \quad \text{TV}_{\text{BC}}(\mathbf{u}) := \sqrt{\sum_{i=1}^n \text{TV}(u_i)^2}.$$

From the Euler–Lagrange equations of this norm,

$$(2.4) \quad \frac{\text{TV}(u_i)}{\text{TV}_{\text{BC}}(\mathbf{u})} \operatorname{div} \left(\frac{\nabla u_i}{|\nabla u_i|_2} \right) = 0 \text{ for all } 1 \leq i \leq n,$$

one can observe that there is a coupling of channels, but it is global and very weak, i.e., the same per-channel weight is used for all image pixels. Furthermore, a regularization is required for the denominator of (2.4) in the diffusion process. The authors demonstrated that their variant has quite a few desirable properties, and since it is convex, there is also a dual formulation available. However, the constraint on the dual variables is global instead of pointwise, so the reprojection on the constraint set cannot be fully parallelized, and an implementation will be not as efficient. Because of this shortcoming and the weakness of the coupling, we do not include this norm in further studies.

2.3. Riemann geometry. In [20], di Zenzo suggests considering a vector-valued image as a parameterized two-dimensional Riemann manifold in nD -space. The metric tensor of this manifold is given by

$$(2.5) \quad g_{\mu\nu} = (\partial_{\mu} \mathbf{u}, \partial_{\nu} \mathbf{u}), \quad \mu, \nu = 1, 2.$$

This is analogous to the structure tensor of an image, and the eigenvector corresponding to the smaller eigenvalue gives the direction of the vectorial edge; see Figure 2. Several variants of anisotropic and edge-enhancing diffusion for color images have been developed using this formulation [37, 40], but in general the diffusion process does not arise as the minimizing flow of an energy.

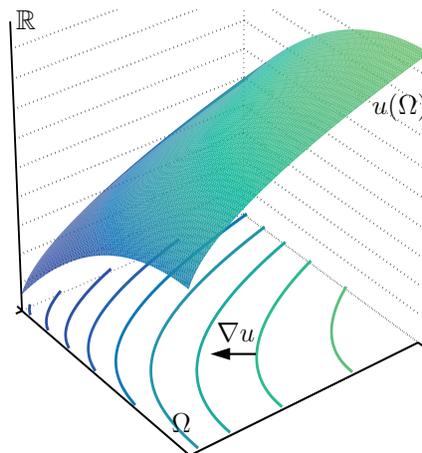


Figure 2. The metric tensor of the image manifold $u(\Omega)$ is given by $(g_{\mu\nu}) = (D\mathbf{u})^T D\mathbf{u}$. The eigenvector corresponding to the largest eigenvalue λ_1 gives the direction of the vectorial edge. In the familiar case $n = 1$ illustrated above, this is the direction of the gradient ∇u , which is always orthogonal to the level lines.

Based on this framework, Sapiro [35] suggests a family of possible definitions for the vectorial TV, which is of the form

$$(2.6) \quad \text{TV}_{\text{SR}}(\mathbf{u}) := \int_{\Sigma} f(\lambda_+, \lambda_-) ds,$$

where λ_{\pm} denote the larger and smaller eigenvalue of $(g_{\mu\nu})$, respectively, and f is a suitable scalar-valued function. It must be noted that TV_{SR} is in general defined only for differentiable functions; only for special cases are dual formulations available to extend it to locally integrable functions.

Another approach based on Riemann geometry was pioneered in the framework by Kimmel, Malladi, and Sochen [25, 38, 26]. In contrast to the approaches above, it considers a higher-dimensional manifold, namely, the graph of the image as a surface in $\mathbb{R}^2 \times \mathbb{R}^n$. The surface area is employed for regularization and is minimized in order to smoothen the image. This approach leads to a diffusion equation with the direction given by the Beltrami flow. It is possible to derive a dual formulation in the case of grayscale images [10]; however, no such dualization is known in the vectorial case. Thus, there are currently no efficient algorithms available to minimize this surface area in the case of vectorial images.

2.4. Pointwise Frobenius norm. A special case of the Sapiro–Ringach TV (2.6) is the choice $f(\lambda_{\pm}) = \sqrt{\lambda_+ + \lambda_-}$, which generalizes to the Frobenius norm of the derivative $D\mathbf{u}$

$$(2.7) \quad \text{TV}_F(\mathbf{u}) := \int_{\Omega} \|D\mathbf{u}(x)\|_F dx.$$

in the case of an n -dimensional image space Ω . The choice above is a remarkable case because there is a convenient dual formulation, which extends the definition from differentiable to

locally integrable functions,

$$(2.8) \quad \text{TV}_F(\mathbf{u}) = \sup_{(\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_n) \in K_F} \left\{ \sum_{i=1}^n \int_{\Omega} u_i \operatorname{div}(\boldsymbol{\xi}_i) \, dx \right\}$$

with $K_F = \mathcal{C}_c^1(\Omega, \mathbb{E}^{n \cdot m})$.

This can be seen as another straightforward generalization of single-channel TV if one compares to its dual formulation. Notably, definitions (2.8) and (2.2) are equivalent in the sense that they lead to the same space $\mathcal{BV}(\Omega, \mathbb{R}^n)$. However, the actual results in image processing algorithms are quite different, since there is no correlation of channels in TV_S , in contrast to TV_F , which has a desirable coupling of channels. In the books [1, 3], both definitions appear depending on the preference of the authors.

Efficient minimization techniques for functionals based on (2.8) have been intensively studied by Bresson and Chan [10] as well as Duval, Aujol, and Vese [22]. Because of its good performance, TV_F has emerged as a favorite candidate for vectorial TV, and it is often referred to as *the* vectorial TV, although it is only a single one in the large family (2.6) proposed by Sapiro and Ringach.

However, it was already noted by Blomgren and Chan [9] that TV_F has some less-than-ideal properties. In particular, it actually favors gray value images over colored ones, which leads to color smearing, for example, in denoising applications. That the coupling of channels is not optimal can also be seen in the dual formulation. While the edge strength is correctly weighted over all channels such that common edges are not overly penalized, the edge directions can be different for the different channels.

3. Vectorial TV. In the following, we will show that in the family of generalizations of TV to vector-valued functions proposed by Sapiro and Ringach [35] is the one which emerges naturally in light of geometric measure theory. The resulting vectorial TV approach supports a common edge direction for all channels, comprises important invariance properties, and comes with a dual formulation that allows for stable and exact minimization schemes.

This section contains our main theoretical contributions. We first introduce the regularizer TV_J from the point of view of geometric measure theory and then show how it relates to the framework of Sapiro and Ringach. The key results are in subsection 3.2, where we compute the convex set of which TV_J is the support functional, and thus the convex conjugate. This will be central to the efficient minimization methods introduced in later parts of the paper. The remainder of the section then collects some other useful properties of the regularizer TV_J .

3.1. Definition. Geometric measure theory [23] is focused on geometric properties of the measures of sets, for example, a set's arc length and area. One of the central concepts in geometric measure theory is the notion of a *Jacobian* J_k , a generalization of the Jacobian determinant to the case $k \leq n$. We will only require the case $k = 1$, which we are going to explain in the following.

For a scalar valued differentiable function u it turns out that $J_1 u = |\nabla u|_2$, which implies that the TV corresponds to the integral of $J_1 u$. Therefore, a natural generalization of the TV

to a vector-valued functions $\mathbf{u} : \Omega \rightarrow \mathbb{R}^n$ is given by the integral

$$(3.1) \quad \text{TV}_J(\mathbf{u}) := \int_{\Omega} J_1 \mathbf{u} \, dx.$$

The precise meaning is made clear by the following proposition, which relates the Jacobian to the singular values $\sigma_1(D\mathbf{u}), \dots, \sigma_m(D\mathbf{u})$ of the derivative matrix $D\mathbf{u}$ in case of differentiable \mathbf{u} .

Proposition 3.1. *For functions $\mathbf{u} \in \mathcal{C}^1(\Omega, \mathbb{R}^n)$, the vectorial total variation $\text{TV}_J(\mathbf{u})$ equals the integral over the largest singular value of the derivative matrix,*

$$(3.2) \quad \text{TV}_J(\mathbf{u}) = \int_{\Omega} \sigma_1(D\mathbf{u}) \, dx.$$

In particular, TV_J is equal to the standard TV for real-valued functions.

Proof. The Jacobian $J_1 \mathbf{u}$ is defined as the operator norm of $\Lambda_1 D\mathbf{u} = D\mathbf{u}$, which can be computed explicitly via

$$(3.3) \quad \begin{aligned} J_1 \mathbf{u} &= \|\Lambda_1 D\mathbf{u}\| \\ &= \sup_{\boldsymbol{\xi} \in \mathbb{E}^m} \left\{ |(\Lambda_1 D\mathbf{u})(\xi_1 \mathbf{e}_1 + \dots + \xi_m \mathbf{e}_m)|_2 \right\} \\ &= \sup_{(\boldsymbol{\xi}, \boldsymbol{\eta}) \in \mathbb{E}^m \times \mathbb{E}^n} \left\{ \sum_{j=1}^m (\xi_j D\mathbf{u}(\mathbf{e}_j), \boldsymbol{\eta}) \right\} \\ &= \sup_{(\boldsymbol{\xi}, \boldsymbol{\eta}) \in \mathbb{E}^m \times \mathbb{E}^n} \left\{ \sum_{i=1}^n \eta_i \sum_{j=1}^m \xi_j \partial_j u_i \right\} \\ &= \sup_{(\boldsymbol{\xi}, \boldsymbol{\eta}) \in \mathbb{E}^m \times \mathbb{E}^n} \left\{ \sum_{i=1}^n \boldsymbol{\eta}^T D\mathbf{u} \cdot \boldsymbol{\xi} \right\}. \end{aligned}$$

The claim follows now from the singular value decomposition of $D\mathbf{u}$. ■

Interestingly, although the motivation for the above formulation comes from a completely different direction, Proposition 3.1 shows an intimate relationship of the proposed formulation to the Sapiro–Ringach approach [37]. To see this, note that the metric tensor of the image manifold is equal to $(g_{\mu\nu}) = (D\mathbf{u})^T D\mathbf{u}$, in particular, $\sigma_1(D\mathbf{u}) = \sqrt{\lambda_+}$; see Figure 2. Thus, TV_J is a special case of (2.6), similar to the model TV_F . In previous works, the norm $\sqrt{\lambda_+}$ has already been employed as a local geometry indicator to construct diffusion flows for image restoration [11, 36, 39]. However, those flows do not appear as gradient flows of an energy, and to our knowledge, our paper is the first one which actually directly uses the norm as a regularizer in energy functionals. This is a large conceptual and in particular algorithmic difference. Note that a correct minimization with diffusion-based methods is not possible, since the function σ_+ is not differentiable—from the explicit formula in the case $m = 2, n = 3$ of color images one can see that a smooth approximation is not as simple to achieve as in the case of TV.

