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

Peter Blomgren
University of California, Los Angeles
Department of Mathematics
blomgren@math.ucla.edu

Tony F. Chan
University of California, Los Angeles
Department of Mathematics
chan@math.ucla.edu

May 23, 1996

Abstract

We propose a new definition of the total variation norm for vector valued functions which can be applied to restore color and other vector valued images. The new TV norm has the desirable properties of (*i*) not penalizing discontinuities (edges) in the image, (*ii*) rotationally invariant in the image space, and (*iii*) reduces to the usual TV norm in the scalar case. Some numerical experiments on denoising simple color images in RGB color space are presented.

1 Introduction

During gathering and transfer of image data some noise and blur is usually introduced into the image. Several reconstruction methods based on the total variation (TV) norm have been proposed and studied for intensity (gray scale) images, *see* [9, 14, 21, 26, 29].

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.

Why do we need color restoration? It can be argued that since color literally only is in the eye of the beholder, color image processing is not important for computer vision in terms of edge detection and targeting. However, in any application where the images are to be viewed by a human, color is an important factor; also, intensity based processing fails to detect iso-luminance edges, *i.e.* edges where there is a "jump" in color, but the intensity (or luminance) is constant.

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 penalize

*This work is supported by the ONR under contract ONR N00017-96-1-0277
†This paper is available in color postscript at `http://www.math.ucla.edu/applied/cam`

against edges, and *(ii)* be 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* [10, 12, 16, 22, 23]. Most of these, however, do not satisfy all of the above criteria. Our approach is most closely related to that of Sapiro's [22, 23, 24], who proposed an anisotropic diffusion model for vector valued images. We will compare our approach to that of Sapiro's in section 3.

In section 2, we extend the TV norm to vector valued functions, and we explain why we choose this extension of the norm. In section 4, we compare our definition to other possible extensions and approaches to the color image restoration problem.

In section 5, we make some remarks regarding the color space used in our numerical experiments, and how a different choice of color space may improve image reconstruction.

In section 6, we present some numerical results showing that denoising using the extended norm give us good reconstruction for *(i)* a space-curve in $\mathbb{R}^3$, *(ii)* a one-dimensional RGB color image, and *(iii)* two two-dimensional RGB color images.

2 The Multi Dimensional Total Variation Norm

2.1 Important Properties of the One Dimensional Total Variation Norm

Before stating our definition of the multidimensional $\mathrm{TV}_{\mathrm{n,m}}$ norm, we first review the definition of the one dimensional $\mathrm{TV}_{\mathrm{n,1}}$ norm, and its properties.

Definition 1 (The One Dimensional $\mathrm{TV}_{\mathrm{n,1}}(\Phi)$ Norm)
The $\mathrm{TV}_{\mathrm{n,1}}(\Phi)$ norm is defined (*see* [21]) for scalar valued functions $\Phi : \mathbb{R}^n \to \mathbb{R}$

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

One important property is most evident when $n = 1$:

Property 1 (Total Variation of Monotone Functions in One Dimension)
All 1D functions satisfying:

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

have the same total variation, $\mathrm{TV}_{1,1}(\Phi) = |\Phi_a - \Phi_b|$; *see* figure 1. $BV([a,b])$ is the class of functions with bounded variation on the interval $[a,b]$, *i.e.*

$\int_a^b |f_x(x)|\,dx < \infty$. Notice that discontinuous functions are included in this class.

This property follows from (1) since the right hand side is exactly integrable if Φ is monotone.

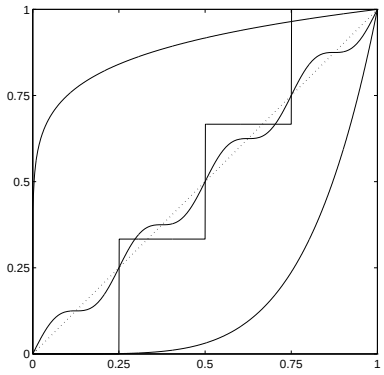

Figure 1: Functions with the same Total Variation in 1D.

Property 2 (Rotationally Invariant)
The $\mathrm{TV}_{\mathrm{n},1}$ norm is rotationally invariant in image space. This is clear, since the norm of the gradient $|\nabla\phi| = \|\nabla\phi\|_{l^2}$ is rotationally invariant.

This means that when we are solving the nonlinear minimization problem

$$\min_{\Phi\in \mathrm{BV}(\Omega)} \mathrm{TV}_{\mathrm{n},1}(\Phi) - \frac{\lambda}{2}\left\|\Phi - \Phi^0\right\|_2^2 \tag{2}$$

which arises in denoising, no monotone function is preferred over another, so edges will not be smoothed. Another important implication is that even though the norm preserves edges it is not overly biased in favor of edges, *i.e.* it will not introduce artificial jumps in the function (over-sharpening).

2.2 A New Definition of the Total Variation Norm for Vector Valued Functions

We are now ready to state our extension of the definition of the TV norm to the vector valued case.

Definition 2 (The Multi Dimensional $\mathrm{TV}_{\mathrm{n,m}}(\mathbf{\Phi})$ Norm)
Let $\mathrm{TV}_{\mathrm{n},1}$ be the usual one dimensional total variation norm, then for any function $\mathbf{\Phi} : \mathbb{R}^n \to \mathbb{R}^m$, the multi dimensional TV norm is

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

The $\mathrm{TV}_{1,\mathrm{m}}(\mathbf{\Phi})$ norm is preserved for the class of vector valued functions with monotone components, Φ^i. Furthermore, the total variation of such a function is the *Euclidean distance* between $\mathbf{\Phi}(a)$ and $\mathbf{\Phi}(b)$, *see* figure 2.

Theorem 1 (Total Variation Preservation for Monotone Functions)
Let $\mathbf{\Phi} : [a, b] \subset \mathbb{R} \rightarrow \mathbb{R}^m$ be a vector valued function with monotone components, *i.e.* $\Phi^i(x)$ are monotone for all i, then $\mathrm{TV}_{1,\mathrm{m}}(\mathbf{\Phi}) = d(\mathbf{\Phi}(a), \mathbf{\Phi}(b))$.

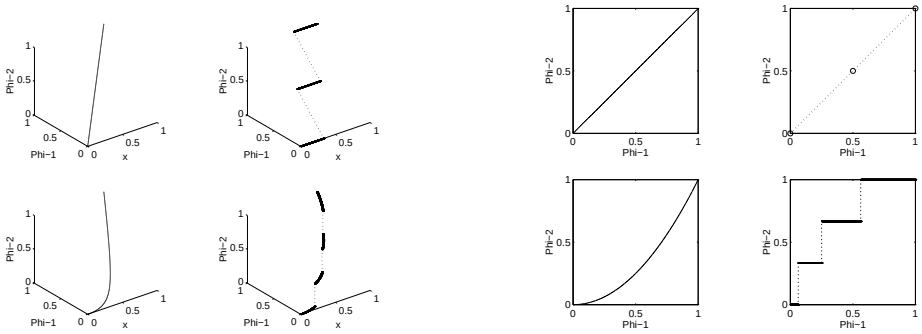

Figure 2: Functions with the same Total Variation (left), and their projections to $\mathbf{\Phi}$-space (right).

