Color TV: Total Variation Methods for Restoration of Vector-Valued Images

Peter Blomgren and Tony F. Chan, *Member, IEEE*

***Abstract*— We propose a new definition of the total variation norm for vector-valued functions that can be applied to restore color and other vector-valued images. The new TV norm has the desirable properties of 1) not penalizing discontinuities (edges) in the image, 2) being rotationally invariant in the image space, and 3) reducing to the usual TV norm in the scalar case. Some numerical experiments on denoising simple color images in red–green–blue (RGB) color space are presented.**

***Index Terms*— Color images, image restoration, multispectral image, total variation.**

I. Introduction

SEVERAL reconstruction methods based on the total variation (TV) norm introduced in [1] have been proposed and studied for restoration of intensity images; see also [2]–[5].

Since these methods have been successful in reducing noise and blur without smearing sharp edges for intensity images, it is natural to extend the TV norm to handle color and other vector valued images.

In applications where images are to be viewed by, or interpreted for, a human, color is an integral factor. Hence, color processing is important for recognition, segmentation, etc. Also, intensity-based processing fails to detect isoluminance edges, i.e., edges where there is a "jump" in color, but not in intensity. Incidentally, humans are not very good at detecting such edges.

Any attempt to extend the scalar TV norm to the vector-valued case should at least preserve two of the basic advantages of TV; namely, i) not penalizing against edges and ii) being rotationally invariant in image space. Moreover, it is also desirable to have the extension reduce to the usual TV norm in the scalar case. There have been several attempts to extend TV related restoration techniques and edge detection to vector-valued images; see [6]–[10]. Most of these, however, do not satisfy all of the above criteria. Our approach is most closely related to that of Sapiro [9]–[11], who proposed an anisotropic diffusion model for vector valued images.

In Section II, we extend the TV norm to vector valued functions. Some numerical results are presented in Section III. We compare our approach to that of Sapiro in Section IV, and in Section V we compare our definition to some other possible extensions.

Manuscript received August 13, 1996; revised March 10, 1997. This work was supported by the ONR under Contract ONR N00017-96-1-0277, and by the NSF under Grant MS-9626755. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Guillermo Sapiro.

The authors are with the Department of Mathematics, University of California, Los Angeles, CA 90095-1555 USA (e-mail: blomgren@math.ucla.edu, chan@math.ucla.edu.).

Publisher Item Identifier S 1057-7149(98)01743-6.

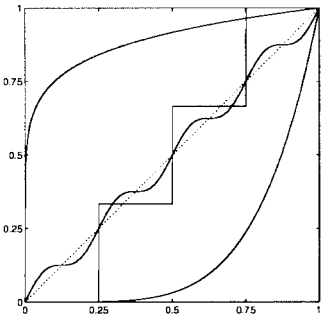

Fig. 1. Functions with the same total variation in one dimension.

II. The Multidimensional Total Variation Norm

We view the restoration problem in terms of solving the nonlinear optimization problem

$$\min_{\mathbf{\Phi}\in \mathrm{BV}(\Omega)} \|\mathbf{\Phi}\|_{\widetilde{\mathrm{TV}}} + \frac{\lambda}{2}\|\mathbf{\Phi}-\mathbf{\Phi}^0\|_2^2 \tag{1}$$

where $\mathbf{\Phi}^0$ is a given noisy image, λ a Lagrange multiplier associated with the noise level, $\|\mathbf{\Phi}\|_2^2 = \sum_{i=1}^m \|\Phi^i\|_2^2$, and $\|\cdot\|_{\widetilde{\mathrm{TV}}}$ an appropriate definition of the norm.

A. Important Properties of the 1-D TV-Norm

Before stating our definition of the multidimensional $\mathrm{TV}_{n,m}$ norm, we first review the definition of the one-dimensional (1-D) $\mathrm{TV}_{n,1}$ norm.

Definition 1—The 1-D $\mathrm{TV}_{n,1}(\Phi)$ Norm: The $\mathrm{TV}_{n,1}(\Phi)$ norm for scalar valued functions Φ: $\mathbb{R}^n \rightarrow \mathbb{R}$ is defined (see [1]) by

$$\mathrm{TV}_{n,1}(\Phi) \stackrel{\mathrm{def}}{=} \int_\Omega |\nabla \Phi|\, d\mathbf{x}, \quad \Omega \subset \mathbb{R}^n. \tag{2}$$

The TV norm satisfies the following two important properties.

Property 1—TV of Monotone Functions in 1-D, $n = 1$: For all functions satisfying

$$\begin{cases} \Phi \in \mathrm{BV}([a,b]) \\ \Phi(a) = \Phi_a,\ \Phi(b) = \Phi_b \\ \Phi(x) \text{ monotone in } [a,b]. \end{cases} \tag{3}$$

$\mathrm{TV}_{1,1}(\Phi) = |\Phi_a - \Phi_b|$; see Fig. 1.

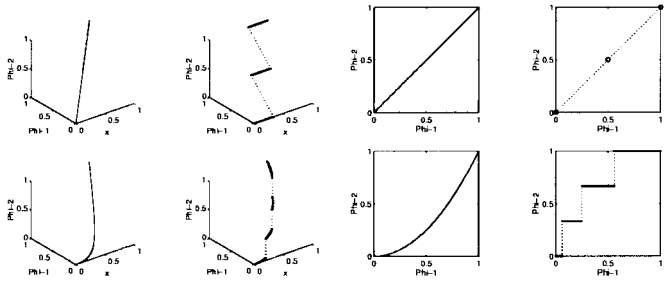

Fig. 2. Functions with the same total variation (left), and their images in $\mathbf{\Phi}$-space (right).