From the new context, we can derive a dual formulation for TV_J , which leads to a very efficient optimization method for the proposed regularizer. Furthermore, this dual formulation allows us to extend the definition to nondifferentiable functions, which is not possible in the Sapiro–Ringach formulation.

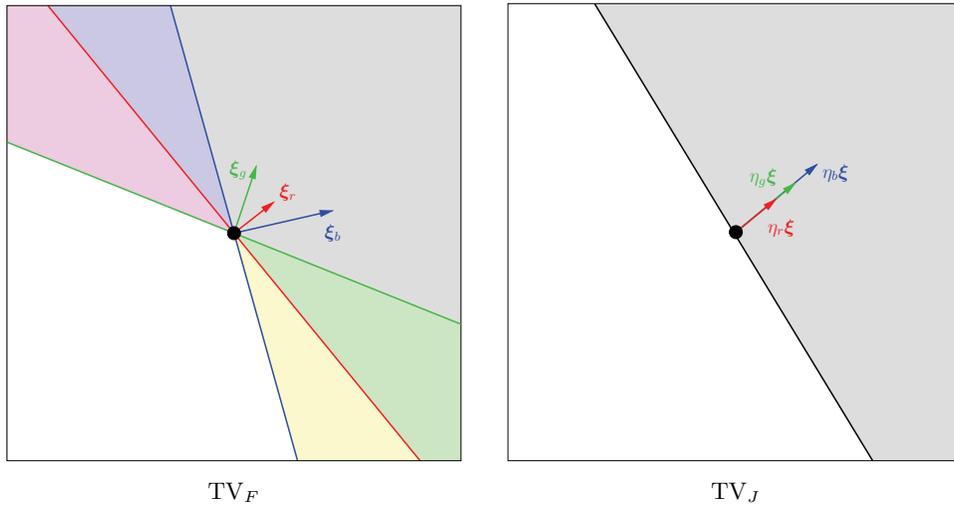


Figure 3. Theoretical advantage of TV_J . The vectorial total variation TV_F based on the Frobenius norm has different edge directions for each color channel, while TV_J yields only one common direction with channel-by-channel weights; see also Figure 4 and Proposition 3.2. This leads to a better preservation of color edges in inverse problems.

Proposition 3.2. On $C^1(\Omega, \mathbb{R}^n)$, the vectorial total variation TV_J can be expressed as

$$(3.4) \quad TV_J(\mathbf{u}) = \sup_{(\boldsymbol{\xi}, \boldsymbol{\eta}) \in K_J} \left\{ \sum_{i=1}^n \int_{\Omega} u_i \operatorname{div}(\eta_i \boldsymbol{\xi}) \, dx \right\}$$

with $K_J = C_c^1(\Omega, \mathbb{E}^m \times \mathbb{E}^n)$.

The right-hand side is well defined for all $\mathbf{u} \in \mathcal{L}_{loc}^1(\Omega, \mathbb{R}^n)$.

Proof. The proof follows from the representation (3.3) for $J_1 \mathbf{u}$ and the Gaussian divergence theorem. ■

Because $\sigma_1(\cdot)$ and $\|\cdot\|_F$ are equivalent norms on $\mathbb{R}^{n \times m}$, TV_J leads to the same space $\mathcal{BV}(\Omega, \mathbb{R}^n)$ of functions of bounded vectorial TV.

Note that in representation (3.4), the term below the integral contains products of $\boldsymbol{\xi}$ and η_i , which means that although K_J is convex, the right-hand side cannot be a support functional of a convex set. We will see how we can rewrite it to a support functional (and thus derive the true dual formulation) in section 3.2. However, from representation (3.4) we can notice one major feature of TV_J . While TV_F and TV_S have different edge directions ξ_i for each color channel, in the formulation TV_J there is a single direction $\boldsymbol{\xi}$ which is weighted among the channels. This is illustrated in Figure 3, while a theoretical comparison of the different variants can be seen in Figure 4. We refer to [26] for a discussion of why the coupled edge direction is ideal for many natural images, in particular, how it follows from a Lambertian assumption about the lighting model.

3.2. Convex analysis. We now compute the convex conjugate of TV_J . From Proposition 3.2, it is already obvious that TV_J is equal to the support functional of a closed convex set $\mathcal{K} \subset \mathcal{L}^2(\Omega, \mathbb{R}^n)$ and that its conjugate TV_J^* is the indicator function of \mathcal{K} . In particu-

<i>Variant</i>	<i>Primal</i>	<i>Dual</i>	<i>Properties</i>
$\text{TV}_S(\mathbf{u})$	$\sum_{i=1}^n \int_{\Omega} \nabla u_i _2 \, dx$	$\sup_{(\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_n) \in K_S} \left\{ \sum_{i=1}^n \int_{\Omega} u_i \operatorname{div}(\boldsymbol{\xi}_i) \, dx \right\}$ with $K_S = \mathcal{C}_c^1(\Omega, \mathbb{E}^m \times \dots \times \mathbb{E}^m)$	separate directions, unweighted
$\text{TV}_F(\mathbf{u})$	$\int_{\Omega} \ D\mathbf{u}(x)\ _F \, dx$	$\sup_{(\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_n) \in K_F} \left\{ \sum_{i=1}^n \int_{\Omega} u_i \operatorname{div}(\boldsymbol{\xi}_i) \, dx \right\}$ with $K_F = \mathcal{C}_c^1(\Omega, \mathbb{E}^{n \cdot m})$	separate directions, weighted
$\text{TV}_J(\mathbf{u})$	$\int_{\Omega} J_1(\mathbf{u}) \, dx$	$\sup_{(\boldsymbol{\xi}, \boldsymbol{\eta}) \in K_J} \left\{ \sum_{i=1}^n \int_{\Omega} u_i \operatorname{div}(\eta_i \boldsymbol{\xi}) \, dx \right\}$ with $K_J = \mathcal{C}_c^1(\Omega, \mathbb{E}^m \times \mathbb{E}^n)$	shared direction, weighted

Figure 4. Comparison of dual formulations and properties.

lar, TV_J is convex, closed, and positive homogeneous on the Hilbert space $\mathcal{L}^2(\Omega, \mathbb{R}^n)$. The remainder of the section is devoted to computing the set \mathcal{K} .

We first require some additional notation. For a vector-valued function $\mathbf{u} \in \mathcal{C}^1(\Omega)$, we define the gradient operator Grad componentwise as

$$(3.5) \quad \text{Grad}(\mathbf{u})(x) = \begin{bmatrix} \nabla u_1(x) \\ \vdots \\ \nabla u_n(x) \end{bmatrix} \in \mathbb{R}^{n \times m}.$$

Similarly, for a tensor field $\boldsymbol{\zeta} \in \mathcal{C}_c^1(\Omega, \mathbb{R}^{n \times m})$, the divergence operator Div is the regular divergence acting on the rows,

$$(3.6) \quad \text{Div}(\boldsymbol{\zeta})(x) = \text{Div} \left(\begin{bmatrix} \boldsymbol{\zeta}_1 \\ \vdots \\ \boldsymbol{\zeta}_n \end{bmatrix} \right) (x) := \begin{bmatrix} \operatorname{div}(\boldsymbol{\zeta}_1)(x) \\ \vdots \\ \operatorname{div}(\boldsymbol{\zeta}_n)(x) \end{bmatrix} \in \mathbb{R}^n.$$

Note that since $\boldsymbol{\zeta}$ has compact support, $\langle \text{Grad}(\mathbf{u}), \boldsymbol{\zeta} \rangle = \langle \mathbf{u}, -\text{Div}(\boldsymbol{\zeta}) \rangle$ for all such \mathbf{u} and $\boldsymbol{\zeta}$, in particular, $\text{Grad}^* = -\text{Div}$. Here, $\langle \cdot, \cdot \rangle$ denotes the inner product for \mathcal{L}^2 functions of the respective dimensionality.

The reformulation of TV_J can best be understood by representing matrices in $\mathbb{R}^{n \times m}$ with Kronecker products. Let $\mathbf{a} \in \mathbb{R}^n$ and $\mathbf{b} \in \mathbb{R}^m$; then their Kronecker product $\mathbf{a} \otimes \mathbf{b}$ is by definition an $n \cdot m$ vector, which throughout the rest of the work we identify with an $n \times m$ matrix

$$(3.7) \quad \mathbf{a} \otimes \mathbf{b} := \begin{bmatrix} a_1 \mathbf{b} \\ \vdots \\ a_n \mathbf{b} \end{bmatrix} = \begin{bmatrix} a_1 \mathbf{b}^T \\ \vdots \\ a_n \mathbf{b}^T \end{bmatrix}.$$

Using this notation, one can rewrite the term in the representation (3.4) of TV_J as

$$(3.8) \quad \sum_{i=1}^n \int_{\Omega} u_i \operatorname{div}(\eta_i \boldsymbol{\xi}) \, dx = \langle \mathbf{u}, \operatorname{Div}(\boldsymbol{\eta} \otimes \boldsymbol{\xi}) \rangle.$$

Thus, one can identify the desired convex set \mathcal{K} as follows.

Proposition 3.3. *Let $\mathbf{u} \in \mathcal{L}^2(\Omega, \mathbb{R}^n)$. Then*

$$(3.9) \quad \begin{aligned} TV_J(\mathbf{u}) &= \sigma_{\mathcal{K}}(\mathbf{u}) = \sup_{\mathbf{v} \in \mathcal{K}} \langle \mathbf{u}, \mathbf{v} \rangle \\ \text{with } \mathcal{K} &:= \{ \operatorname{Div}(\boldsymbol{\zeta}) : \boldsymbol{\zeta} \in \mathcal{C}_c^1(\Omega, \operatorname{co}(\mathbb{E}^n \otimes \mathbb{E}^m)) \}; \end{aligned}$$

thus TV_J is convex, closed, and positive homogeneous on the Hilbert space $\mathcal{L}^2(\Omega, \mathbb{R}^n)$. Above, $\operatorname{co}(\mathbb{E}^n \otimes \mathbb{E}^m)$ denotes the convex hull of the Kronecker product $\mathbb{E}^n \otimes \mathbb{E}^m$, which is formed element by element according to (3.7).

Proof. If we rewrite (3.4) using (3.8), we can see that

$$(3.10) \quad TV_J(\mathbf{u}) = \sup_{(\boldsymbol{\xi}, \boldsymbol{\eta}) \in \mathbb{E}^n \otimes \mathbb{E}^m} \langle \mathbf{u}, \operatorname{Div}(\boldsymbol{\eta} \otimes \boldsymbol{\xi}) \rangle.$$

We arrive at the result by observing that a support functional of a set A is equal to the support functional of its convex hull $\operatorname{co}(A)$ and taking the convex hull commutes with the linear mapping Div . ■

What remains to be shown is how to characterize $\operatorname{co}(\mathbb{E}^n \otimes \mathbb{E}^m)$. The characterization can be found in the literature, but we review the arguments here for the sake of completeness. The following proposition gives the abstract conditions for a matrix to lie in this set.