Proof of Theorem 1.
Since the vector components are monotone, we have

$$\mathrm{TV}_{1,1}(\Phi^i) = \left|\Phi^i(b) - \Phi^i(a)\right| = d\left(\Phi^i(a), \Phi^i(b)\right)$$

by Property 1 of the scalar TV norm. Therefore

$$\mathrm{TV}_{1,\mathrm{m}}(\mathbf{\Phi}) = \sqrt{\sum_{i=1}^{m} \left[\mathrm{TV}_{1,1}(\Phi^i)\right]^2} = \sqrt{\sum_{i=1}^{m} \left[\Phi^i(b) - \Phi^i(a)\right]^2} = d(\mathbf{\Phi}(a), \mathbf{\Phi}(b)).$$

□

As in the gray scale case, this definition of $\mathrm{TV}_{\mathrm{n,m}}(\mathbf{\Phi})$ allows discontinuities (edges). Since all component-monotone functions have the same norm it preserves edges, in the sense that no monotone function is preferred over another in the minimization problem

$$\min_{\mathbf{\Phi} \in \mathrm{BV}(\Omega)} \mathrm{TV}_{\mathrm{n,m}}(\mathbf{\Phi}) - \frac{\lambda}{2} \left\| \mathbf{\Phi} - \mathbf{\Phi}^0 \right\|_2^2 . \tag{3}$$

Furthermore, it is not biased against smooth functions so artificial edges will not be introduced, and it is also rotationally invariant in image space. The rotational

invariance follows from the fact that the $\mathrm{TV}_{1,1}$ norms of the components, Φ^i, are rotationally invariant.

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

3 Other Approaches

Other approaches to the vector valued restoration problem, *e.g.* anisotropic diffusion [2, 3, 5, 18, 24, 30], edge detection and segmentation [2, 5, 12, 16, 27, 32] as well as segmentation methods related to level-set methods [4, 15, 17, 22, 23] can be found in the literature. Here we will briefly describe the general anisotropic diffusion approach, and Sapiro-Ringach's [22, 23, 24] approach in particular.

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

$$\min_{\mathbf{\Phi}\in \mathrm{BV}(\Omega)} \|\mathbf{\Phi}\|_{\widetilde{TV}} - \frac{\lambda}{2} \left\| \mathbf{\Phi} - \mathbf{\Phi}^0 \right\|_2^2 \tag{4}$$

where $\mathbf{\Phi}^0$ is a given initial (noisy) image, λ a Lagrange multiplier associated with the noise level, and $\|\cdot\|_{\widetilde{TV}}$ an appropriate definition of the norm.

Notice that solving the nonlinear system of equations

$$\frac{\partial \mathbf{\Phi}}{\partial t} = \mathbf{F}\left(\mathbf{\Phi}, \partial/\partial x_1, \ldots, \partial/\partial x_n; \lambda\right) \tag{5}$$

where $\mathbf{F}(\cdot) = 0$ is the Euler-Lagrange equations associated with (4), is a form of anisotropic diffusion.

However, before comparing $\mathrm{TV}_{\mathrm{n,m}}$ and the anisotropic diffusion approach, we explore the relationship between $\mathrm{TV}_{\mathrm{n,m}}$ and the simplest extension of the total variation approach to the vector valued case — channel-by-channel TV restoration.

3.1 $\mathrm{TV}_{\mathrm{n,m}}$ 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 $\mathrm{TV}_{\mathrm{n,m}}$ norm and give an indication of how the performance of this norm is different from the channel-by-channel TV norm.

Consider the following denoising problems:

Channel-By-Channel TV

$$\min_{\Phi^i \in BV(\Omega)} \left\{ \mathrm{TV}_{\mathrm{n},1}(\Phi^i) - \frac{\lambda}{2} \left\| \Phi^i - \Phi^{i,0} \right\|_2^2 \right\}, \tag{6}$$

Color TV

$$\min_{\mathbf{\Phi} \in BV(\Omega)} \left\{ \mathrm{TV}_{\mathrm{n,m}}(\mathbf{\Phi}) - \frac{\lambda}{2} \left\| \mathbf{\Phi} - \mathbf{\Phi}^0 \right\|_2^2 \right\}, \tag{7}$$

where $\|\mathbf{\Phi}\|_2^2$ is the channel-mixed-l_2 norm, $\sum_{j=1}^m \|\Phi^i\|_2^2$.

The corresponding *Euler-Lagrange* equations are

Channel-by-Channel TV

$$\nabla \circ \left(\frac{\nabla \Phi_i}{\|\nabla \Phi_i\|} \right) - \lambda \left(\Phi_i - \Phi_i^0 \right) = 0, \tag{8}$$

Color TV

$$\frac{\mathrm{TV}_{\mathrm{n},1}(\Phi_i)}{\mathrm{TV}_{\mathrm{n,m}}(\mathbf{\Phi})} \nabla \circ \left(\frac{\nabla \Phi_i}{\|\nabla \Phi_i\|} \right) - \lambda \left(\Phi_i - \Phi_i^0 \right) = 0. \tag{9}$$

The $\mathrm{TV}_{\mathrm{n,m}}$ coupling takes the form of a global channel-wise modification of the Lagrange-multiplier λ:

$$\lambda_i \leftarrow \lambda \cdot \frac{\mathrm{TV}_{\mathrm{n,m}}(\mathbf{\Phi})}{\mathrm{TV}_{\mathrm{n},1}(\Phi_i)}. \tag{10}$$

This has the implication that a channel with large TV will be smoothed more than a channel with small TV. Consider the simple model example, with $m = 2$, $n = 1$, depicted in figure 3; we see that for a particular λ the Channel-By-Channel TV almost completely wipe out the weaker of the channels, whereas the $\mathrm{TV}_{\mathrm{n,m}}$ regularization maintains a balance of how much each channel is smoothed. Figure 4 shows reconstruction from a noisy signal.

In both these examples we are measuring the noise, or allowed deviation from the initial signals, in the channel-mixed-l_2 norm. In cases where we have separate noise measures for the separate channels, the channel-by-channel approach may be more successful. For a more complete discussion of how the choice of regularization parameter λ affects the scale of TV smoothing, see Strong-Chan [26].

3.2 Anisotropic Diffusion Based on Riemannian Geometry

We describe the the *Riemannian Geometry* based *anisotropic diffusion* in general, and Sapiro-Ringach's [22, 23, 24] approach in particular, since their approach is closely related to ours in the sense that the implied definition of the vector valued TV norm is closely related to our definition.

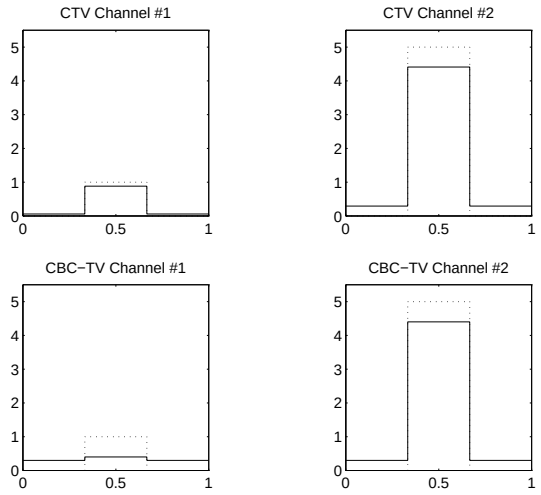

Figure 3: Application of the denoising (TV smoothing) to a simple model problem where the two channels have different strength. Here $\lambda = 10$. The upper two pictures show the result of the Color TV regularization, and the lower pictures show the result of Channel-By-Channel application of the TV norm. The dotted line shows the initial data.

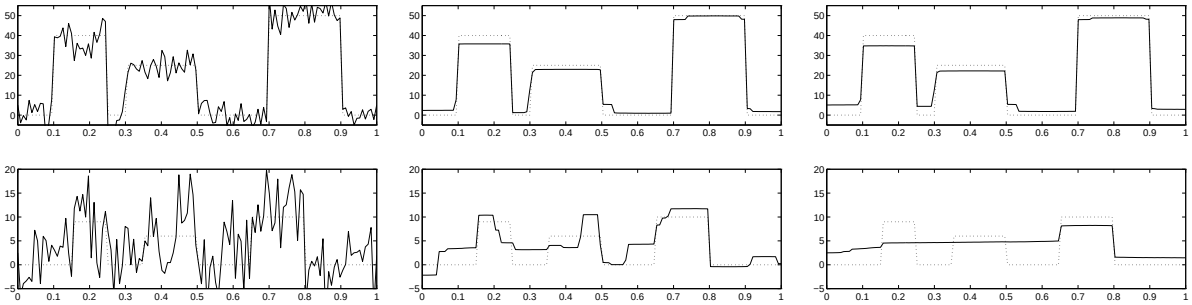

Figure 4: Comparison of Color TV, and Channel-By-Channel TV. Regularization parameter $\lambda = 2$. `Left-to-right:` The initial signals, the Color TV reconstruction, and the Channel-By-Channel reconstruction. Notice that channel #1 is about 5 times as strong as channel #2.