The property follows from (2), since the right hand side is exactly integrable if Φ is monotone. $BV([a, b])$ contains functions with bounded variation on $[a, b]$, i.e., $\int_a^b |f_x(x)|\,dx < \infty$. Discontinuous functions are included in this class.

When we are solving the nonlinear minimization problem (1), with $\|\Phi\|_{\widetilde{\mathrm{TV}}} = \mathrm{TV}_{n,1}(\Phi)$, no monotone function is preferred over another, so edges will not be smoothed. However, the contrast will be reduced. Even though the norm preserves edges, it is not biased in favor of edges. It will not introduce artificial jumps in the function, although a staircasing effect my occur in the presence of noise.

Property 2—Rotational Invariance: The $\mathrm{TV}_{n,1}$ norm is rotationally invariant in image space.

B. Definition of the TV-Norm for Vector-Valued Functions

We want an extension of the TV norm that is rotationally invariant, allows discontinuities, and satisfies a modification of Property 1: The norm should be invariant for functions with monotone components, and fixed end points. With this in mind, we define the TV norm for vector-valued functions:

Definition 2—The Multidimensional $TV_{n,m}(\mathbf{Phi})$ Norm: For any function $\mathbf{\Phi}$: $\mathbb{R}^n \to \mathbb{R}^m$, we define the multidimensional TV norm to be

$$\mathrm{TV}_{n,m}(\mathbf{\Phi}) \stackrel{\mathrm{def}}{=} \sqrt{\sum_{i=1}^{m} [\mathrm{TV}_{n,1}(\Phi^i)]^2}. \tag{4}$$

The invariance for functions with monotone components is easily verified; the TV of such a function is the *Euclidean distance* between a) $\mathbf{\Phi}$ and b) $\mathbf{\Phi}$, see Fig. 2.

As in the gray scale case, this definition of $\mathrm{TV}_{n,m}(\mathbf{\Phi})$ allows discontinuities in each channel. It preserves edges in the sense that no component-monotone function is preferred over another in the minimization problem (1), with $\|\mathbf{\Phi}\|_{\widetilde{\mathrm{TV}}} = \mathrm{TV}_{n,m}(\mathbf{\Phi})$.

We notice that if $m = 1$, $\mathrm{TV}_{n,m}(\mathbf{\Phi}) \equiv \mathrm{TV}_{n,1}(\Phi^1)$, i.e., we recover the TV norm for a scalar valued function.

III. Numerical Experiments

A. Noise Reduction Using $TV_{n,m}(\Phi)$

Given a noisy image $\mathbf{\Phi}^0$, we are interested in minimizing (1), with $\|\mathbf{\Phi}\|_{\widetilde{\mathrm{TV}}} = \mathrm{TV}_{n,m}(\mathbf{\Phi})$. The corresponding

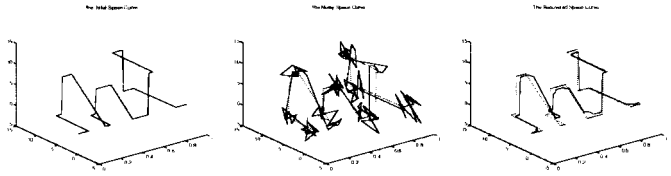

Fig. 3. Initial, noisy (SNR = 4.62 dB), and recovered space curves. Notice how the edges are recovered.

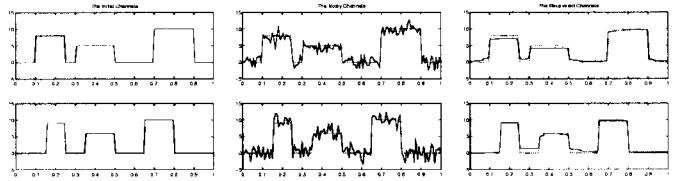

Fig. 4. Channel separation for initial, noisy, and recovered data. (Same data as in Fig. 3.).

Euler–Lagrange equations are

$$\frac{\mathrm{TV}_{n,1}(\Phi_i)}{\mathrm{TV}_{n,m}(\mathbf{\Phi})}\,\nabla \circ \left(\frac{\nabla \Phi_i}{\|\nabla \Phi_i\|}\right) - \lambda\big(\Phi_i - \Phi_i^0\big) = 0. \tag{5}$$

The solutions were computed using an explicit time marching scheme, a slight modification of the one introduced in [1], applied to the following system of equations:

$$\frac{\partial \Phi_i}{\partial t} = \frac{\mathrm{TV}_{n,1}(\Phi_i)}{\mathrm{TV}_{n,m}(\mathbf{\Phi})}\,\nabla \circ \left(\frac{\nabla \Phi_i}{\sqrt{\beta + \|\nabla \Phi_i\|^2}}\right) - \lambda\big(\Phi_i - \Phi_i^0\big) \tag{6}$$

where β is a small regularization parameter, introduced to avoid division by zero. In all the computations, λ is determined by a gradient projection method, cf. [1], [12].