Proposition 3.4. *We identify vectors in $\mathbb{R}^{n \cdot m}$ with matrices $\mathbb{R}^{n \times m}$. Then*

$$(3.11) \quad \operatorname{co}(\mathbb{E}^n \otimes \mathbb{E}^m) = \bigcap_{B \in \mathbb{R}^{n \times m}} \{ A \in \mathbb{R}^{n \times m} : \langle A, B \rangle \leq \sigma_1(B) \},$$

where $\sigma_1(B)$ denotes the largest singular value of B .

Proof. Let $E := \mathbb{E}^n \otimes \mathbb{E}^m \subset \mathbb{R}^{n \cdot m} = \mathbb{R}^{n \times m}$. According to the Eidelheit separation theorem (which is itself a direct consequence of the theorem of Mazur, or geometric Hahn–Banach) [33], the convex hull of E can be written as the intersection of all half-spaces which contain E . Using the support functional² $\sigma_E : \mathbb{R}^{n \times m} \rightarrow \mathbb{R}$ of E , we can write this intersection as

$$(3.12) \quad \operatorname{co}(\mathbb{E}^n \otimes \mathbb{E}^m) = \bigcap_{B \in \mathbb{R}^{n \times m}} \{ A \in \mathbb{R}^{n \times m} : \langle A, B \rangle \leq \sigma_E(B) \}.$$

Since the support functional can be computed explicitly as

$$(3.13) \quad \sigma_E(B) = \sup_{\boldsymbol{\eta} \otimes \boldsymbol{\xi} \in \mathbb{E}^n \otimes \mathbb{E}^m} \langle \boldsymbol{\eta} \otimes \boldsymbol{\xi}, B \rangle = \sup_{\boldsymbol{\eta} \in \mathbb{E}^n, \boldsymbol{\xi} \in \mathbb{E}^m} \sum_{i=1}^n \sum_{j=1}^m \eta_i b_{ij} \xi_j = \sigma_1(B),$$

we arrive at the claim of the proposition. ■

²The standard notation σ_E for the support functional of the set E is not to be confused with the notation $\sigma_i(A)$ for the singular values of a matrix A . Note that the subscript is a set in the first case, an integer in the second.

We have now a full characterization of the convex hull; however, it has an infinite number of constraints. We can replace it with a single constraint when we take into account that the matrix norm dual to the spectral norm $\sigma_1(\cdot)$ is the *nuclear norm*, which is the sum of all singular values,

$$(3.14) \quad |A|_* := \sum \sigma_i(A).$$

To make this precise, we can formulate the following duality theorem.

Proposition 3.5. *For every $A \in \mathbb{R}^{n \times m}$,*

$$(3.15) \quad |A|_* = \max\{\langle A, B \rangle : \sigma_1(B) \leq 1\}.$$

Proof. See [32, Proposition 2.1]. ■

As a consequence, we can write the desired convex hull as follows.

Corollary 3.6. *The convex hull of $\mathbb{E}^n \otimes \mathbb{E}^m$ is the nuclear unit ball,*

$$(3.16) \quad \text{co}(\mathbb{E}^n \otimes \mathbb{E}^m) = \{A \in \mathbb{R}^{n \times m} : |A|_* \leq 1\}.$$

Proof. For the proof, we note that from (3.15) it follows that for all $B \in \mathbb{R}^{n \times m}$,

$$(3.17) \quad \langle A, B \rangle \leq |A|_* \cdot \sigma_1(B)$$

with equality holding for certain B . Plugging this into the support plane characterization (3.11), we arrive at the result. ■

For minimization algorithms later on, it will be necessary to project back onto \mathcal{K} . We can compute the orthogonal projection via the following theorem.

Theorem 3.7. *Let $A \in \mathbb{R}^{n \times m}$ with singular value decomposition $A = U\Sigma V^T$. Then the orthogonal projection of A onto $\text{co}(\mathbb{E}^n \otimes \mathbb{E}^m)$ is*

$$(3.18) \quad P_{\text{co}(\mathbb{E}^n \otimes \mathbb{E}^m)}(A) = U \text{diag}(\sigma_p) V^T,$$

where σ_p is the projection of $(\sigma_1, \dots, \sigma_n)$ onto the l_1 -unit ball

$$(3.19) \quad \left\{x \in \mathbb{R}^n : \sum |x_i| \leq 1\right\}.$$

Proof. This follows from the main theorem proved in [30], which is in turn a consequence of Mirsky’s inequality [29]. ■

An important special case is to compute the projection for the 3×2 matrices which appear in color image processing. We give an efficient explicit algorithm for this case in the appendix.

3.3. Relations to other norms. The proposed vectorial total variation TV_J is systematically smaller than TV_F , which is in turn smaller than TV_S . This means that when using these for regularization, different scaling factors have to be used to reach a similar level of smoothing, which makes results not directly comparable. We compensate for this in the experiments by choosing an optimal value for each regularizer separately.

Proposition 3.8. *For all $\mathbf{u} \in \mathcal{BV}(\Omega, \mathbb{R}^n)$,*

$$(3.20) \quad TV_J(\mathbf{u}) \leq TV_F(\mathbf{u}) \leq TV_S(\mathbf{u}).$$

Proof. For differentiable \mathbf{u} , we can actually see that the inequality holds pointwise for the integrands, since

$$\begin{aligned} \sigma_1(D\mathbf{u}) &\leq \sqrt{\sigma_1(D\mathbf{u})^2 + \cdots + \sigma_m(D\mathbf{u})^2} \\ (3.21) \qquad &= \sqrt{|\nabla u_1|_2^2 + \cdots + |\nabla u_n|_2^2} \\ &\leq |\nabla u_1|_2 + \cdots + |\nabla u_n|_2. \end{aligned}$$

By a standard density argument, it follows for all of $\mathcal{BV}(\Omega, \mathbb{R}^n)$. ■

3.4. Invariance properties. We show that $\text{TV}_J(\mathbf{u})$ is invariant with respect to both orthogonal transformations in *image space* as well as in *color space*. Those important properties are not shared by all variants of the vectorial TV; see Blomgren and Chan [9] for an analysis. The first proposition shows that TV_J is invariant under orthogonal reparametrizations of Ω .

Proposition 3.9. *Let $\Omega' \subset \mathbb{R}^m$ and $A : \Omega' \rightarrow \Omega$ be an orthogonal transformation. Then for any $\mathbf{u} \in \mathcal{L}_{loc}^1(\Omega, \mathbb{R}^n)$,*

$$\text{TV}_J(\mathbf{u} \circ A) = \text{TV}_J(\mathbf{u}).$$

Proof. Since the divergence is invariant under orthogonal transformations, this follows from representation (3.4) and the integral transformation formula. ■

Furthermore, $\text{TV}_J(\mathbf{u})$ is invariant under isometries of the color space \mathbb{R}^n . Note that this is incorrect for most variants of vectorial TV, a notable exception being TV_F .

Proposition 3.10. *Let $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be an isometry with respect to the Euclidean norm. Then*

$$\text{TV}_J(T\mathbf{u}) = \text{TV}_J(\mathbf{u}).$$

Proof. The proof follows from the representation (3.3). Note that $D(T\mathbf{u})(\mathbf{e}_j) = T \cdot D\mathbf{u}(\mathbf{e}_j)$ and T maps the unit sphere in \mathbb{R}^n onto itself. ■

4. TV_J in inverse problems. In image processing problems, one frequently considers inverse problems where the goal is to recover an original image $\mathbf{u} \in L^2(\Omega, \mathbb{R}^n)$ from a noisy observation $\mathbf{f} \in L^2(\Omega, \mathbb{R}^n)$ of a possibly further degraded version $A\mathbf{u}$. Since the model $A\mathbf{u} = \mathbf{f}$ is usually underdetermined and in order to correctly model the noise, it is necessary to impose a prior model for the original image \mathbf{u} and impose $A\mathbf{u} = \mathbf{f}$ only as a soft constraint. We thus want to solve the problem

$$(4.1) \qquad \operatorname{argmin}_{\mathbf{u} \in L^2(\Omega, \mathbb{R}^n)} \left\{ \text{TV}_J(\mathbf{u}) + \frac{1}{2\lambda} \|A\mathbf{u} - \mathbf{f}\|_p^p \right\},$$

where $\text{TV}_J(\mathbf{u})$ serves as an objective, improper prior. In the case $p = 2$, the above model can be interpreted as a maximum a posteriori estimate under the assumption of Gaussian noise on the observation, while the choice of $p = 1$ is suitable for salt-and-pepper noise. The constant $\lambda > 0$ controls the desired smoothness of the result—the larger, the greater the influence of the regularizer.

In this section, we verify that the usual inverse problems with the regularizer TV_J are well defined, among them the proximation operator, which is important for algorithms. Most results are standard, shown in a similar way as for the grayscale or other vectorial TVs, but for the sake of completeness we reproduce them here and adapt the arguments to our case.

4.1. Existence and uniqueness of solutions. In order to guarantee existence of solutions, we need to make further assumptions on the operator A .

Theorem 4.1. *Let $A : L^2(\Omega, \mathbb{R}^n) \rightarrow L^2(\Omega, \mathbb{R}^n)$ be bounded, symmetric, and coercive with constant $\alpha > 0$, i.e., for all $\mathbf{u} \in L^2(\Omega, \mathbb{R}^n)$ we have*

$$(4.2) \quad \langle A\mathbf{u}, \mathbf{u} \rangle \geq \alpha \|\mathbf{u}\|_2.$$

Then problem (4.1) admits a unique minimizer $u \in L^2(\Omega, \mathbb{R}^n)$.

Proof. Note that coercivity of A implies in particular

$$(4.3) \quad \|\mathbf{A}\mathbf{u}\|_2 \geq \frac{\langle A\mathbf{u}, \mathbf{u} \rangle}{\|\mathbf{u}\|_2} \geq \alpha \|\mathbf{u}\|_2 \rightarrow \infty \text{ as } \|\mathbf{u}\|_2 \rightarrow \infty.$$

Together with the inverse triangle inequality, this implies that the energy of (4.1) is coercive on $L^2(\Omega, \mathbb{R}^n)$. Proposition 3.3 implies that it is lower semicontinuous and strictly convex as the sum of the TV_J seminorm and \mathcal{L}^2 norm. These properties are sufficient to guarantee a unique minimizer on the Hilbert space $L^2(\Omega, \mathbb{R}^n)$; see Theorems 3.2.5 and 3.3.3 in [3]. ■

Condition (4.2) is probably too strong; see the discussion in the case of grayscale TV in [16]. However, this analysis is beyond the scope of this work.