The Riemannian geometry framework for edge detection in vector valued images was first suggested in [32], see also [12, 16, 27, 30, 32]. Given a multi dimensional image $\mathbf{\Phi}$, then the *first fundamental form* [10, 11]

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

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

$$g_{ij} = \frac{\partial \mathbf{\Phi}}{\partial x_i} \circ \frac{\partial \mathbf{\Phi}}{\partial x_j}. \tag{12}$$

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

$$\frac{\partial \mathbf{\Phi}}{\partial t} = g\left(\lambda^+, \lambda^-\right) \frac{\partial^2 \mathbf{\Phi}}{\partial \xi_-^2}, \tag{13}$$

where, for $n = 2$,

$$\lambda^{\pm} = \frac{g_{11} + g_{22} \pm \sqrt{(g_{11} - g_{22})^2 + 4g_{12}^2}}{2} \tag{14}$$

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

$$\xi_+ = \frac{1}{2} \arctan\left(\frac{2g_{12}}{g_{11} - g_{22}}\right) \tag{15}$$

$$\xi_- - \xi_+. \tag{16}$$

$g(\cdot)$ 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 $\frac{\partial^2 \mathbf{\Phi}}{\partial \xi_-^2} = 0$ is the Euler-Lagrange equations for the total variation norm, which can be written as:

$$\mathrm{TV}_{\mathrm{n},1}(\Phi) \equiv \int_{\Omega \in \mathbb{R}^n} \sqrt{\lambda_1^+}\, d\mathbf{x}. \tag{17}$$

For a general $g(\cdot)$, (13) is not a time marching scheme associated with the Euler-Lagrange equations of a norm.

An alternative, but equivalent, formulation is given by Chambolle in [8], who suggests using the diffusion equation (13) with the directions

$$\xi^+ = \arg \max_{\substack{\xi \in \mathbb{R}^2 \\ |\xi|=1}} \left\| \begin{matrix} (\nabla \Phi^1,\, \xi) \\ (\nabla \Phi^2,\, \xi) \\ (\nabla \Phi^3,\, \xi) \end{matrix} \right\| \tag{18}$$

and $\xi^- - \xi^+$. It is straight forward to show that equations (11) and (18) define the same directions, $\xi^{\pm}$.

3.3 Sapiro's Generalization of the TV Norm

In [22], Sapiro suggests the most general extension of the TV norm based on Riemannian geometry:

$$\|\mathbf{\Phi}\|_{\widetilde{\mathrm{TV}}} = \int_{\Omega \in \mathbb{R}^2} f(\lambda^+, \lambda^-) d\mathbf{x}, \tag{19}$$

where $\lambda^{\pm}$ are the eigenvalues of the matrix G in (11), and $f(\cdot)$ is a non-decreasing function of $\lambda_{\pm}$; the most natural choice is $f(\cdot) = \sqrt{\lambda^+ + \lambda^-}$, *i.e.* the square root of the trace of the matrix G, *see* section 4. This choice is equivalent to the TV norm in the scalar case ($\lambda^- \equiv 0$). We shall show, however, that this definition of the norm has a tendency to smooth edges, and is therefore not the optimal choice for vector valued image reconstruction in applications where edge preservation is important. In this paper we do not discuss the norm proposed by Sapiro [22], generated by the function $f(\cdot) = \psi(\lambda^+ - \lambda^-)$.

We can give an intuitive explanation for the smoothing caused by this definition of the norm: a function $f(\lambda_+, \lambda_-)$, non-decreasing in both arguments, is essentially $f(d\mathbf{\Phi}^2)$ where $d\mathbf{\Phi}$ is defined in (11). $d\mathbf{\Phi}$ is the arc element in $\mathbf{\Phi}$-space, and the norm $\int f(d\mathbf{\Phi}^2)\, d\mathbf{x}$ is a measure of the arc length. Minimizing using 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 avoids an inner product like the one in (11) and is therefore not concerned with the arc length.

As shown in figure 2, the projection of the embedding of the map $\mathbf{\Phi} : \mathbb{R}^n \to \mathbb{R}^m$ in $\mathbb{R}^{n+m}$ projected into $\mathbb{R}^m$ ($\mathbf{\Phi}$-space) can be a discrete set and therefore the arc length is not well defined. It turns out that the norm measures the generalized arc length, where jumps in $\mathbf{\Phi}$ space are connected by straight lines, *see* figures 6 and 7.

We give three simple examples which illustrate the differences between our approach and the Riemannian geometry approach.

Example 1
Let $\mathbf{\Phi} : [0,1] \subset \mathbb{R} \to \mathbb{R}^2$ be monotone in its components:

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

We get:

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

and we have

$$\begin{aligned} \lim_{\gamma \searrow 0} N(\gamma) &= \sqrt{2} \\ \lim_{\gamma \nearrow 1} N(\gamma) &= 2. \end{aligned}$$

As shown in figure 5 this definition of the norm "prefers" solutions with small gradients. Since the components are monotone, our extension of the TV norm would yield the same value independent of γ.

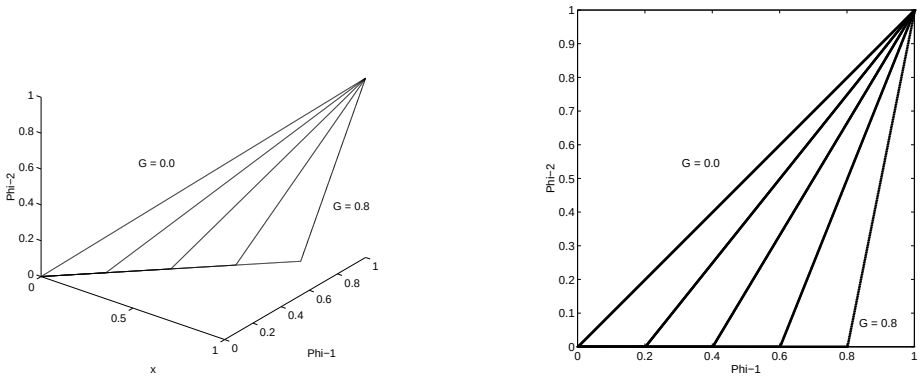

Figure 5: Functions with the same Total Variation in the Blomgren-Chan TV-norm and different (increasing in γ) $\sqrt{\lambda^+ + \lambda^-}$-norm (left); and their projections to $\mathbf{\Phi}$-space (right).

Example 2

Consider a discontinuous $\mathbf{\Phi}$, *e.g.*

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

which has the projection $\{(0,0),\,(1,1)\}$ into $\mathbf{\Phi}$-space. In this case the Blomgren-Chan TV norm and the $\sqrt{\lambda^+ + \lambda^-}$ norm coincide with value $\sqrt{2}$; *see* figure 6.

Now, consider the perturbed problem

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

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

which has the projection $\{(0,0),(1,0),(1,1)\}$ into $\mathbf{\Phi}$-space. In this case the Blomgren-Chan TV norm takes on the value $\sqrt{2}$ and the $\sqrt{\lambda^+ + \lambda^-}$ norm the value 2; *see* figure 7.

We observe that the $\sqrt{\lambda^+ + \lambda^-}$ norm is biased in favor of vector valued functions whose projections into $\mathbf{\Phi}$-space are closer to the line parameterized by $\Phi^i(s) = s$, *i.e.* functions with small arc length in $\mathbf{\Phi}$-space. In RGB space this means that the norm prefers gray scale images over color images!