The numerical experiments were all performed in Euclidean spaces, which for color images represent the linear red–green–blue (RGB) space. This is not an optimal space for perceptual uniformity, but the results are of interest anyway. For discussion about alternative color spaces, see [13].

Example 1—Denoising of a Space Curve (Figs. 3 and 4): We define a curve (with several discontinuities) embedded in $\mathbb{R}^3$ by defining a mapping $\mathbf{\Phi}$: $[0, 1] \subset \mathbb{R} \to \mathbb{R}^2$. We add Gaussian noise (SNR = 4.62 dB) to the data, and apply the TV-denoising to the noisy curve. The time step was $dt = 7.75 \cdot 10^{-6}$. The process was stopped after 7290 iterations when the l_2-norm of the correction was less than $1.5 \cdot 10^{-5}$.

As can be seen, the sharp edges are indeed recovered, and the reconstruction is remarkably good.

Example 2—Denoising of a 1-D Color Image (Figs. 5 and 6): A 1-D image was created in RGB space, i.e., $\mathbf{\Phi}$: $\mathbb{R}^1 \to \mathbb{R}^3$ where each channel represents a color intensity. Gaussian noise (SNR = −3.01 dB) was added.

The images are visualized by extending the one dimensional lines into two dimensional strips. Again the reconstruction is good, with sharp edges.

Example 3—Denoising of a 2-D Color Image (Fig. 7): We created a two-dimensional (2-D) test image in RGB space, i.e., $\mathbf{\Phi}$: $\mathbb{R}^2 \to \mathbb{R}^3$; Gaussian noise was added (SNR = 3.98 dB), and we ran the explicit time marching ($dt = 1.0 \cdot 10^{-6}$) with

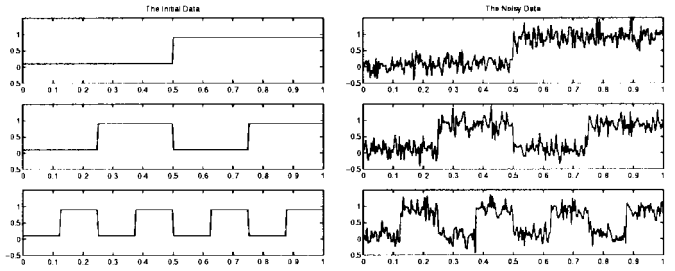

Fig. 5. Channel separation for the initial and noisy (SNR $= -3.01$ dB) images.

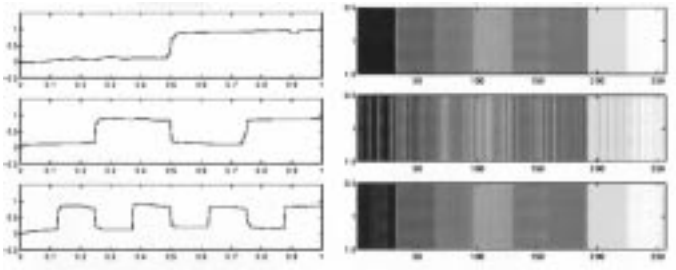

Fig. 6. Channel separation for the recovered image, and color visualization for the original, noisy, and recovered images. Left: recovered data. Right: Initial, noisy, and recovered color images. (The color version of this figure is available at www.math.ucla.edu/-blomgren/camreports.shtml.)

the regularizing parameter $\beta = 1.0 \cdot 10^{-9}$. As in the 1-D case, the reconstruction is good, with sharp edges retained in the correct locations.

Example 4—Denoising of the Lena Image (Fig. 8): We denoise (SNR $= 3.01$ dB) a 128×128 color Lena image. We used the parameters $dt = 10^{-6}$, and $\beta = 10^{-10}$.

We observe, in all examples, that the reconstruction does not smear edges. The quality increases as the regularization parameter β and the time step dt shrink. It is well known that the explicit time marching approach we are using here has less than optimal convergence properties. It is straightforward to implement more efficient numerical methods, e.g., fixed point (cf., Vogel [5], [14]) and primal-dual (cf., Chan *et al.* [2]).

IV. Other Approaches

Other approaches to vector-valued image processing, e.g., anisotropic diffusion [11], [15]–[19], edge detection and segmentation [7], [8], [15], [17], [20], [21] as well as segmentation methods related to level-set methods [9], [10], [22]–[24], can be found in the literature. Here we will briefly describe the general anisotropic diffusion approach, and the Sapiro–Ringach approach in particular [9]–[11].

Notice that solving the nonlinear system of equations (6) is a form of anisotropic diffusion. Before comparing $\mathrm{TV}_{n,m}$ and the anisotropic diffusion approach, we explore the relationship between $\mathrm{TV}_{n,m}$ and the simplest extension of the TV approach to the vector valued case—channel-by-channel TV restoration.

A. $\mathrm{TV}_{n,m}$ and Channel-by-Channel TV

A common belief in the image processing community is that when processing vector-valued images, there should be a coupling between the channels. In general, however, it is not clear what the correct coupling is. Here we describe the nature of the coupling of the channels in the $\mathrm{TV}_{n,m}$ norm and give an indication of how the performance of this norm is different from direct application of $\mathrm{TV}_{n,1}$ separately to the channels.