4.2. The $\text{TV}_J\text{-}\mathcal{L}^2$ Model. The ROF model, named after Rudin, Osher, and Fatemi, who introduced it in [34], is a very popular approach to image denoising, designed to restore an image which was degraded by Gaussian noise. It is designed for grayscale images and employs TV as a prior. With the vectorial TV, it can easily be generalized to vector-valued images. This leads to the $\text{TV}_J\text{-}\mathcal{L}^2$ model for denoising, which is to solve for a given noisy image $\mathbf{f} \in L^2(\Omega, \mathbb{R}^n)$ the minimization problem

$$(4.4) \quad \operatorname{argmin}_{\mathbf{u} \in \mathcal{BV}(\Omega, \mathbb{R}^n)} \left\{ \text{TV}_J(\mathbf{u}) + \frac{1}{2\lambda} \|\mathbf{u} - \mathbf{f}\|_2^2 \right\}.$$

Choosing A as the identity I , we can see that it is a special case of (4.1) with coercivity constant $\alpha = 1$. In this special case, we can refine the proof for existence of a solution and show an additional property of the minimizer.

Theorem 4.2. *Denote the (componentwise) mean of a function \mathbf{u} on Ω by $\mathbf{u}_\Omega := \frac{1}{|\Omega|} \int_\Omega \mathbf{u} \, dx$. Problem (4.4) admits a unique solution $\bar{\mathbf{u}}$ which has the same mean as \mathbf{f} .*

Proof. We give the basic ideas of the proof and refer the reader to [22] for more details. The energy is strictly convex as a sum of the convex TV_J seminorm (see Proposition 3.3) and the strictly convex norm of the L^2 -Hilbertspace. Thus, what remains to be shown is existence of a solution on the space

$$V_f := \{\mathbf{u} \in L^2(\Omega, \mathbb{R}^n) : \mathbf{u}_\Omega = \mathbf{f}_\Omega\}.$$

We already know that the energy is lower semicontinuous on V_f (Proposition 3.3). In addition, the energy is coercive on V_f : from the Poincaré inequality (Theorem 5.3.1 in [3]) we have

$$\|\mathbf{u} - \mathbf{f}_\Omega\|_2 = \|\mathbf{u} - \mathbf{u}_\Omega\|_2 \leq C \cdot \text{TV}_J(\mathbf{u})$$

with a constant $C > 0$, since the Frobenius and operator norms are equivalent norms on $\mathbb{R}^{n \times m}$. Weak lower semicontinuity and coercivity imply the existence of a minimizer $\bar{\mathbf{u}}$ in V_f (Theorem 3.2.5 in [3]). Since the constant function \mathbf{f}_Ω is of bounded variation and $\mathbf{f}_\Omega \in V_f$, it follows that $\bar{\mathbf{u}} \in \mathcal{BV}(\Omega, \mathbb{R}^n)$. ■

5. Efficient minimization. In this section, we present algorithms to minimize the general inverse problem of the form

$$(5.1) \quad \min_{\mathbf{u} \in \mathcal{L}^2(\Omega, \mathbb{R}^n)} \left\{ \text{VTV}(\mathbf{u}) + \frac{1}{2\lambda} \|A\mathbf{u} - \mathbf{f}\|_p^p \right\},$$

where VTV is either channel-by-channel vectorial total variation TV_S , the vectorial total variation TV_F based on the Frobenius norm of the derivative matrix, or our novel regularizer TV_J . All three regularizers are closed and convex, which allows us to apply powerful algorithms from convex optimization to solving inverse problems of the type (5.1). In this paper, we discuss and compare a selected few of them. All algorithms are explicitly detailed in Figure 8 in the appendix.

5.1. Bermúdez–Moreno algorithm for the VTV- \mathcal{L}^2 model. The Bermúdez–Moreno (BM) algorithm was introduced in [7] and subsequently reintroduced in [4] in the context of image processing problems. It can be shown that for the grayscale ROF model, it is the same as the projection scheme proposed by Chambolle in [13]. Because VTV is convex and lower semicontinuous, the BM algorithm can be directly adapted to the VTV- \mathcal{L}^2 model as well and leads to a very similar algorithm. The sequence of iterations is detailed in Figure 8 in the appendix. In [7, 4], it was shown that a step size of $\sigma < 1/(4\lambda) = 2/(\|\text{Div}\|^2 \cdot \lambda)$ is optimal, in the sense that convergence is proved and experimentally the scheme becomes unstable for any $\sigma \geq 1/(4\lambda)$. Note that we use simple forward and backward differences to discretize gradient and divergence. The norm of the discrete operator Div is the same as $\|\text{div}\| = \frac{1}{\sqrt{8}}$, since

$$(5.2) \quad \begin{aligned} \|\text{Div}(\boldsymbol{\zeta})\| &= \sqrt{\|\text{div}(\boldsymbol{\zeta}_1)\|^2 + \dots + \|\text{div}(\boldsymbol{\zeta}_n)\|^2} \\ &\leq \|\text{div}\| \sqrt{\|\boldsymbol{\zeta}_1\|^2 + \dots + \|\boldsymbol{\zeta}_n\|^2} \\ &= \|\text{div}\| \cdot \|\boldsymbol{\zeta}\|. \end{aligned}$$

No estimate on the rate of convergence is given in the above references; however, the observed convergence rate in experiments is $\mathcal{O}(1/n)$.

5.2. Primal-dual methods for the general inverse problem. The conjugates of the various VTV regularizers allow us to transform problem (5.1) into the saddle point problem

$$(5.3) \quad \begin{aligned} &\min_{\mathbf{u} \in \mathcal{L}^2(\Omega, \mathbb{R}^n)} \max_{\boldsymbol{\zeta} \in \mathcal{L}^2(\Omega, \mathbb{R}^{n \times m})} \left\{ -\langle \mathbf{u}, \text{Div}(\boldsymbol{\zeta}) \rangle - \delta_K(\boldsymbol{\zeta}) + \frac{1}{2\lambda} \|A\mathbf{u} - \mathbf{f}\|_p^p \right\}, \\ &\text{where } K = \begin{cases} \mathcal{C}_c^1(\Omega, \text{co}(\mathbb{E}^n \otimes \mathbb{E}^m)) & \text{for } \text{TV}_J, \\ \mathcal{C}_c^1(\Omega, \mathbb{E}^{n \cdot m}) & \text{for } \text{TV}_F, \\ \mathcal{C}_c^1(\Omega, \mathbb{E}^m \times \dots \times \mathbb{E}^m) & \text{for } \text{TV}_S, \end{cases} \\ &\text{and the indicator function } \delta_K(\boldsymbol{\zeta}) := \begin{cases} 0 & \text{for } \boldsymbol{\zeta} \in K, \\ \infty & \text{otherwise.} \end{cases} \end{aligned}$$

To solve this formulation, a number of first order primal-dual algorithms were suggested, of which we compare the following two. Both work for the general inverse problem with both \mathcal{L}^1 and \mathcal{L}^2 regularization in the data term. For handling general operators A and the non-differentiable \mathcal{L}^1 -norm, we modified the saddle point problem according to a suggestion by Chambolle and Pock (section 6.3.1 in [17]) to also dualize the norm in the data term. We arrive at the saddle point problem

$$(5.4) \quad \min_{\mathbf{u} \in \mathcal{L}^2(\Omega, \mathbb{R}^n)} \max_{\substack{\boldsymbol{\zeta} \in \mathcal{L}^2(\Omega, \mathbb{R}^{n \times m}) \\ \mathbf{q} \in \mathcal{L}^2(\Omega, \mathbb{R}^n)}} \left\{ -\langle \mathbf{u}, \text{Div}(\boldsymbol{\zeta}) \rangle + \langle A\mathbf{u} - \mathbf{f}, \mathbf{q} \rangle - \delta_K(\boldsymbol{\zeta}) - n_p^*(\mathbf{q}) \right\}.$$

Here, n_p^* is the convex conjugate of the weighted norm $\mathbf{v} \mapsto \frac{1}{2\lambda} \|\mathbf{v}\|_p^p$, i.e.,

$$(5.5) \quad n_p^*(\mathbf{q}) = \begin{cases} \delta_{\mathbb{E}^n} \left(\frac{\mathbf{q}}{2\lambda} \right) & \text{if } p = 1, \\ \frac{\lambda}{2} |\mathbf{q}|_2^2 & \text{if } p = 2, \end{cases}$$

with the equalities understood pointwise.

5.2.1. The Arrow–Hurwicz method. For reference, we test the algorithm which we employed in our original conference paper [24]. It comprises a straightforward subgradient descent in the primal variable \mathbf{u} and ascent in the dual variable $\boldsymbol{\zeta}$, as detailed in Figure 8 in the appendix. This technique was already suggested in [2] and is named the Arrow–Hurwicz (AH) method after the authors. Unfortunately, it lacks a proof of optimal step sizes for which the scheme converges as well as a convergence rate. Experiments suggest that the algorithm remains stable up to the suggested step sizes $\tau, \sigma = 1/\sqrt{8} = 1/\|\text{Div}\|$, and behaves asymptotically exactly as the BM scheme.

5.2.2. Acceleration and variable step sizes. Chambolle and Pock suggest a number of improvements to the simple subgradient descent and ascent scheme in [17], as a variant of the algorithm presented previously in [31]. From their work, we implemented algorithm 2, which enjoys a proven convergence rate of $\mathcal{O}(1/n^2)$.

5.3. Fast iterative shrinking and thresholding for the general inverse problem. A well-known method to solve problems of the form

$$(5.6) \quad \min_u J(u) + F(u)$$

with closed and convex J, F and differentiable F is called *iterative shrinking and thresholding* (ISTA). One ISTA step comprises a gradient descent in the differentiable F followed by a subgradient descent in J . The step size in both cases is $1/L$, where L is a Lipschitz constant of ∇F . Since a subgradient descent step is equal to the evaluation of the proximation operator for J , this means that a sequence of VTV- \mathcal{L}^2 problems needs to be solved when the idea is applied to problem (5.1). Typically, the inner ROF problems are not solved exactly. Instead, a fixed number k of steps of any iterative ROF solver is applied. In our experiments, we chose $k = 10$. For the general inverse problem, $L = \|A^*A\|/\lambda$.

Beck and Teboulle showed in [6] how to accelerate this simple scheme by applying an acceleration step with variable step sizes after each ISTA iteration. They called this new

scheme fast ISTA (FISTA), and it is capable of solving the general inverse problem with \mathcal{L}^2 data term while achieving the same optimal $\mathcal{O}(1/n^2)$ convergence as Chambolle and Pock. Beck and Teboulle also proposed an accelerated algorithm for ROF obtained by applying FISTA to the dual problem, which they call FGP [5]. We adapted this algorithm to the VTV- \mathcal{L}^2 problem as well as to compare it with the other schemes, see Figure 9.