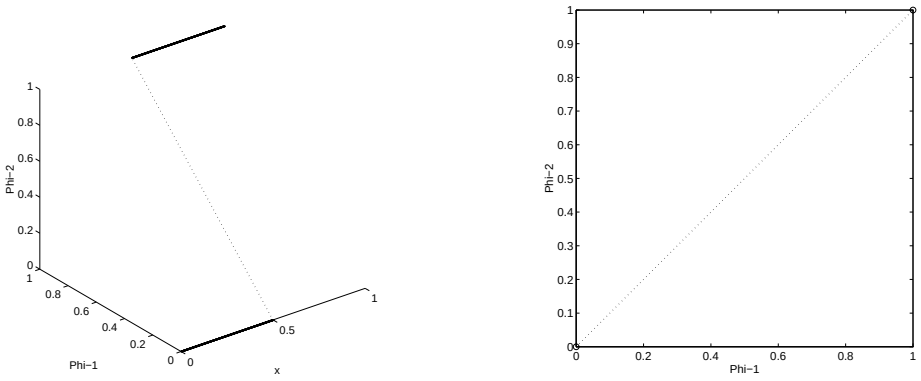

Figure 6: A function for which the Blomgren-Chan TV-norm and the $\sqrt{\lambda^+ + \lambda^-}$-norm coincide. Notice that the projection only contains the two points $(0,0)$ and $(1,1)$, the (generalized) arc length is the straight line connecting those two points.

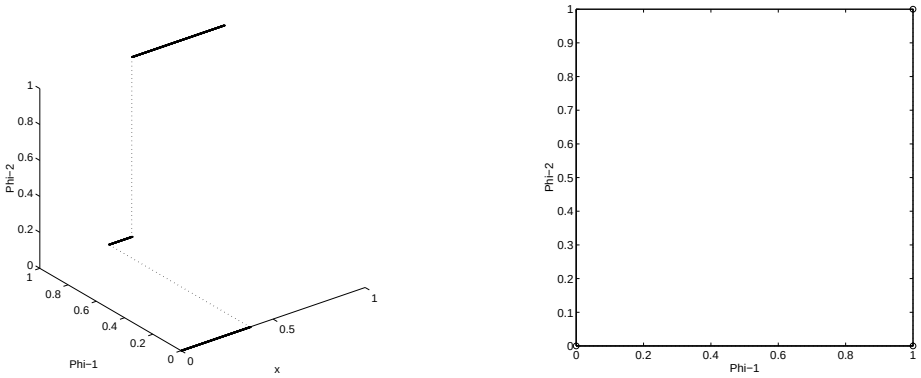

Figure 7: A for which the Blomgren-Chan TV-norm and the $\sqrt{\lambda^+ + \lambda^-}$-norm are different.

Example 3

We compare numerical restorations using the $\mathrm{TV}_{\mathrm{n,m}}$ and $\sqrt{\lambda^+ + \lambda^-}$ norms. We choose the functions $\Phi^{1,2}(x)$ as in example 1 with $\gamma = 0.5$. We solve the problem

$$\min_{\substack{\Phi \in \mathrm{BV}(\Omega) \\ \Phi^1(0)=0,\, \Phi^1(1)=1}} \alpha \|\Phi\| - \frac{1}{2} \left\| \Phi - \Phi^0 \right\|_2^2$$

for some values of α. Figure 8 shows the restorations visualized in the $\mathbf{\Phi}$-space, and figure 9 shows the visualization where the $(R,G,B) = (\Phi^1, \Phi^2, \Phi^2)$. We notice that, as expected, the $\sqrt{\lambda^+ + \lambda^-}$ norm has a tendency to smooth the kink in the curve much more than the $\mathrm{TV}_{\mathrm{n,m}}$ norm does, especially for large α.

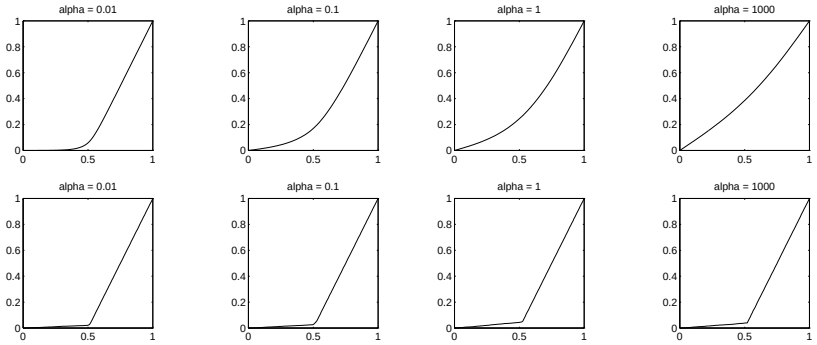

Figure 8: Restorations visualized in $\mathbf{\Phi}$-space for some values of α. (SNR $= \infty$, no noise) The upper graphs correspond to the $\sqrt{\lambda^+ + \lambda^-}$ norm, and the lower to the $\mathrm{TV}_{\mathrm{n,m}}$ norm, the small "lift" from the exact solution for the $\mathrm{TV}_{\mathrm{n,m}}$ norm depends on the numerical regularization parameter β, here $\beta = 10^{-8}$. (See section 6.1 for explanation of β.)

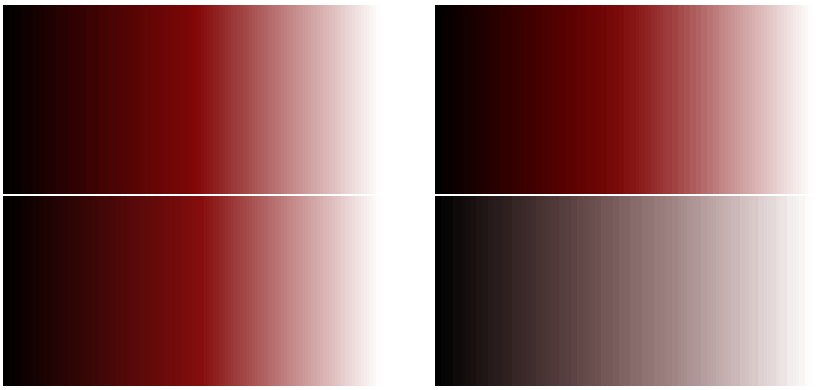

Figure 9: Restorations visualized in RGB color space, the upper two pictures show the TV (left) and $\sqrt{\lambda^+ + \lambda^-}$ (right) reconstructions for $\alpha = 0.01$, and the lower two pictures show the reconstructions for $\alpha = 1000$. We observe a clear shift toward a gray scale image for the $\sqrt{\lambda^+ + \lambda^-}$ norm when $\alpha = 1000$. (This is a color figure)

4 A General Framework for Vector Norms

Our definition of the vector valued TV-norm as the l^2-norm of the TV-norm of the separate channels may seem arbitrary. It may seem possible that there exist other definitions which also preserve the stated desired properties of a vector TV-norm. In this section we will show that our definition is the most natural within a quite general class of norms.

First, we notice that we can write the norm as a composition of l^2, and

L^1-norms:

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

In this framework, the six permutations of

$$l^{p_m}(m,i) \circ L^{p_\Omega}(\Omega) \circ l^{p_n}(n,j)$$

define six three-parameter families of norms.

Lemma 1
Composition of two norms $\|\cdot\|_{1,p_1}$ and $\|\cdot\|_{2,p_2}$ commute if and only if $p_1 = p_2$.