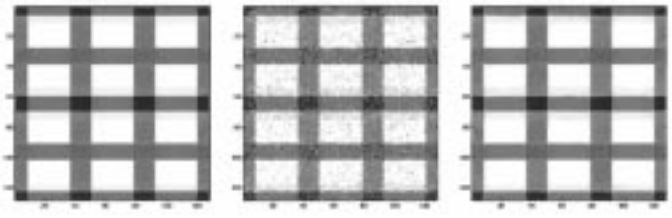

Fig. 7. Initial, noisy, and denoised color images. Left: initial color image. Middle: noisy color image. Right: Denoised color image. (The color version of this figure is available at www.math.ucla.edu/-blomgren/camreports.shtml.)

Fig. 8. Left to right: noisy (SNR $= 3.01$ dB) image, recovered image, and noiseless image. (The color version of this figure is available at www.math.ucla.edu/-blomgren/camreports.shtml.)

Consider the two denoising approaches. First, in *channel-by-channel TV*, apply the minimization (1), with $\|\Phi^i\|_{\widetilde{\mathrm{TV}}} = \mathrm{TV}_{n,1}(\Phi^i)$, separately to each channel. Second, in *color TV*, apply the minimization (1), with $\|\mathbf{\Phi}\|_{\widetilde{\mathrm{TV}}} = \mathrm{TV}_{n,m}(\mathbf{\Phi})$. The corresponding Euler–Lagrange equations are as follows.

Channel-by-Channel TV:

$$\nabla \circ \left(\frac{\nabla \Phi_i}{\|\nabla \Phi_i\|} \right) - \lambda \left(\Phi_i - \Phi_i^0 \right) = 0. \tag{7}$$

Color TV:

$$\frac{\mathrm{TV}_{n,1}(\Phi_i)}{\mathrm{TV}_{n,m}(\mathbf{\Phi})} \nabla \circ \left(\frac{\nabla \Phi_i}{\|\nabla \Phi_i\|} \right) - \lambda \left(\Phi_i - \Phi_i^0 \right) = 0. \tag{8}$$

Comparing (7) and (8), we see that the $\mathrm{TV}_{n,m}$ coupling takes the form of a global channel-wise scaling of the diffusion coefficient by the factor $\mathrm{TV}_{n,1}(\Phi_i)/\mathrm{TV}_{n,m}(\mathbf{\Phi})$ Thus, a channel with larger TV will be smoothed more than a channel with smaller TV. Consider the example, with $m = 2$, $n = 1$, as depicted in Fig. 9. For the same λ, the channel-by-channel TV almost completely wipes out the weaker of the channels, whereas the $\mathrm{TV}_{n,m}$ regularization maintains a balance of how much each channel is smoothed. Fig. 10 shows reconstruction from a noisy signal. Notice how the TVn, m approach retains more detail in the weaker channel.

In cases where we have separate noise measures for the separate channels (multiple constraints), the channel-by-channel approach may be more successful. For a discussion of how the choice of regularization parameter λ affects the scale of TV smoothing, see Strong–Chan [4].

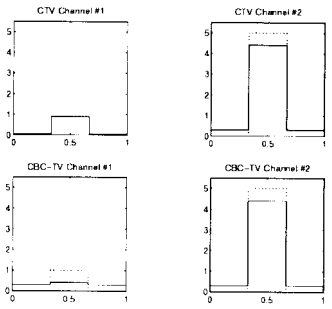

Fig. 9. Application of the denoising (TV smoothing) to a simple model problem where the two channels have different strength. Here $\lambda = 10$. The upper two pictures show the result of the color TV regularization, and the lower pictures show the result of channel-by-channel application of the TV norm. The dotted line shows the initial data.

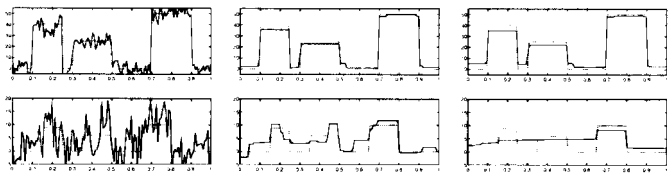

Fig. 10. Comparison of color TV, and channel-by-channel TV. Regularization parameter $\lambda = 2$. Left to right: Initial signals, color TV reconstruction, and channel-by-channel reconstruction. Notice that channel 1 is about five times as strong as channel 2.

B. Anisotropic Diffusion Based on Riemannian Geometry

The Riemannian geometry framework for edge detection in vector-valued images was first suggested in [21]; see also [7], [8], [19], [20]. Given a multidimensional image $\mathbf{\Phi}$, the *first fundamental form* [6], [25]

$$d\mathbf{\Phi}^2 = \sum_{j=1}^{2}\sum_{i=1}^{2} \frac{\partial \mathbf{\Phi}}{\partial x_i} \circ \frac{\partial \mathbf{\Phi}}{\partial x_j}\, dx_i\, dx_j = \begin{bmatrix} dx_1 \\ dx_2 \end{bmatrix}^T \underbrace{\begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix}}_{G} \begin{bmatrix} dx_1 \\ dx_2 \end{bmatrix} \tag{9}$$

is the squared norm of the arc element $d\mathbf{\Phi}$. Here

$$g_{ij} = \frac{\partial \mathbf{\Phi}}{\partial x_i} \circ \frac{\partial \mathbf{\Phi}}{\partial x_j} = \sum_{k=1}^{m} \frac{\partial \Phi^k}{\partial x_i} \frac{\partial \Phi^k}{\partial x_j}. \tag{10}$$