6. Experimental results.

6.1. Denoising: The vectorial ROF model. In our first experiment, we tested the performance of both regularizers and algorithms in solving the basic VTV- \mathcal{L}^2 model. Since this model can be interpreted as an image restoration method for images degraded by Gaussian noise, we first tested which of the regularizers gives the best denoising results. Since the weight λ is not directly comparable, we solved each denoising problem with a range of different λ and only reported the best result for each regularizer in terms of peak signal-to-noise ratio (PSNR). The results can be observed in Figure 10 in the appendix. We can see that TV_J systematically yields a higher PSNR, with TV_F being a close second. As expected, results from TV_S are significantly worse than from the other two, since there is no coupling of channels and significant smearing of colors. As previously observed by Blomgren and Chan [9], TV_F also exhibits a tendency to smear colors across edges, but it is less pronounced. The visually best results are achieved with TV_J . Since all color channels share a common edge direction, color edges are preserved and no color degradation can be observed.

Besides quality of results, algorithmic performance in solving VTV- \mathcal{L}^2 is also very important. The reason is that solving the VTV- \mathcal{L}^2 model is equivalent to computing the proximation operator for VTV, which can be interpreted as a subgradient descent step via the Moreau–Yosida approximation of the subgradient [4]. Thus, solving VTV- \mathcal{L}^2 is an integral step of some algorithms designed for more general functionals, with one example being FISTA. A full performance comparison of the algorithms detailed in the last section is displayed in Figure 9 in the appendix. We can see the the $\mathcal{O}(1/n^2)$ algorithms FGP and CP are systematically superior to BM and AH by a large margin. Since in particular the implementation of CP only amounts to a simple adaption of step sizes and acceleration step compared to AH, there is no reason not to use it. Comparing FGP and CP, it seems that FGP is slightly faster in number of iterations and total time required, although a single iteration is more costly. CP is slightly more easy to implement, however, and requires less memory. One should also note that BM has a relatively fast convergence in the first iterations, before acceleration fully kicks in. Thus, we use it in the core of our FISTA implementation, since in the inner iteration only a few steps of an ROF solver are applied. We provide source code for all experiments in a GPL-licensed library.³

6.2. Denoising: \mathcal{L}^1 data term. Taking the \mathcal{L}^2 norm (i.e., $p = 2$) in the data term of (5.1) is the suitable choice in the case of Gaussian noise, since in this case the energy corresponds to the maximum a posteriori estimate. However, if the image is degraded with salt-and-pepper type noise, then the choice of $p = 2$ fails and we need to implement the more general model (5.4) for $p = 1$. This is illustrated in Figure 11 in the appendix with a performance comparison.

³<http://www.sourceforge.net/p/cocolib>.



Figure 5. Deblurring using TV_J . Colors are reproduced correctly in the result, and sharp edges are restored.

6.3. Deblurring: The general inverse problem. To demonstrate the performance on general inverse problems, we implemented vectorial versions of TV-deblurring [21] and TV-based superresolution [28]. Deblurring is based on the general inverse problem (5.1) with $A\mathbf{u} := \mathbf{b} * \mathbf{u}$ and the convolution defined componentwise. Note that $\|A\| = \|A^*\| = \|\mathbf{b}\|_1$. Superresolution slightly generalizes the data term to accommodate for several translated input images; see [28] for details. Results are displayed in Figure 5 and in Figure 12 in the appendix for deblurring and in Figure 6 for superresolution. One can again note that TV_J reproduces sharp color edges correctly without smearing colors.

6.4. Inpainting and different color spaces. All previous experiments were performed using the standard RGB color space. However, the regularizer TV_J penalizes the different color channels uniformly. For this reason, it makes sense to assume that the visually most pleasing results can be achieved if one uses a perceptually uniform color space, a property which the common RGB space does not have. The color space CIE- $L^*a^*b^*$ was defined with this property as a goal. We test if this choice makes a perceptual difference in another standard application, where we inpaint damaged regions of an image. Note that the transformation

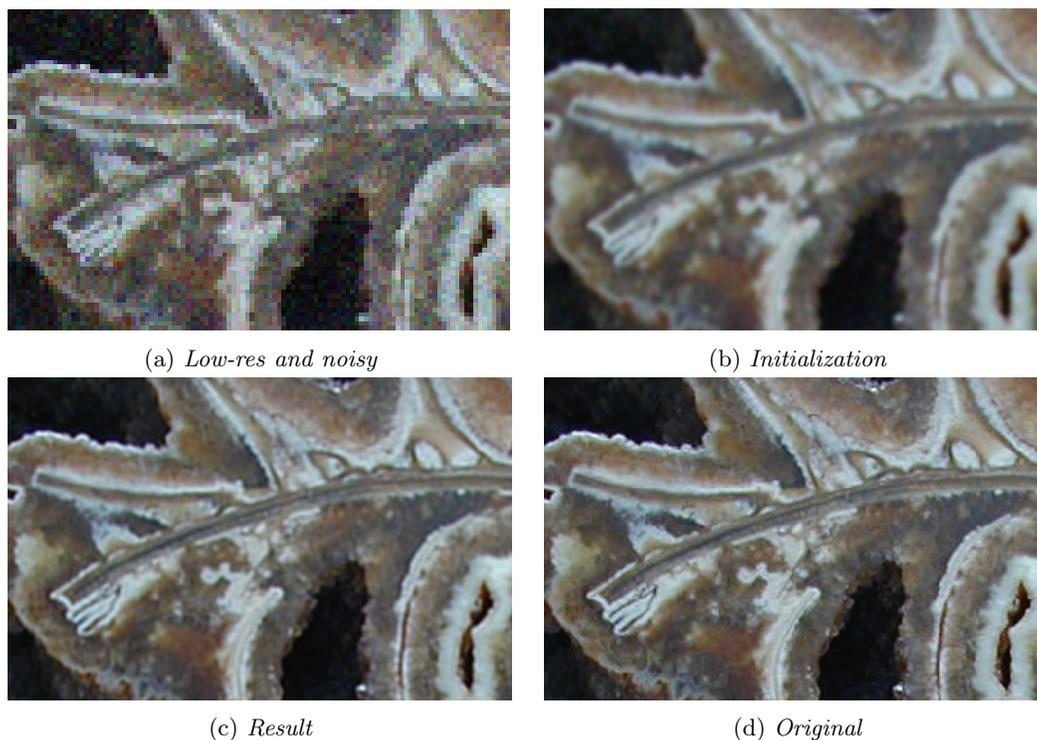


Figure 6. Superresolution using TV_J . (a) One of 16 low-resolution, noisy input images. (b) Initialization obtained by averaging the input images. (c) Result from running the algorithm. It can be observed that color edges are preserved very well and no color smearing occurs. (d) Original image.

between RGB and CIE- $L^*a^*b^*$ is very complex; in particular, none of the invariance properties applies.

As a simple inpainting model, we chose the $VTV-\mathcal{L}^2$ functional with the data term set to zero inside the damaged regions. Since the data term is differentiable, optimization can efficiently be performed with the FISTA algorithm. Results are presented in Figure 13 in the appendix. One can observe that the choice of the uniform color space leads to a very slight but nevertheless visually noticeable improvement, independent of the choice of λ . We conclude that it probably pays off to perform a color space transformation before running image processing algorithms on color images.

7. Conclusion. By referring to geometric measure theory, we have derived a definition for TV_J , a TV of vector-valued functions, as an integral over a generalized Jacobian. We showed that it corresponds to a special case of a family of norms proposed by Sapiro. We provided a complete mathematical characterization of TV_J . In particular, we proved that similar to the Frobenius norm TV_F , the norm TV_J also admits a dual formulation in the sense that it can be written as the support functional of a closed and convex set. This formulation identifies TV_J as a closed and convex functional, which allows us to use powerful minimization algorithms from convex optimization to solve inverse problems. In experiments, we demonstrated that denoising results are superior compared to other variants of vectorial TV

in that color edges are preserved better. We also showed that TV_J can serve as a regularizer in more general energy functionals, which makes it applicable to general inverse problems like deblurring, zooming, inpainting, and superresolution.

7.1. Projection Π_S for TV_S . Since each channel is treated separately, we can compute the well-known projection for the scalar TV for each color channel. Let $A \in \mathbb{R}^{n \times m}$ with rows $a_1, \dots, a_n \in \mathbb{R}^m$. Then Π_S is defined rowwise as

$$(7.1) \quad \Pi_S(a_i) = \frac{a_i}{\max(1, |a_i|_2)}.$$

7.2. Projection Π_F for TV_F . Let $A \in \mathbb{R}^{n \times m}$ with elements $a_{ij} \in \mathbb{R}$. From (2.8) we see that we need to compute the projection onto the unit ball in $\mathbb{R}^{n \cdot m}$ when (a_{ij}) is viewed as a vector in $\mathbb{R}^{n \cdot m}$. Thus,

$$(7.2) \quad \Pi_F(A) = \frac{A}{\max\left(1, \sqrt{\sum_{i=1}^n \sum_{j=1}^m a_{ij}^2}\right)}.$$

7.3. Projection Π_J for TV_J . Let $A \in \mathbb{R}^{n \times m}$ with singular value decomposition $A = U\Sigma V^T$ and $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_m)$. We assume that the singular values are ordered with σ_1 being the largest. If the sum of the singular values is less than or equal to one, A already lies in $\text{co}(\mathbb{E}^n \otimes \mathbb{E}^m)$. Otherwise, according to Theorem 3.18,

$$(7.3) \quad \Pi(A) = U\Sigma_p V^T \text{ with } \Sigma_p = \text{diag}(\sigma_p).$$

To compute the matrix V and the singular values, note that the Eigenvalue decomposition of the $m \times m$ matrix $A^T A$ is given by $V\Sigma^2 V^T$, which is more efficient to compute than the full singular value decomposition since $m < n$. For images, $m = 2$, so there is even an explicit formula available. We can now simplify the formula (7.3) to make the computation of U unnecessary. Let Σ^+ denote the pseudoinverse of Σ which is given by

$$(7.4) \quad \Sigma^+ = \text{diag}\left(\frac{1}{\sigma_1}, \dots, \frac{1}{\sigma_k}, 0, \dots, 0\right),$$

where σ_k is the smallest nonzero singular value. Then $U = AV\Sigma^+$, and from (7.3) we conclude

$$(7.5) \quad \Pi(A) = AV\Sigma^+ \Sigma_p V^T.$$

For the special case of color images, where $n = 3$ and $m = 2$, the implementation of (7.5) is detailed in Figure 7.