Proof of Lemma 1.
Let $\mathcal{L}_1$ and $\mathcal{L}_2$ be the linear operators defining $\|\cdot\|_{1,p_1}$ and $\|\cdot\|_{2,p_2}$, and let $p_1 = p_2 = p$. We get:

$$\begin{aligned} \| \|\Psi\|_{1,p} \|_{2,p} &= \sqrt[p]{\mathcal{L}_2(\sqrt[p]{\mathcal{L}_1(|\Psi|^p)})^p} \\ &= \sqrt[p]{\mathcal{L}_2(\mathcal{L}_1(|\Psi|^p))} = \sqrt[p]{\mathcal{L}_1(\mathcal{L}_2(|\Psi|^p))} \\ &= \sqrt[p]{\mathcal{L}_1(\sqrt[p]{\mathcal{L}_2(|\Psi|^p)})^p} = \| \|\Psi\|_{2,p} \|_{1,p}. \end{aligned}$$ □

Using this general characterization, we argue as follows (*see also* [1]): a necessary, but not sufficient, condition for rotational invariance in physical space is $p_n = 2$; for rotational invariance in projection ($\mathbf{\Phi}$) space, we need $p_m = 2$*, and finally, for any choice $p_\Omega \neq 1$ there is no hope of defining a composition norm which is unbiased with respect to discontinuous functions (edges). Therefore the correct composition norm from this collection of norms is one of the following four possibilities:

1. $(l^2(m,j) \circ L^1(\Omega) \circ l^2(n,i))(\mathbf{\Phi})$ Blomgren-Chan
2. $(L^1(\Omega) \circ l^2(m,j) \circ l^2(n,i))(\mathbf{\Phi})$ Sapiro [22], with $f(\lambda^+, \lambda^-) = \sqrt{\lambda^+ + \lambda^-}$
3. $(l^2(n,i) \circ L^1(\Omega) \circ l^2(m,j))(\mathbf{\Phi})$
4. $(l^2(m,j) \circ l^2(n,i) \circ L^1(\Omega))(\mathbf{\Phi})$.

The choice $f(\lambda^+, \lambda^-) = \sqrt{\lambda^+ + \lambda^-}$ is the most natural in the sense that it fits into this general framework. Other choices of $f(\lambda^+, \lambda^-)$ do not fit. In an image where $n > 2$, $f(\lambda^1, \lambda^2, \ldots, \lambda^n) = \sqrt{\lambda^1 + \lambda^2 + \cdots + \lambda^n} = \sqrt{\mathrm{trace}(G)}$ is the most natural choice of $f(\cdot)$.

It is worth noticing that when $m = 1$, the first two reduce to

*This property is not crucial, since vision, and hence color-spaces are non-linear; *i.e* rotational invariance in color space has no meaning.

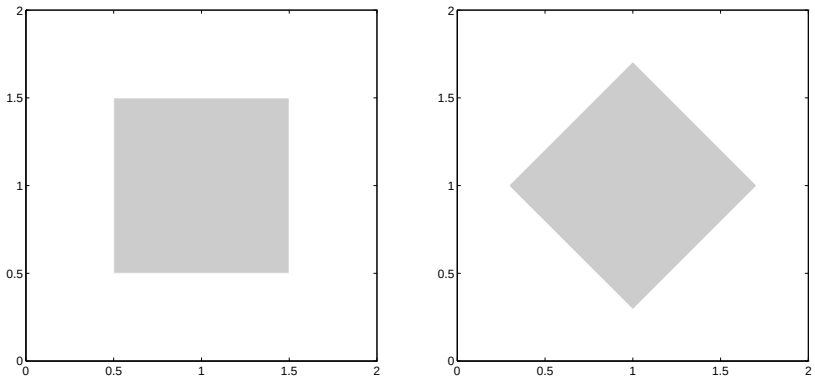

Figure 10: The $(l^2(n,i) \circ L^1(\Omega))$-norm takes the value $2\sqrt{2}$ for the image to the left, and 4 for the image to the right. The classical TV-norm takes the value 4 for both images. (The function value is 0 in the white region, and 1 in the gray region.)

- $(L^1(\Omega) \circ l^2(n,i))(\mathbf{\Phi})$,

which is the "classical" gray scale TV-norm, and the last two reduce to

- $(l^2(n,i) \circ L^1(\Omega))(\mathbf{\Phi}) \equiv \sqrt{\left(\int_\Omega |\Phi_x|\, dx\, dy\right)^2 + \left(\int_\Omega |\Phi_y|\, dx\, dy\right)^2}$

which is an alternate definition for the gray scale TV-norm; however, this definition is *not* rotationally invariant in physical space, *see* figure 4.

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 $n = 1$, *see* Theorem 1.

Finally, we note that, in our notation

- $(L^1(\Omega) \circ l^1(n,i))(\mathbf{\Phi})$, is the "taxi," or "Manhattan-distance" TV-norm of Li-Santosa [14]; *see also* LeVeque [13, chapter 15].

Summary of Properties of the Norms

In Table 1, we summarize the properties of the different definitions of vector TV-norms. We see that the definition proposed in this paper is the only one (from this class) which satisfies all four of the desirable properties.

5 Color and Color spaces

The human retina has three types of color receptor cells which respond to incident radiation with different spectral response curves. Color is the *perceptual*

Norm	Reduces to Scalar TV	Rotationally Invariant	Allows Edges	Unbiased wrt. Color
Blomgren-Chan, $\mathrm{TV}_{\mathrm{n,m}}$ $(l^2(m,j) \circ L^1(\Omega) \circ l^2(n,i))$	Yes	Yes	Yes	Yes
Sapiro $\sqrt{\lambda^+ + \lambda-}$ $(L^1(\Omega) \circ l^2(m,j) \circ l^2(n,i))$ $(L^1(\Omega) \circ l^2(n,i) \circ l^2(m,j))$	Yes	Yes	Yes	No
$(l^2(n,i) \circ L^1(\Omega) \circ l^2(m,j))$	No	No	Yes	No
$(l^2(m,j) \circ l^2(n,i) \circ L^1(\Omega))$ $(l^2(n,i) \circ l^2(m,j) \circ L^1(\Omega))$	No	No	Yes	Yes

Table 1: Properties of different vector valued TV norms. Rotational invariance refers to physical (image) space. Since all the compositions include $L^1(\Omega)$, they are all unbiased with respect to discontinuities. The Sapiro, and $(l^2(n,i) \circ L^1(\Omega) \circ l^2(m,j))$ norms have a tendency to smooth colors (bias against color images) as discussed in Examples 1 and 3 in section 3.3.

result of light in the visible region of the spectrum, having wavelengths between 400 nm and 700 nm. Since there are three color receptors, we can describe color using three numerical values.

Intensity is a linear measure of the total energy over some interval of the electro-magnetic spectrum.

Luminance is a measure of the radiant power weighted by a spectral sensitivity function characteristic of vision, denoted by Y.

Human vision responds nonlinearly to luminance; the perceptual response to luminance is called *lightness*, denoted by L^*.

$$L^* \quad = \quad 116 \left(\frac{Y}{Y_n} \right)^{\frac{1}{3}} - 16, \quad Y > 0.008856\, Y_n$$

where Y_n is the luminance of the white reference, for $Y < 0.008856\, Y_n$ the L^* is a linear function of L.

There exist several color specifications, including CIE XYZ, CIE xyY, CIE $L^*u^*v^*$, CIE $L^*a^*b^*$, linear RGB, nonlinear $R'G'B'$. [6, 7, 25, 31]

A color system is *perceptually uniform* if a perturbation, δ to a component value is equally perceptible across the entire range of component values. The naive XYZ and RGB systems are far from uniform, *i.e.* the perceptibility of a perturbation $(\Delta R, \Delta G, \Delta B)$ is dependent of the value of (R, G, B). The CIE $L^*u^*v^*$ and CIE $L^*a^*b^*$ spaces are first approximations for perceptually uniform color spaces. Formulas for transforming between RGB and CIE $L^*a^*b^*$ can be found in [22] and the very accessible introduction to color [19].

The computations in section 6 were performed in linear RGB space, which is a perceptually nonlinear space. Perceptually better results could be obtained by performing the calculations in CIE $L^*a^*b^*$.