The idea is to preserve edges by smoothing in the direction of minimal change, ξ_-, parallel to the edge. The anisotropic diffusion model is described by the evolution equation

$$\frac{\partial \mathbf{\Phi}}{\partial t} = g(\lambda^+, \lambda^-) \frac{\partial^2 \mathbf{\Phi}}{\partial \xi_-^2} \tag{11}$$

where, for $n = 2$

$$\lambda^{\pm} = \frac{g_{11} + g_{22} \pm \sqrt{(g_{11} - g_{22})^2 + 4g_{12}^2}}{2} \tag{12}$$

are the largest, resp., smallest eigenvalues of the matrix G, and ξ_- the *direction of minimal change*, the eigenvector corresponding to λ^-

$$\xi_+ = \frac{1}{2} \arctan\left(\frac{2g_{12}}{g_{11} - g_{22}}\right),\ \xi_- \perp \xi_+. \tag{13}$$

$g(\cdot)$ is any decreasing function of either $(\lambda^+ - \lambda^-)$ or λ^+/λ^-, used to penalize against smoothing edges.

It can be shown that in the gray scale case, $\lambda^- = 0$, and $\partial^2 \mathbf{\Phi}/\partial \xi_-^2 = \nabla \circ (\nabla \Phi / \|\nabla \Phi\|) = 0$ is exactly the Euler–Lagrange equations for the TV norm. In addition the TV norm can be written as

$$\mathrm{TV}_{n,1}(\Phi) \equiv \int_{\Omega \in \mathbb{R}^n} \sqrt{\lambda_1^+}\, d\mathbf{x}. \tag{14}$$

Thus, in the gray-scale case, this approach is strongly related to the TV regularization.

An alternative, but equivalent, formulation is given by Chambolle [26], who suggests using the diffusion equation (11) with the directions

$$\xi^+ = \arg \max_{|\xi|=1} \sum_{i=1}^{3} (\nabla \Phi^i, \xi)^2 \tag{15}$$

and $\xi^- \perp \xi^+$. It is straight forward to show that (9) and (15) define the same directions, $\xi^{\pm}$.

C. Sapiro's Generalization of the TV Norm

In [9], Sapiro suggests, but does not explore in detail, the general extension of the TV norm based on Riemannian geometry

$$\|\mathbf{\Phi}\|_{\widehat{\mathrm{TV}}} = \int_{\Omega \in \mathbb{R}^2} f(\lambda^+, \lambda^-)\, d\mathbf{x}. \tag{16}$$

However, Sapiro did not recommend a specific f. To define a meaningful norm, $f(\cdot)$ should be a nondecreasing function of $\lambda_{\pm}$. One natural choice is $f(\cdot) = \sqrt{\lambda^+ + \lambda^-}$, i.e., the square root of the trace of the matrix G; see Section V. This choice is equivalent to the TV norm in the scalar case $(\lambda^- \equiv 0)$. In [9] the *anisotropic diffusion model*, with $g() = \psi(\lambda^+ - \lambda^-)$, $\psi()$ nondecreasing, in (11), was suggested. However, Sapiro did not use $f(\cdot) = \psi(\lambda^+ - \lambda^-)$ to define a norm. Our experiments indicate that norms of this form do not work well within the variational framework of (1), therefore we do not discuss these norms in this paper. We restrict our attention to the norm generated by $f(\cdot) = \sqrt{\lambda^+ + \lambda^-}$. We shall show, however, that this definition of the norm has a tendency to smooth colors.

We can give an intuitive explanation for the smoothing caused by this definition of the norm: a function $f(\lambda^+, \lambda^-)$, nondecreasing in both arguments, is essentially $f(d\mathbf{\Phi}^2)$ since f yields a different metric G' which is a rescaling of G in (9). Since $d\mathbf{\Phi}$ is the arc element in $\mathbf{\Phi}$-space, the norm $\int f(d\mathbf{\Phi}^2)\, d\mathbf{x}$ is a measure of the arc length. Therefore minimizing this norm will give a preference to curves with shorter arc length in $\mathbf{\Phi}$-space, which leads to a smoothing of edges. The key difference between our approach and the Riemannian geometry approach is that our definition is not concerned with the arc length. We now give some simple examples to illustrate this difference.

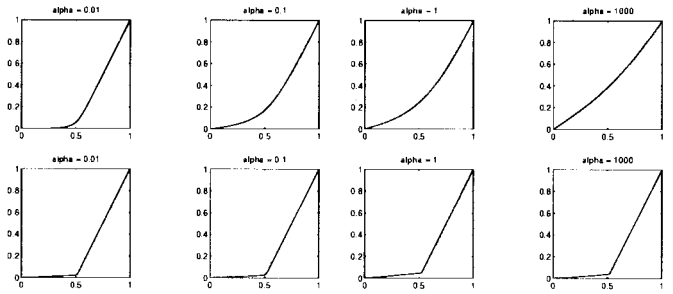

Fig. 11. Restorations visualized in $\mathbf{\Phi}$-space for some values of $\lambda = 1/\alpha$. (SNR $= \infty$, no noise.) The upper graphs correspond to the $\sqrt{\lambda^+ + \lambda^-}$ norm, and the lower to the $\mathrm{TV}_{n,m}$ norm, the small "lift" from the exact solution for the $\mathrm{TV}_{n,m}$ norm depends on the numerical regularization parameter β, here $\beta = 10^{-8}$.