Appendix A. In this appendix we show explicitly how to compute the projection $\Pi_K : \mathbb{R}^{n \times m} \rightarrow \mathbb{R}^{n \times m}$ required for the algorithms in Figure 8 for the different types of vectorial total variation. In all cases, Π_K is the orthogonal projection onto a closed convex set K , which is given for the different regularizers in (5.3).

```

// Input : 3x2 matrix A in double variables a11 ... a32
// Output: projection \Pi(A) in double variables a11 ... a32

// Compute D = A^T A
double d11 = a11*a11 + a21*a21 + a31*a31;
double d12 = a12*a11 + a22*a21 + a32*a31;
double d22 = a12*a12 + a22*a22 + a32*a32;

// Compute Eigenvalues (lmax, lmin)
// = squares of singular values (smax, smin)
double trace = d11 + d22;
double det = d11*d22 - d12*d12;
double d = sqrt( max( 0.0, 0.25*trace*trace - det ) );
double lmax = max( 0.0, 0.5 * trace + d );
double lmin = max( 0.0, 0.5 * trace - d );
double smax = sqrt( lmax );
double smin = sqrt( lmin );

// Check whether matrix lies outside K and needs to be projected
if ( smax + smin > 1.0 ) {

    // Compute orthonormal system of Eigenvectors, such that
    // (v11, v21) belongs to lmax and (v12, v22) belongs to lmin
    double v11, v12, v21, v22;
    if ( d12 == 0.0 ) {
        if ( d11 >= d22 ) {v11 = 1.0; v21 = 0.0; v12 = 0.0; v22 = 1.0;}
        else {v11 = 0.0; v21 = 1.0; v12 = 1.0; v22 = 0.0;}
    }
    else {
        v11 = lmax - d22; v21 = d12;
        double l1 = hypot( v11, v21 );
        v11 /= l1; v21 /= l1;
        v12 = lmin - d22; v22 = d12;
        double l2 = hypot( v12, v22 );
        v12 /= l2; v22 /= l2;
    }

    // Project (smax,smin) to line (0,1) + tau * (1,-1), 0<=tau<=1.
    double tau = 0.5 * (smax - smin + 1.0);
    double s1 = min( 1.0, tau );
    double s2 = 1.0 - s1;
    // Compute \Sigma^+ * \Sigma_p
    s1 /= smax;
    s2 = (smin > 0.0) ? s2 / smin : 0.0;

    // Compute T = V * \Sigma^+ * \Sigma_p * V^T
    double t11 = s1*v11*v11 + s2*v12*v12;
    double t12 = s1*v11*v21 + s2*v12*v22;
    double t21 = s1*v21*v11 + s2*v22*v12;
    double t22 = s1*v21*v21 + s2*v22*v22;

    // Result \Pi(A) = A*T
    a11 = a11 * t11 + a12 * t21;
    a21 = a21 * t11 + a22 * t21;
    a31 = a31 * t11 + a32 * t21;
    a12 = a11 * t12 + a12 * t22;
    a22 = a21 * t12 + a22 * t22;
    a32 = a31 * t12 + a32 * t22;
}

```

Figure 7. Computation of the projection $\Pi(A)$ onto $\text{co}(\mathbb{E}^3 \otimes \mathbb{E}^2)$ for a 3×2 matrix A .

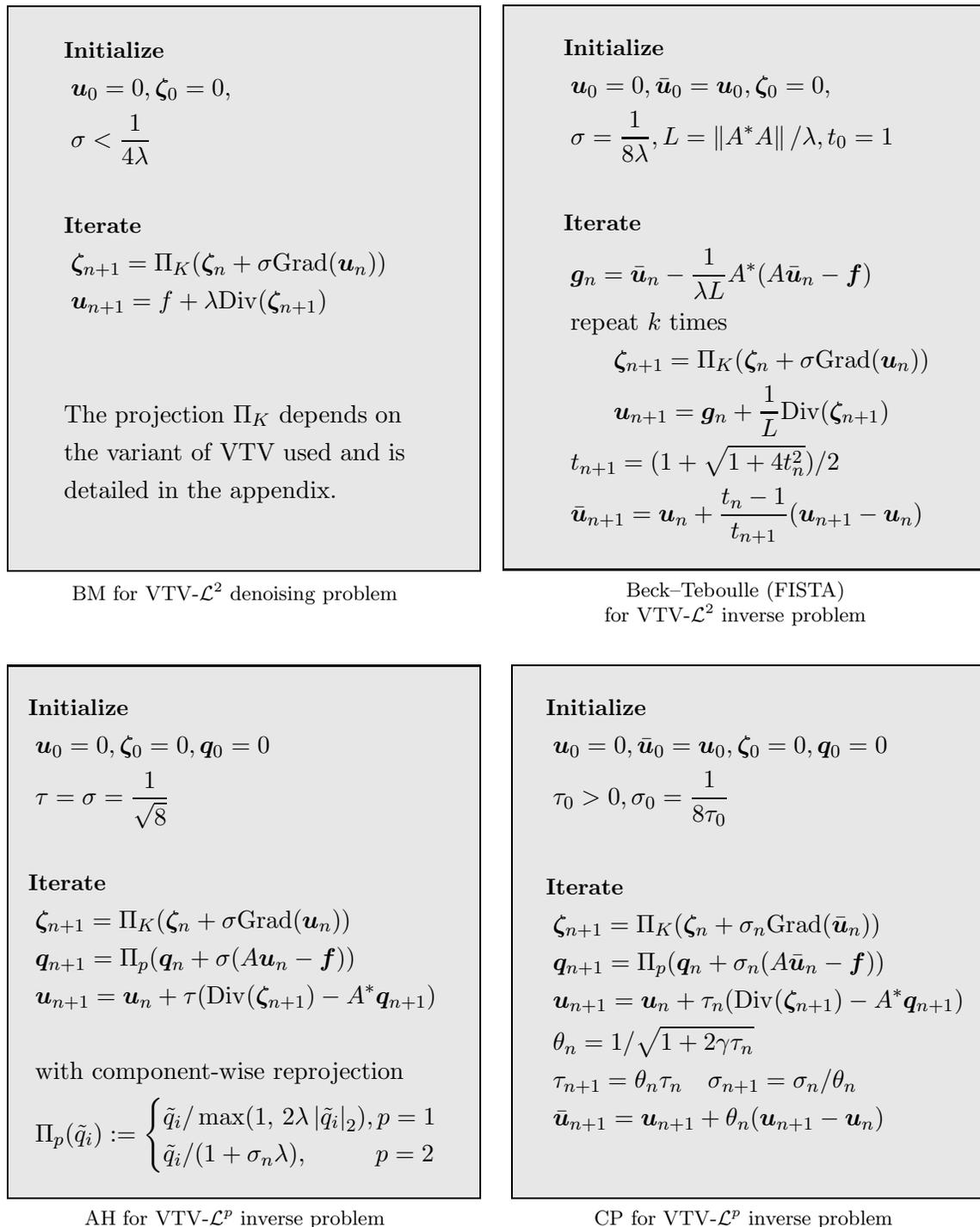
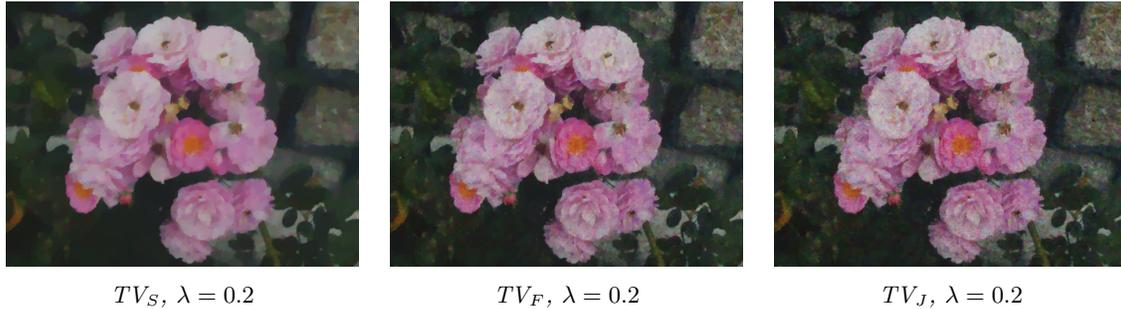
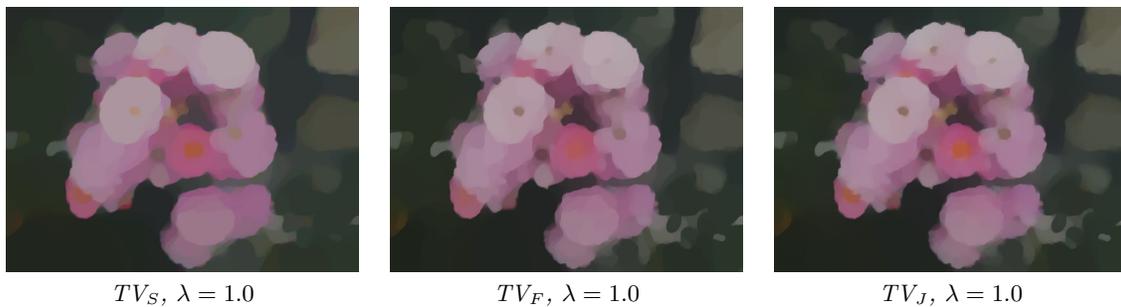


Figure 8. Algorithms to solve inverse problems with vectorial TV regularization. The BM algorithm is suitable only for the VTV- \mathcal{L}^2 ROF model, while FISTA can handle general inverse problems of the type VTV- \mathcal{L}^2 . Only the AH and CP methods can in addition deal with nondifferentiable \mathcal{L}^1 data terms. See [12] for the correct discretization of the gradient and divergence operators.



$\lambda = 0.2$	TV_J		TV_F		TV_S	
	$\epsilon = 10^{-4}$	$\epsilon = 10^{-6}$	$\epsilon = 10^{-4}$	$\epsilon = 10^{-6}$	$\epsilon = 10^{-4}$	$\epsilon = 10^{-6}$
AH	31 (0.15s)	285 (1.37s)	42 (0.19s)	493 (2.24s)	122 (0.53s)	2220 (9.7s)
BM	31 (0.14s)	284 (1.28s)	41 (0.17s)	493 (2.05s)	121 (0.5s)	2220 (9.11s)
CP	19 (0.09s)	164 (0.79s)	20 (0.09s)	166 (0.75s)	26 (0.11s)	217 (0.95s)
FGP	11 (0.06s)	38 (0.2s)	12 (0.06s)	62 (0.3s)	22 (0.11s)	98 (0.47s)



$\lambda = 1.0$	TV_J		TV_F		TV_S	
	$\epsilon = 10^{-4}$	$\epsilon = 10^{-6}$	$\epsilon = 10^{-4}$	$\epsilon = 10^{-6}$	$\epsilon = 10^{-4}$	$\epsilon = 10^{-6}$
AH	297 (1.42s)	4230 (20.2s)	408 (1.85s)	9236 (41.9s)	889 (3.89s)	-
BM	290 (1.31s)	4228 (19.1s)	402 (1.67s)	9235 (38.5s)	880 (3.61s)	-
CP	119 (0.57s)	1078 (5.2s)	129 (0.59s)	1419 (6.4s)	168 (0.74s)	1459 (6.4s)
FGP	31 (0.16s)	338 (1.8s)	71 (0.35s)	492 (2.4s)	149 (0.71s)	631 (3.0s)

Figure 9. Comparison of convergence speed of several algorithms and regularizers. The tables give the number of iterations and the time it takes the algorithms shown in Figure 8 to reach an accuracy of ϵ for the VTV-ROF model. More complicated rejections have only a small impact on time per iteration. For the same λ , the more complex regularizers are smaller, so less smoothing occurs, which tends to reduce the total number of iterations until convergence.