6 Numerical Experiments

6.1 Noise Reduction Using $\mathrm{TV}_{\mathrm{n,m}}(\Phi)$

Given an initial, noisy, image Φ^0 on a domain $\Omega \subset \mathbb{R}^n$, we are interested in minimizing the total variation norm, $\mathrm{TV}_{\mathrm{n,m}}(\Phi)$ subject to constraints on Φ given by the variance of the noise, σ^2. We want to solve:

$$\min_{\mathbf{\Phi} \in BV(\Omega)} \left\{ \mathrm{TV}_{\mathrm{n,m}}(\mathbf{\Phi}) - \frac{\lambda}{2} \left\| \mathbf{\Phi} - \mathbf{\Phi}^0 \right\|_2^2 \right\}. \tag{20}$$

It is straight forward to show that the corresponding Euler-Lagrange equations are

$$\frac{\mathrm{TV}_{\mathrm{n},1}(\Phi_i)}{\mathrm{TV}_{\mathrm{n,m}}(\mathbf{\Phi})} \nabla \circ \left(\frac{\nabla \Phi_i}{\|\nabla \Phi_i\|} \right) - \lambda \left(\Phi_i - \Phi_i^0 \right) = 0. \tag{21}$$

The following numerical experiments were all performed in Euclidean spaces, which for color images represent the linear RGB space. As discussed in section 5 this is not the optimal space for perceptual uniformity, but the results are of interest anyway.

The solutions were computed using an explicit time marching scheme, a slight modification of the one introduced in [21], applied to the system of equations

$$\frac{\partial \Phi_i}{\partial t} = \frac{\mathrm{TV}_{\mathrm{n},1}(\Phi_i)}{\mathrm{TV}_{\mathrm{n,m}}(\mathbf{\Phi})} \nabla \circ \left(\frac{\nabla \Phi_i}{\sqrt{\beta + \|\nabla \Phi_i\|^2}} \right) - \lambda \left(\Phi_i - \Phi_i^0 \right). \tag{22}$$

where β is a small numerical regularization parameter, introduced to avoid division by zero.

The signal-to-noise ratio (SNR) is defined by

$$\mathrm{SNR} = \frac{\sum_{k=1,i=1}^{m,N} \left(\Phi_i^k - \overline{\mathbf{\Phi}} \right)^2}{\sum \left(n_i^k \right)^2}, \tag{23}$$

where $\overline{\mathbf{\Phi}}$ is the average of $\mathbf{\Phi}$ over all data points (enumerated by the subscript i), and all channels (superscript k). In all the computations the Lagrange multiplier, λ, is determined by gradient projection, *cf.* [20, 21].

Example 1. Reconstruction of a Space Curve

In our first numerical experiment we define a curve (with several discontinuities) embedded in $\mathbb{R}^3$ by defining a mapping $\mathbf{\Phi} : [0,1] \subset \mathbb{R} \rightarrow \mathbb{R}^2$. We add Gaussian noise (SNR = 2.90) to the data, and apply the TV-denoising to the noisy curve.

The time step was $\texttt{dt} = 7.75 \cdot 10^{-6}$, and the iteration was stopped after 7290 iterations when the l_2-norm of the correction was less than $1.5 \cdot 10^{-5}$.

Admittedly, this is a very artificial setting. The rationale for this experiment was to convince ourselves that denoising using the extended definition of the TV norm would indeed preserve sharp edges. Also it provides a computationally "cheap" example which is easy to visualize.

As shown in figures 11–12 the reconstruction is quite good. Since we are dealing with a space curve in $\mathbb{R}^3$, we are pushing the question of perceptual uniformity aside.

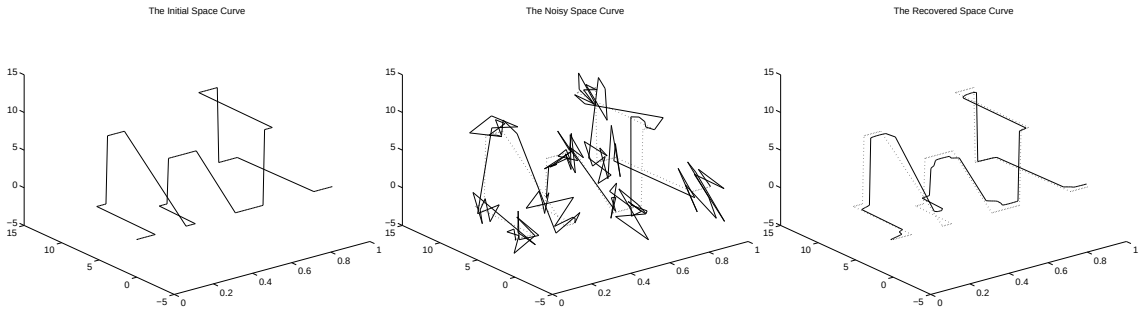

Figure 11: The initial, noisy (SNR = 2.90), and recovered space curves. Notice how the edges are recovered.

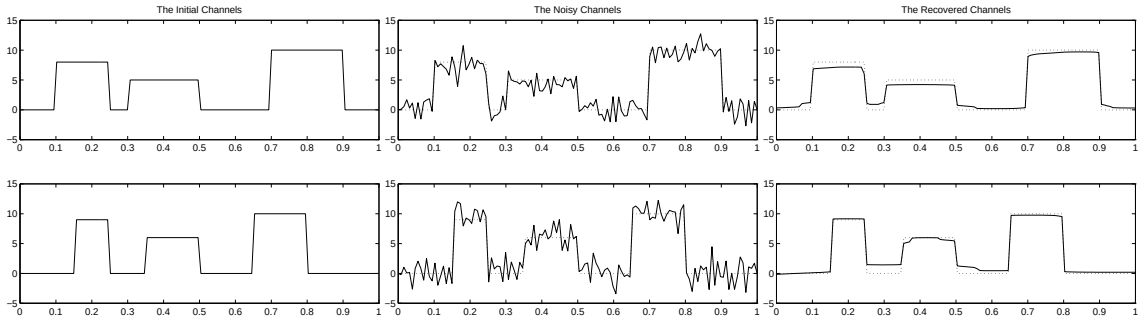

Figure 12: Channel separation for initial, noisy, and recovered data. (Same data as in figure 11.)

Example 2. Reconstruction of a One-Dimensional Color Image

A one dimensional 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 = 0.5) was added and an explicit time marching scheme was applied to reduce the noise (*see* figures 13–14).

The images are visualized by extending the one dimensional lines into two dimensional strips.

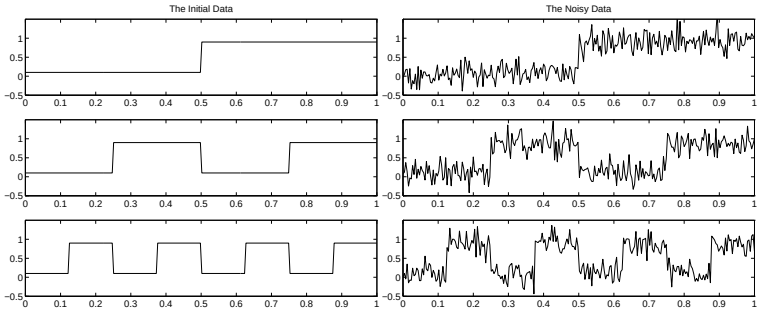

Figure 13: Channel separation for the initial and noisy (SNR = 0.5) images.