Example 1—Let $\mathbf{\Phi}$: $[0, 1] \subset \mathbb{R} \to \mathbb{R}^2$ *Be Monotone in Its Components:*

$$\Phi^1(x) = x, \qquad \Phi^2(x) = \begin{cases} 0 & x < \gamma \\ \frac{1}{1-\gamma}(x-\gamma) & x > \gamma. \end{cases}$$

We get

$$\int_\Omega \sqrt{\lambda^+ + \lambda^-}\, d\mathbf{x} = \gamma + \sqrt{(1-\gamma)^2 + 1} = N(\gamma)$$

and we have

$$\lim_{\gamma \searrow 0} N(\gamma) = \sqrt{2}, \ \lim_{\gamma \nearrow 1} N(\gamma) = 2.$$

Thus, this norm "prefers" solutions with a smaller γ, corresponding to small gradients in $\mathbf{\Phi}$. Since the components are monotone, our extension of the TV norm would yield the same value independent of γ.

We next compare the numerical restorations using the $\mathrm{TV}_{n,m}$ and $\sqrt{\lambda^+ + \lambda^-}$ norms. We choose the functions $\Phi^{1,2}(x)$ as described, with $\gamma = 0.5$. We solve the minimization problem (1) with the additional conditions $\Phi^i(0) = 0$, $\Phi^i(1) = 1$. Fig. 11 shows the restorations visualized in $\mathbf{\Phi}$-space, and Fig. 12 shows the visualization in RGB space, where $(R, G, B) = (\Phi^1, \Phi^2, \Phi^2)$. We notice that, as expected, the $\sqrt{\lambda^+ + \lambda^-}$ norm has a tendency to smooth the kink in the curve much more than the $\mathrm{TV}_{n,m}$ norm does, especially for small λ (more regularization).

Example 2—Consider a Discontinuous $\mathbf{\Phi}$*:*

$$\phi^1(x) = \phi^2(x) = \begin{cases} 0 & x < 1/2 \\ 1 & x > 1/2 \end{cases}$$

which has the image $\{(0, 0), (1, 1)\}$ in $\mathbf{\Phi}$-space. The $\mathrm{TV}_{n,m}$ norm and the $\sqrt{\lambda^+ + \lambda^-}$ norm coincide with value $\sqrt{2}$.

Now, consider the perturbed problem

$$\phi^1(x) = \begin{cases} 0 & x < 1/2 - \epsilon \\ 1 & x > 1/2 - \epsilon, \end{cases}, \quad \phi^2(x) = \begin{cases} 0 & x < 1/2 \\ 1 & x > 1/2 \end{cases}$$

which has the image $\{(0, 0), (1, 0), (1, 1)\}$ in $\mathbf{\Phi}$-space. The $\mathrm{TV}_{n,m}$ norm takes on the value $\sqrt{2}$ and the $\sqrt{\lambda^+ + \lambda^-}$ norm the value 2.

We observe that the $\sqrt{\lambda^+ + \lambda^-}$ norm is biased in favor of vector valued functions whose channels have similar transitions, i.e., $\|\nabla \Phi^i\| = \|\nabla \Phi^j\|$. This leads to redistribution on energy across the channels, i.e., color smearing.

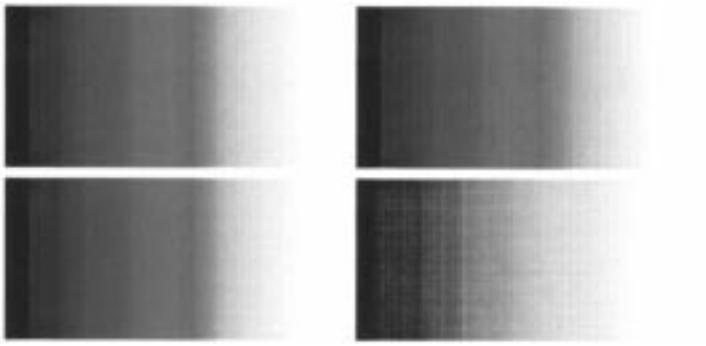

Fig. 12. Restorations visualized in RGB color space, the upper two pictures show the TV (left) and $\sqrt{\lambda^+ + \lambda^-}$ (right) reconstructions for $\lambda = 100$, and the lower two pictures show the reconstructions for $\lambda = 0.001$. We observe a clear shift toward a gray-scale image for the $\sqrt{\lambda^+ + \lambda^-}$ norm when $\lambda = 0.001$. (The color version of this figure is available at www.math.ucla.edu/-blomgren/camreports.shtml.)

V. A General Framework for Vector Norms

The definition of the $\mathrm{TV}_{n,m}$-norm as the l^2-norm of the TV-norm of the separate channels may seem arbitrary. In this section we will show that our definition is the most natural within a quite general class of norms, derived from compositions of p-norms. Thus, alternative vector-TV extensions should be sought elsewhere, e.g., in the $\int f(\lambda^+, \lambda^-)$ framework of Sapiro.