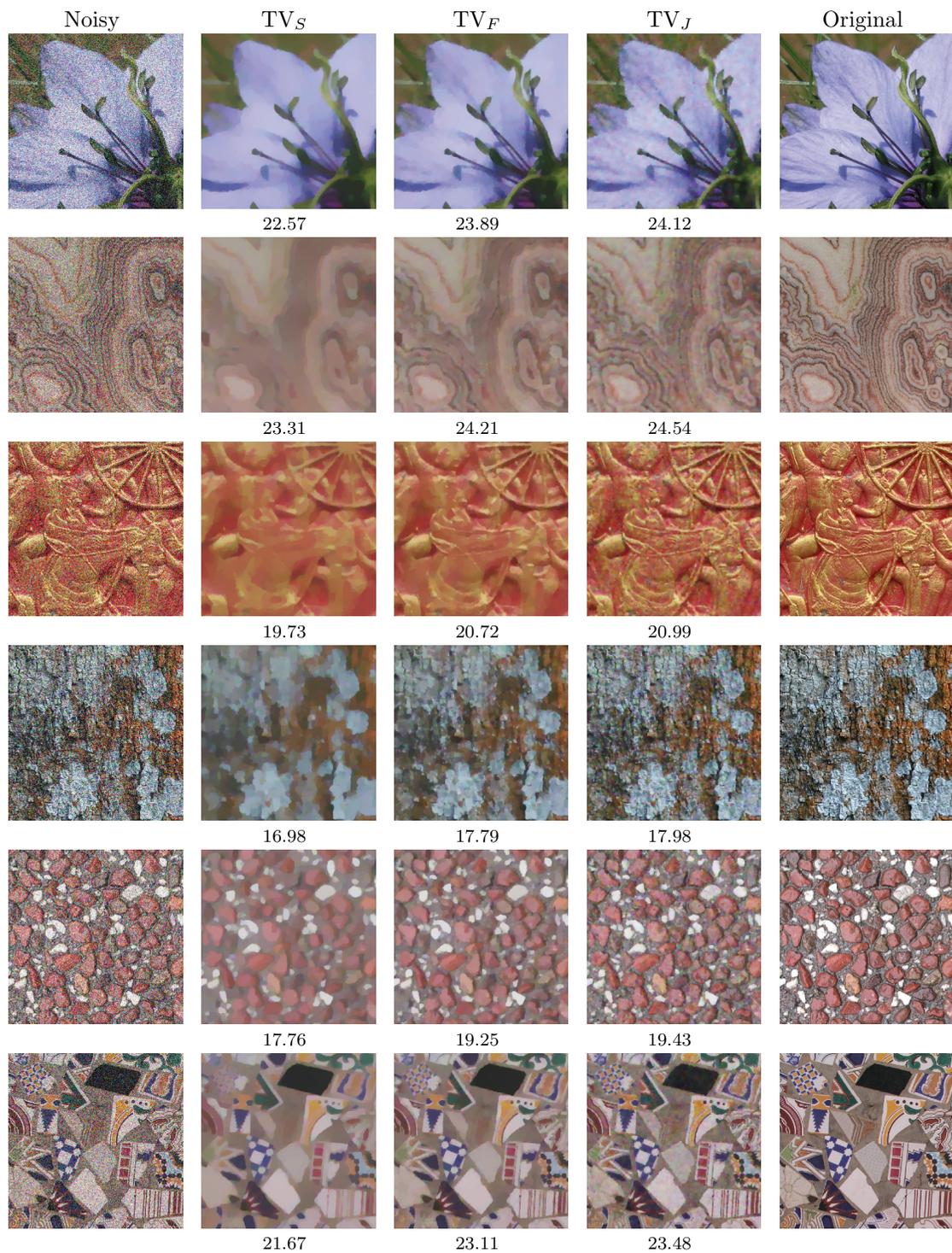
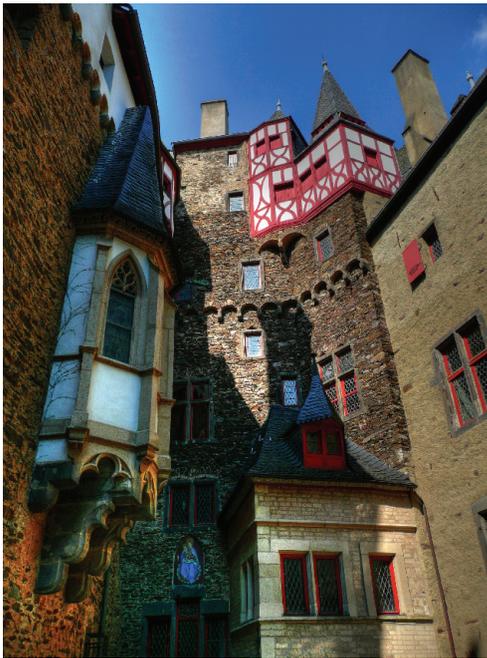
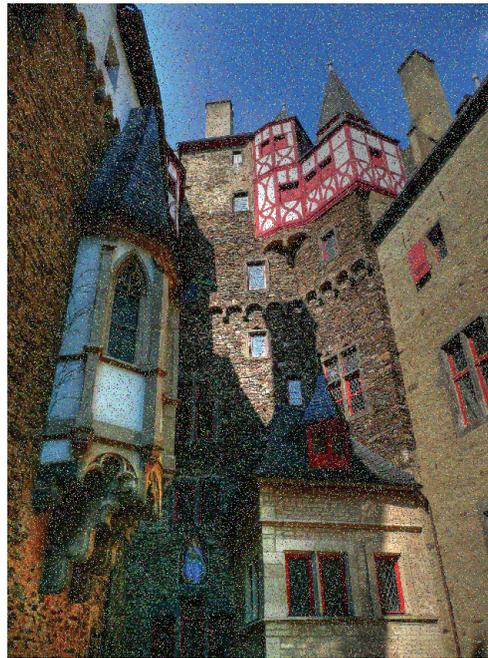


Figure 10. Denoising results for input images with additive Gaussian noise, standard deviation $\sigma = 0.2$. For each method, the value of λ which gave the best results was determined experimentally. The PSNR for each result is noted below the image. These experiments show that compared to both TV_S and TV_F , the proposed TV_J gives rise to significantly better visual quality and to systematically superior PSNR values.



Original



15% of pixels with random color

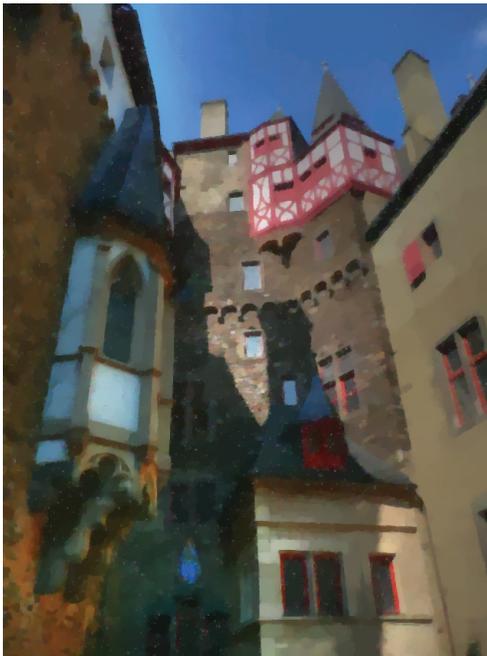
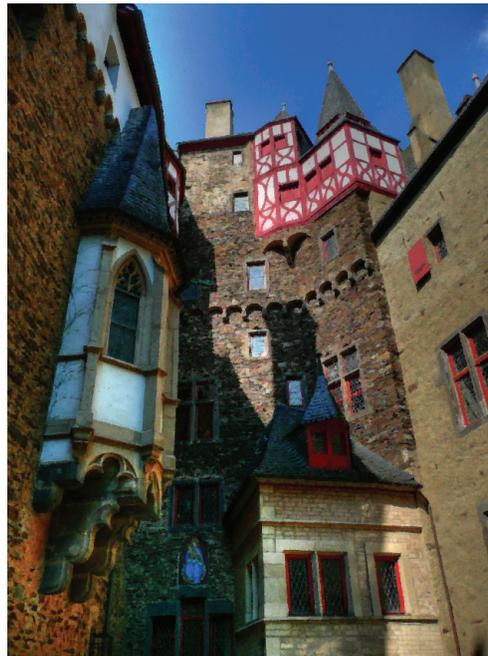
 $TV_J\text{-}\mathcal{L}^2$ denoising (CP algorithm) $TV_J\text{-}\mathcal{L}^1$ denoising (CP algorithm)

Figure 11. Denoising salt-and-pepper noise using TV_J . The regularization weight λ was optimized for the visually best result in both cases. The \mathcal{L}^2 model in the data term is clearly unsuitable for salt-and-pepper noise, since it requires too much regularization to even partially get rid of the noise. On the other hand, the \mathcal{L}^1 model works very well and the regularizer preserves colors and sharp edges. Note that this model requires algorithms CP or AH, where CP exhibits a much faster convergence in practice.



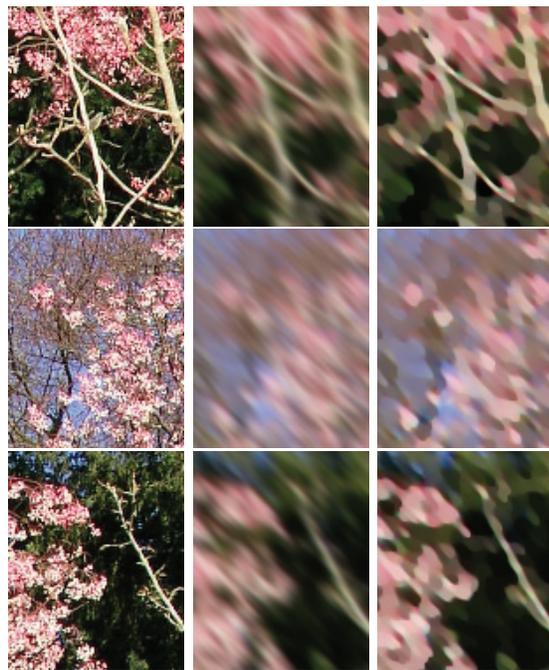
Original



Blurred and noisy



$TV_J\text{-}\mathcal{L}^2$ deblurring (CP algorithm)



Closeup comparison

Figure 12. Deblurring using TV_J . This example shows much stronger anisotropic blur. Colors are still preserved very well.