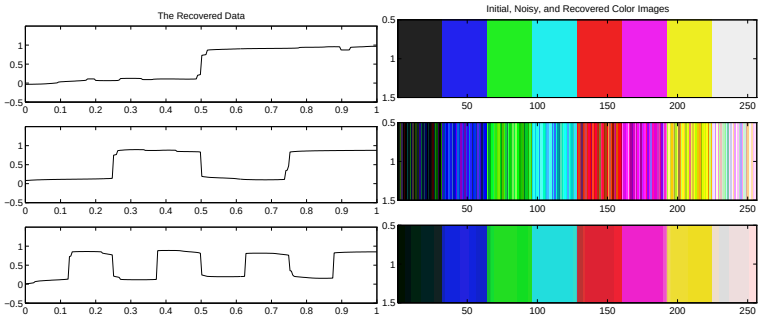

Figure 14: Channel separation for the recovered image, and color visualization for the original, noisy and recovered images. (This is a color figure)

Example 3. Denoising of a Two-Dimensional Color Image

For the full two-dimensional denoising we created a test image in RGB space, *i.e.* $\mathbf{\Phi} : \mathbb{R}^2 \to \mathbb{R}^3$; Gaussian noise was added (SNR = 2.5), and we ran the explicit time marching ($\texttt{dt} = 1.0 \cdot 10^{-6}$) with the regularizing parameter $\beta = 1.0 \cdot 10^{-9}$. After 4150 iterations, the TV-norm of the noisy image had been reduced by 82%. Channel separated initial, noisy, and restored images in figures 15–17, color realization in figure 18.

Example 4. Reconstruction of the Color Lena Image

Our final example is denoising (SNR = 2.0) of a 128×128 color "Lena" image. We used the parameters $\texttt{dt} = 10^{-6}$, and $\beta = 10^{-10}$, the results are shown in figures 19-20.

We observe that the reconstructions do not smear edges. The quality of the reconstruction increases as the size of the regularization parameter β and the timestep `dt` in the explicit time marching shrink. It is well known that the steepest descent approach we are using here has less than optimal convergence

properties. It is straight-forward to implement more efficient numerical methods, *e.g.* fixed point (cf. Vogel [28, 29]) and primal-dual (cf. Chan-Golub-Mulet [9]) methods.

7 Concluding Remarks

We have introduced a new definition of the total variation norm, $\mathrm{TV}_{\mathrm{n,m}}(\mathbf{\Phi})$, for vector valued functions, $\mathbf{\Phi} : \mathbb{R}^n \to \mathbb{R}^m$. This definition has a number of properties which may be desirable in applications, *(i)* it allows discontinuous functions – edges, *(ii)* it is rotationally invariant in physical space, and *(iii)* 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 [22]. 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 show how our norm fits into this framework, and explain why the $\sqrt{\lambda^+ + \lambda^-}$ norm of Sapiro is the most "natural" choice from the class of norms $f(\lambda^+, \lambda^-)$. Many promising possible norm definitions fall outside this framework, and are not discussed in this paper. We show how the classical gray scale TV, as well as Li-Santosa's [14] norms can be described in this framework.

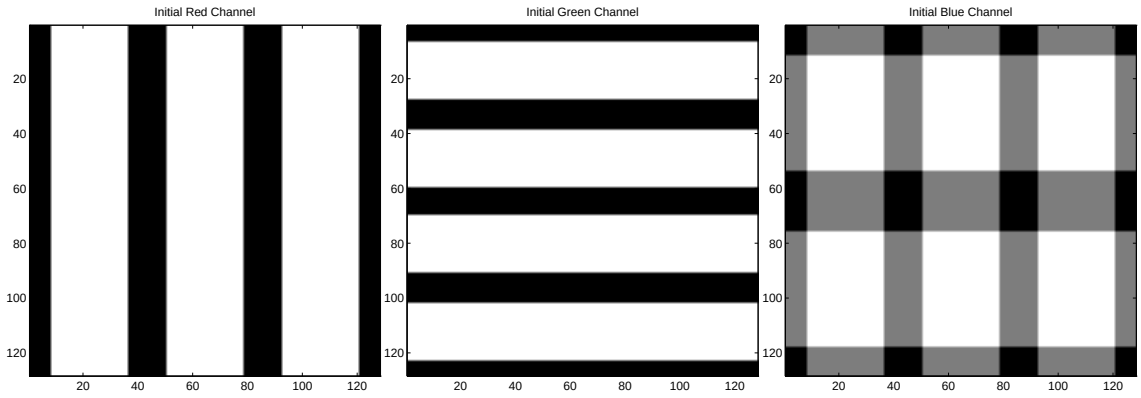

Figure 15: RGB intensity separation for the initial image.

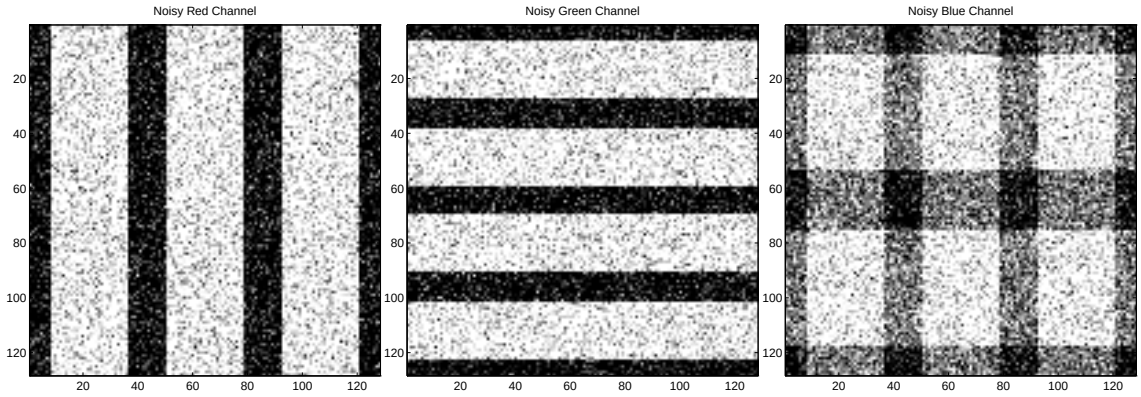

Figure 16: RGB intensity separation for the noisy (SNR = 2.5) image.

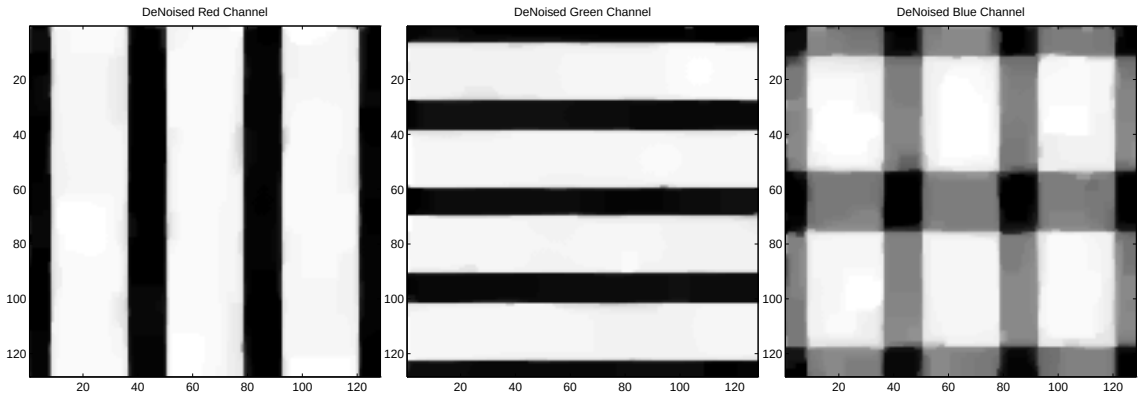

Figure 17: RGB intensity separation for the denoised image.