First, we write the $\mathrm{TV}_{n,m}$-norm as a composition of l^2, and L^1-norms

$$\mathrm{TV}_{n,m}(\mathbf{\Phi}) \stackrel{\text{def}}{=} l^2(i) \circ L^1(\Omega) \circ l^2(j) \left(\frac{\partial \Phi^i}{\partial x_j} \right).$$

In this framework, the six permutations of compositions of the norms $l^{p_m}(i) \circ L^{p_\Omega}(\Omega) \circ l^{p_n}(j)$ define six three-parameter families of norms. Composition of any two p-norms commute if an only if the corresponding p's are equal.

Using this general characterization, we argue as follows: a necessary, but not sufficient, condition for rotational invariance in physical space is $p_n = 2$; for rotational invariance in image space, we need $p_m = 2$;[1] for any choice $p_\Omega \neq 1$ there is no hope of obtaining a norm which is unbiased with respect to discontinuous functions. That leaves us with the following four possibilities:

1) $[l^2(i) \circ L^1(\Omega) \circ l^2(j)](\mathbf{\Phi})$ Blomgren–Chan, $\mathrm{TV}_{n,m}$
2) $[L^1(\Omega) \circ l^2(i) \circ l^2(j)](\mathbf{\Phi})$ Sapiro–$\sqrt{\lambda^+ + \lambda^-}$ [9]
3) $[l^2(j) \circ L^1(\Omega) \circ l^2(i)](\mathbf{\Phi})$
4) $[l^2(i) \circ l^2(j) \circ L^1(\Omega)](\mathbf{\Phi})$.

The choice $f(\lambda^+, \lambda^-) = \sqrt{\lambda^+ + \lambda^-}$ is natural in the sense that it fits into this general framework. Other choices of $f(\lambda^+, \lambda^-)$ do not fit. When $n > 2$, $f(\lambda^1, \lambda^2, \cdots, \lambda^n) = \sqrt{\lambda^1 + \lambda^2 + \cdots + \lambda^n} = \sqrt{\mathrm{trace}(G)}$ is the natural choice. Definitions 3 and 4 are not rotationally invariant in physical space.

Notice that when $m = 1$, the first two reduce to $[L^1(\Omega) \circ l^2(j)](\mathbf{\Phi})$, which is the "classical" gray scale TV-norm, and the last two reduce to $[l^2(j) \circ L^1(\Omega)](\mathbf{\Phi}) \equiv$

[1] This property is not crucial, since vision, and hence color-spaces are nonlinear; i.e, rotational invariance in color space has no meaning.

TABLE I
PROPERTIES OF DIFFERENT NORMS. ROTATIONAL INVARIANCE REFERS TO IMAGE SPACE. SINCE ALL THE COMPOSITIONS INCLUDE $L^1(\Omega)$, THEY ARE UNBIASED WRT DISCONTINUITIES. THE $\sqrt{\lambda^+ + \lambda^-}$ AND $[l^2(j) \circ L^1(\Omega) \circ l^2(i)]$ NORMS HAVE A TENDENCY TO SMEAR COLORS (BIAS AGAINST COLOR IMAGES) AS DISCUSSED IN EXAMPLES 1 AND 2 IN SECTION IV-C

Norm	Reduces to 1D TV	Rot. Inv.	Allow Edge	Unbiased wrt. Color
$\mathrm{TV}_{n,m}$	Yes	Yes	Yes	Yes
$\sqrt{\lambda^+ + \lambda^-}$	Yes	Yes	Yes	No
$l^2(j) \circ L^1(\Omega) \circ l^2(i)$	No	No	Yes	No
$l^2(i) \circ l^2(j) \circ L^1(\Omega)$	No	No	Yes	Yes

$\sqrt{(\int_\Omega |\Phi_x|)^2 + (\int_\Omega |\Phi_y|)^2}$, which is an alternate definition for the gray scale TV-norm.

Our choice of definition is based on the two requirements that the norm should reduce to the gray scale TV-norm for $m = 1$ and monotonicity invariance for $n = 1$.

In this framework $[L^1(\Omega) \circ l^1(j)](\mathbf{\Phi})$, is the "taxi," or "Manhattan-distance" TV-norm of Li–Santosa [3].

In Table I, we summarize the properties of the different definitions of vector TV-norms.

VI. CONCLUDING REMARKS

We have introduced a new definition of the total variation norm, $\mathrm{TV}_{n,m}(\mathbf{\Phi})$, for vector-valued functions. This definition has a number of properties that may be desirable in applications: 1) it allows discontinuous functions—edges; 2) it is rotationally invariant in image space; and 3) it reduces to the classical TV norm in the scalar case.

We have compared the properties of the norm to other definitions, in particular to the approach of Sapiro [9]. By studying simple examples, i.e., reduction to one dimension in either physical, or color space, we have illustrated some differences between the norms.

A general framework for vector norms was introduced. We find two natural candidates: the $\mathrm{TV}_{n,m}$, and the $\sqrt{\lambda^+ + \lambda^-}$ norms. Of these two, $\mathrm{TV}_{n,m}$ does a better job of preserving color transitions. Many promising possible norm definitions fall outside this framework, and are not discussed in this paper.

The color versions of Figures 6–8 and 12 are available at www.math.ucla.edu/-blomgren/camreports.shtml.

REFERENCES

[1] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, pp. 259–268, 1992.

[2] T. F. Chan, G. Golub, and P. Mulet, "A nonlinear primal-dual method for TV-based image restoration," in *Proc. 12th Int. Conf. Analysis and Optimization of Systems: Images, Wavelets, and PDE's*, Paris, France, June 26–28, 1996, M. Berger *et al.*, Eds., no. 219, pp. 241–252.