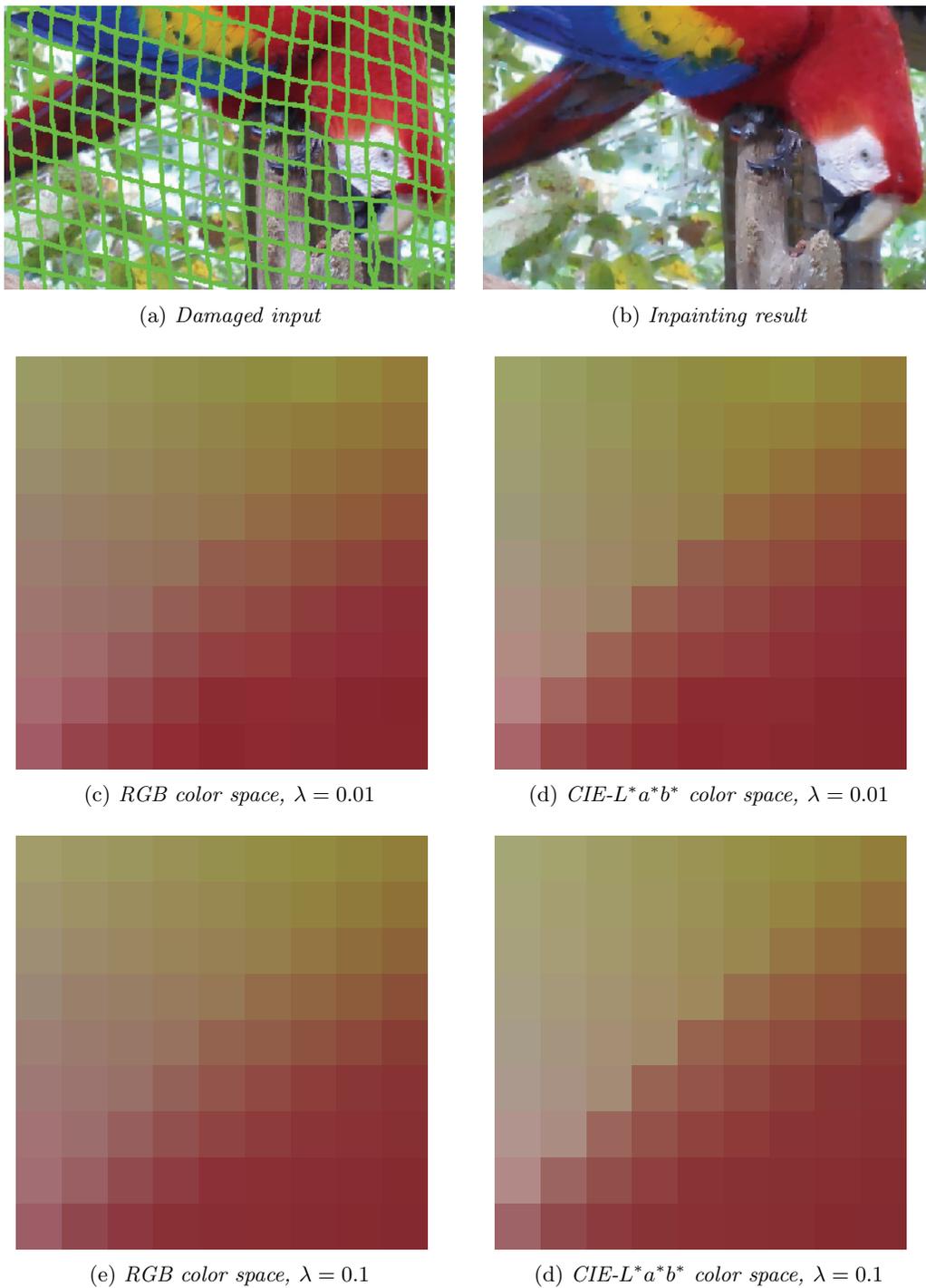


Figure 13. Inpainting using TV_J with different color spaces. The closeups show using the RGB color space leads to a slight bleeding of the color red across the edge. It is particularly visible in the bottom left corner of the closeups, giving some of the green pixels an undesired red taint. The effect is very small but visually noticeable and is independent of the choice of the regularization parameter λ . It vanishes when transforming the image to the perceptually uniform CIE- $L^*a^*b^*$ color space before inpainting.

REFERENCES

- [1] L. AMBROSIO, N. FUSCO, AND D. PALLARA, *Functions of Bounded Variation and Free Discontinuity Problems*, Oxford University Press, Oxford, UK, 2000.
- [2] K. ARROW, *Studies in Linear and Non-linear Programming*, Stanford Mathematical Studies in the Social Sciences 11, Stanford University Press, Stanford, CA, 1958.
- [3] H. ATTOUCH, G. BUTTAZZO, AND G. MICHAILLE, *Variational Analysis in Sobolev and BV Spaces*, MPS/SIAM Ser. Optim., SIAM, Philadelphia, 2006.
- [4] J.-F. AUJOL, *Some first-order algorithms for total variation based image restoration*, J. Math. Imaging Vision, 34 (2009), pp. 307–327.
- [5] 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.
- [6] A. BECK AND M. TEBoulLE, *Fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM J. Imaging Sci., 2 (2009), pp. 183–202.
- [7] A. BERMÚDEZ AND C. MORENO, *Duality methods for solving variational inequalities*, Comput. Math. Appl., 7 (1981), pp. 43–58.
- [8] M. BERTALMIO, L. VESE, G. SAPIRO, AND S. OSHER, *Simultaneous structure and texture image inpainting*, IEEE Trans. Image Process., 12 (2003), pp. 882–889.
- [9] 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.
- [10] X. BRESSON AND T. F. CHAN, *Fast dual minimization of the vectorial total variation norm and applications to color image processing*, Inverse Prob. Imaging, 2 (2008), pp. 455–484.
- [11] A. CHAMBOLLE, *Partial differential equations and image processing*, in Proceedings of the ICIP, 1994, pp. 16–20.
- [12] A. CHAMBOLLE, *An algorithm for total variation minimization and applications*, Math. Imaging Vision, 20 (2004), pp. 89–97.
- [13] A. CHAMBOLLE, *Total variation minimization and a class of binary MRF models*, in Proceedings of the EMMCVPR, 2005, pp. 136–152.
- [14] A. CHAMBOLLE, V. CASELLES, D. CREMERS, M. NOVAGA, AND T. POCK, *An introduction to total variation for image analysis*, in Theoretical Foundations and Numerical Methods for Sparse Recovery, De Gruyter, Berlin, 2010.
- [15] A. CHAMBOLLE, D. CREMERS, AND T. POCK, *A convex approach for computing minimal partitions*, Technical report TR-2008-05, Department of Computer Science, University of Bonn, Bonn, Germany, 2008.
- [16] A. CHAMBOLLE AND P.-L. LIONS, *Image recovery via total variation minimization and related problems*, Numer. Math., 76 (1997), pp. 167–188.
- [17] A. CHAMBOLLE AND T. POCK, *A first-order primal-dual algorithm for convex problems with applications to imaging*, J. Math. Imaging Vis., 40 (2011), pp. 120–145.
- [18] T. CHAN, S. ESEDOĞLU, AND M. NIKOLOVA, *Algorithms for finding global minimizers of image segmentation and denoising models*, SIAM J. Appl. Math., 66 (2006), pp. 1632–1648.
- [19] T. F. CHAN, G. H. GOLUB, AND P. MULET, *A nonlinear primal-dual method for total variation-based image restoration*, SIAM J. Sci. Comput., 20 (1999), pp. 1964–1977.
- [20] S. DI ZENZO, *A note on the gradient of a multi-image*, CVGIP, 33 (1986), pp. 116–125.
- [21] D. DOBSON AND F. SANTOSA, *Recovery of blocky images from noisy and blurred data*, SIAM J. Appl. Math., 56 (1996), pp. 1181–1198.
- [22] V. DUVAL, J.-F. AUJOL, AND L. VESE, *Projected gradient based color image decomposition*, in Scale Space and Variational Methods in Computer Vision, Lecture Notes in Comput. Sci. 5567, Springer, Berlin, 2009, pp. 295–306.
- [23] H. FEDERER, *Geometric Measure Theory*, Classics Math. 153, Springer, New York, 1996.
- [24] B. GOLDLUECKE AND D. CREMERS, *An approach to vectorial total variation based on geometric measure theory*, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2010.
- [25] R. KIMMEL, *A natural norm for color processing*, in Proceedings of the ACCV, 1998.
- [26] R. KIMMEL, R. MALLADI, AND N. SOCHEN, *Images as embedded maps and minimal surfaces: Movies, color, texture and volumetric medical images*, Internat. J. Comput. Vision, 39 (2000), pp. 111–129.

- [27] K. KOLEV, M. KLODT, T. BROX, AND D. CREMERS, *Continuous global optimization in multiview 3d reconstruction*, Internat. J. Comput. Vision, (2009).
- [28] A. MARQUINA AND S. OSHER, *Image super-resolution by TV-regularization and Bregman iteration*, J. Sci. Comput., 37 (2008), pp. 367–382.
- [29] L. MIRSKY, *Symmetric gauge functions and unitarily invariant norms*, Quart. J. Math. Oxford, 2 (1960), pp. 50–59.
- [30] Y. NIEVERGELT, *Schmidt-Mirsky matrix approximation with linearly constrained singular values*, Linear Algebra Appl., 261 (1997), pp. 207–219.
- [31] T. POCK, D. CREMERS, H. BISCHOF, AND A. CHAMBOLLE, *An algorithm for minimizing the piecewise smooth Mumford-Shah functional*, in Proceedings of the IEEE International Conference on Computer Vision, Kyoto, Japan, 2009.
- [32] B. RECHT, M. FAZEL, AND P. PARRILO, *Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization*, SIAM Rev., 52 (2010), pp. 471–501.
- [33] T. ROCKAFELLAR, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.
- [34] L. I. RUDIN, S. OSHER, AND E. FATEMI, *Nonlinear total variation based noise removal algorithms*, Phys. D, 60 (1992), pp. 259–268.
- [35] G. SAPIRO, *Vector-valued active contours*, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1996, pp. 680–685.
- [36] G. SAPIRO, *Color snakes*, Comput. Vision Image Understanding, 68 (1997), pp. 247–253.
- [37] G. SAPIRO AND D. L. RINGACH, *Anisotropic diffusion of multi-valued images with applications to color filtering*, IEEE Trans. Image Process., 5 (1996), pp. 1582–1586.
- [38] N. SOCHEN, R. KIMMEL, AND R. MALLADI, *A general framework for low level vision*, IEEE Trans. Image Process., 7 (1998), pp. 310–318.
- [39] D. TSCHUMPERLÉ AND R. DERICHE, *Constrained and unconstrained PDEs for vector image restoration*, in Proceedings of the Scandinavian Conference on Image Analysis, 2001.
- [40] J. WEICKERT, *Coherence-enhancing diffusion of colour images*, Image Vision Comput., 17 (1999), pp. 201–212.
- [41] C. ZACH, T. POCK, AND H. BISCHOF, *A duality based approach for realtime TV – L^1 optical flow*, in Pattern Recognition, Proceedings of DAGM, 2007, pp. 214–223.