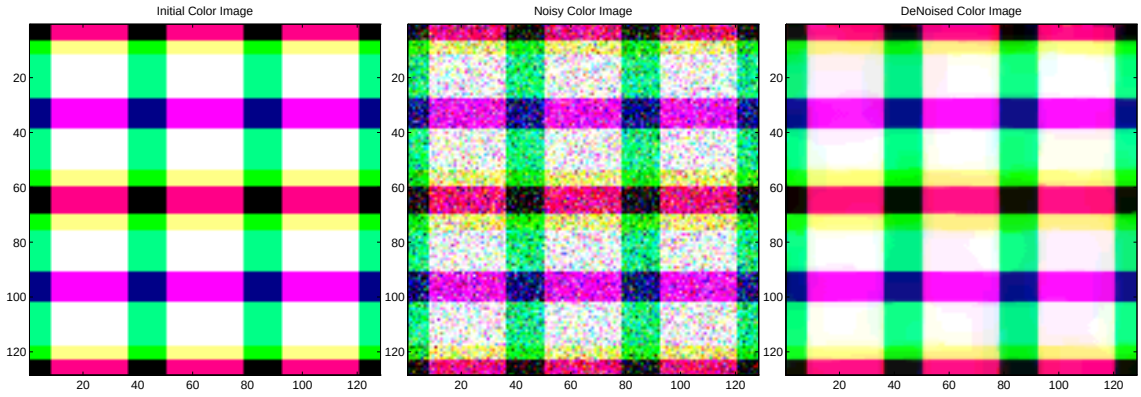

Figure 18: The initial, noisy, and denoised color images. (This is a color figure)

Figure 19: The noisy (SNR = 2.0) image (left) and two stages of denoising: 40% and 60% $\mathrm{TV}_{\mathrm{n,m}}$-norm reduction. (This is a color figure)

Figure 20: Two stages of denoising: 70% and 80% $\mathrm{TV}_{\mathrm{n,m}}$-norm reduction, and the initial image (right). We notice that even when we demand a big reduction in the norm, edges (main features) are preserved. (This is a color figure)

Acknowledgment

The authors would like to thank Dr. Guillermo Sapiro for providing us with the technical report [22] and the pre-print [24], as well as valuable discussions at the UCLA Applied Mathematics conference in Lake Arrowhead, February 22–23, 1996.

References

[1] Luis Alvarez, Frédéric Guichard, Pierre-Louis Lions, and Jean-Michel Morel. Axioms and Fundamental Equations of Image Processing. *Archive for Rational Mechanics and Analysis*, 123:199–257, 1993.

[2] Luis Alvarez, Pierre-Louis Lions, and Jean-Michel Morel. Image Selective Smoothing and Edge Detection by Nonlinear Diffusion. II. *SIAM Journal on Numerical Analysis*, 29(3):845–866, June 1992.

[3] Luis Alvarez and Luis Mazorra. Signal and Image Restoration Using Shock Filters and Anisotropic Diffusion. *SIAM Journal on Numerical Analysis*, 31(2):590–605, April 1994.

[4] Francine Catté, Françoise Dibos, and Georges Koepfler. A Morphological Scheme for Mean Curvature Motion and Applications to Anisotropic Diffusion and Motion of Level Sets. *SIAM Journal on Numerical Analysis*, 32(6):1895–1909, December 1995.

[5] Francine Catté, Pierre-Louis Lions, Jean-Michel Morel, and Tomeu Coll. Image Selective Smoothing and Edge Detection by Nonlinear Diffusion. *SIAM Journal on Numerical Analysis*, 29(1):182–193, February 1992.

[6] Austria Central Bureau of the Commission Internationale de L'Èclairage, Vienna. Publication CIE No 17.4, International Lighting Vocabulary.

[7] Austria Central Bureau of the Commission Internationale de L'Èclairage, Vienna. Publication CIE No 15.2, Colorimetry, Second Edition, 1986.

[8] A. Chambolle. Partial Differential Equations and Image Processing. In *IEEE International Conference on Image Processing*, Austin, Texas, November 1994.

[9] Tony F. Chan, Gene Golub, and Pep Mulet. A Nonlinear Primal-Dual Method for TV-Based Image Restoration. Technical Report CAM 95-43, UCLA Department of Mathematics, September 1995.

[10] A. Cumani. Edge Detection in Multispectral Images. *CVGIP: Graphical Models and Image Processing*, 53:40–51, 1991.

[11] E. Kreyszig. *Differential Geometry*. University of Toronto Press, 1959.

[12] H-C. Lee and D. R. Cok. Detecting Boundaries in a Vector Field. *IEEE Trans. Signal Proc.*, 39:1181–1194, 1991.

[13] Randall J. LeVeque. *Numerical Methods for Conservation Laws.* Lectures in Mathematics – ETH Zürich. Birkhäuser Verlag, second edition, 1992.

[14] Yuying Li and Fadil Santosa. An Affine Scaling Algorithm for Minimizing Total Variation in Image Enchancement. Technical Report CTC94TR201, Cornell University, Center for Theory and Simulation in Science and Engineering, 1994.

[15] Barry Merriman, James Bence, and Stanley Osher. Diffusion Generated Motion by Mean Curvature. Technical Report CAM 92-18, UCLA Department of Mathematics, April 1992.

[16] R. Nevatia. A Color Edge Detector and Its Use in Scene Segmentation. *IEEE Trans. Syst. Man, Cybern.*, 7:820–826, 1977.

[17] Stanley Osher and J. Sethian. Fronts Propagating with Curvature Dependent Speed: Algorithms Based on the Hamliton-Jacobi Formulation. *Journal on Computational Physics*, 79:12–49, 1988.

[18] P. Perona and J. Malik. Scale-space and Edge Detection Using Anisotropic Diffusion. *IEEE Trans. Pattern Anal. Machine Intell.*, 12:629–639, 1990.

[19] Charles A. Poynton. Frequently Asked Questions about Colour. ftp://www.inforamp.net/pub/users/poynton/doc/colour/ColorFAQ.ps.gz, 1995.

[20] J. G. Rosen. The Gradient Projection Methods for Nonlinear Programming, Part II, Nonlinear Constraints. *J. Soc. Indust. Appl. Math*, 9:514, 1961.

[21] Leonid I. Rudin, Stanley Osher, and Emad Fatemi. Nonlinear Total Variation Based Noise Removal Algorithms. *Physica D*, 60:259–268, 1992.

[22] Guillermo Sapiro. Color Snakes. Technical Report HPL-95-113, Hewlett Packard Computer Peripherals Laboratory, September 1995.

[23] Guillermo Sapiro. Vector-Valued Active Contours. In *Proceedings CVPR*, San Francisco, June 1996.

[24] Guillermo Sapiro and Dario L. Ringach. Anisotropic Diffusion in Color Space. *IEEE Trans. Image Proc.*, to appear.

[25] W. F. Schreiber. *Fundamentals of Electronic Imaging Systems.* Springer-Verlag, second edition, 1991.

[26] David M. Strong and Tony F. Chan. Relation of Regularization Parameter and Scale in Total Variation Based Image Denoising. In *IEEE Workshop on Mathematical Methods in Biomedical Image Analysis*, 1996. submitted.

[27] V. Torre and T. Poggio. On Edge Detection. *IEEE Trans. PAMI*, 8:147–163, 1986.

[28] C.R. Vogel and M. E. Oman. Fast Total Variation-Based Image Reconstruction. In *Proceedings of the ASME Symposium on Inverse Problems*, 1995. file://ftp.math.montana.edu/pub/vogel/postscript/asme95.ps.

[29] C.R. Vogel and M. E. Oman. Iterative Methods for Total Variation Denoising. *SIAM Journal of Scientific Computing*, to appear. file://ftp.math.montana.edu/pub/vogel/postscript/denoise.ps.

[30] Ross Whitaker and Guido Gerig. *Geometry Driven Diffusion in Computer Vision*, chapter Vector-Valued Diffusion. Kluwer Academic Publishers, 1994.

[31] Günter Wyszecki and W.S. Styles. *Color Science: Concepts and Methods, Quantitative Data and Formulae*. John Wiley & Sons, second edition, 1982. ISBN 0-387-53272-2.

[32] S. Di Zenzo. A Note on the Gradient of a Multi-Image. *CVGIP*, 33:116–125, 1986.