[3] Y. Li and F. Santosa, "An affine scaling algorithm for minimizing total variation in image enhancement," Tech. Rep. CTC94TR201, Ctr. Theory Simulat. Sci. Eng., Cornell Univ., Ithaca, NY, 1994.

[4] D. M. Strong and T. F. Chan, "Relation of regularization parameter and scale in total variation based image denoising," in *Proc. IEEE Workshop on Mathematical Methods in Biomedical Image Analysis*, 1996.

[5] C. R. Vogel and M. E. Oman, "Iterative methods for total variation denoising," *SIAM J. Sci. Comput.*, vol. 17, Jan. 1996.

[6] A. Cumani, "Edge detection in multispectral images," *CVGIP: Graph. Models Image Processing*, vol. 53, pp. 40–51, 1991.

[7] H-C. Lee and D. R. Cok, "Detecting boundaries in a vector field," *IEEE Trans. Signal Processing*, vol. 39, pp. 1181–1194, 1991.

[8] R. Nevatia, "A color edge detector and its use in scene segmentation," *IEEE Trans. Syst., Man, Cybern.*, vol. 7, pp. 820–826, 1977.

[9] G. Sapiro, "Color snakes," Tech. Rep. HPL-95-113, Hewlett Packard Comput. Periph. Lab., Sept. 1995.

[10] ——, "Vector-valued active contours," in *Proc. Conf. Computer Vision and Pattern Recognition, IEEE Computer Society*, pp. 520–525, 1996.

[11] G. Sapiro and D. L. Ringach, "Anisotropic diffusion of multivalued images with applications to color filtering," *IEEE Trans. Image Processing*, vol. 5, pp. 1582–1586, Nov. 1996.

[12] J. G. Rosen, "The gradient projection methods for nonlinear programming, part II, nonlinear constraints," *J. Soc. Indust. Appl. Math.*, vol. 9, p. 514, 1961.

[13] G. Wyszecki and W. S. Styles, *Color Science: Concepts and Methods, Quantitative Data and Formulae*, 2nd ed. New York: Wiley, 1982.

[14] C. R. Vogel and M. E. Oman, "Fast total variation-based image reconstruction," in *Proc. ASME Symp. Inverse Problems*, 1995.

[15] L. Alvarez, P. L. Lions, and J. M. Morel, "Image selective smoothing and edge detection by nonlinear diffusion. II," *SIAM J. Numer. Anal.*, vol. 29, pp. 845–866, June 1992.

[16] L. Alvarez and L. Mazorra, "Signal and image restoration using shock filters and anisotropic diffusion," *SIAM J. Numer. Anal.*, vol. 31, pp. 590–605, Apr. 1994.

[17] F. Catté, P. L. Lions, J. M. Morel, and T. Coll, "Image selective smoothing and edge detection by nonlinear diffusion," *SIAM J. Numer. Anal.*, vol. 29, pp. 182–193, Feb. 1992.

[18] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. Pattern Anal. Machine Intell,*. vol. 12, pp. 629–639, 1990.

[19] R. Whitaker and G. Gerig, "Vector-valued diffusion," in *Computational Imaging and Vision: Geometry Driven Diffusion in Computer Vision*, B. M. ter Haar Romeny, Ed. Boston, MA: Kluwer, 1994.

[20] V. Torre and T. Poggio, "On edge detection," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-8, pp. 147–163, 1986.

[21] S. Di Zenzo, "A note on the gradient of a multi-image," *CVGIP*, vol. 33, pp. 116–125, 1986.

[22] F. Catté, F. Dibos, and G. Koepfler, "A morphological scheme for mean curvature motion and applications to anisotropic diffusion and motion of level sets," *SIAM J. Numer. Anal.*, vol. 32, pp. 1895–1909, Dec. 1995.

[23] B. Merriman, J. Bence, and S. Osher, "Diffusion generated motion by mean curvature," Tech. Rep. CAM 92-18, Dept. Math., Univ. Calif., Los Angeles, Apr. 1992.

[24] S. Osher and J. Sethian, "Fronts propagating with curvature dependent speed: Algorithms based on the Hamilton–Jacobi formulation," *J. Computat. Phys.*, vol. 79, pp. 12–49, 1988.

[25] E. Kreyszig, *Differential Geometry*. Toronto, Canada: Univ. Toronto Press, 1959.

[26] A. Chambolle, "Partial differential equations and image processing," in *Proc. IEEE Int. Conf. Image Processing*, Austin, TX, Nov. 1994.

Peter Blomgren received the M.Sc. degree in engineering physics from the Royal Institute of Technology, Stockholm, Sweden, in 1994. He is currently a Ph.D. candidate at the University of California, Los Angeles.

His research interests include PDE methods for image restoration and efficient numerical algorithms for image processing.

Tony F. Chan (M'98) received the B.S. degree in engineering and the M.S. degree in aerospace engineering, both in 1973, from the California Institute of Technology, Pasadena, and the Ph.D. degree in computer science from Stanford University, Stanford, CA, in 1978.

He is currently the Department Chair of the Department of Mathematics, University of California, Los Angeles, where he has been a Professor since 1986. His research interests include PDE methods for image processing, multigrid, and domain decomposition algorithms, iterative methods, Krylov subspace methods, and parallel algorithms.