SIAM J. IMAGING SCIENCES
Vol. 3, No. 4, pp. 981–1014

Image Sharpening via Sobolev Gradient Flows*

J. Calder†, A. Mansouri†, and A. Yezzi‡

Abstract. Motivated by some recent work in active contour applications, we study the use of Sobolev gradients for PDE-based image diffusion and sharpening. We begin by studying, for the case of isotropic diffusion, the gradient descent/ascent equation obtained by modifying the usual metric on the space of images, which is the L^2 metric, to a Sobolev metric. We present existence and uniqueness results for the Sobolev isotropic diffusion, derive a number of maximum principles, and show that the differential equations are stable and well-posed both in the forward and backward directions. This allows us to apply the Sobolev flow in the backward direction for sharpening. Favorable comparisons to the well-known shock filter for sharpening are demonstrated. Finally, we continue to exploit this same well-posed behavior both forward and backward in order to formulate new constrained gradient flows on higher order energy functionals which preserve the first order energy of the original image for interesting combined smoothing and sharpening effects.

Key words. image diffusion, partial differential equations, gradient descent, gradient ascent, Sobolev spaces, image sharpening

AMS subject classifications. 94A08, 35A01, 35A02, 35A35

DOI. 10.1137/090771260

1. Introduction.

We revisit the problem of image diffusion, which plays a key role in image enhancement and denoising applications [3, 17, 1, 7, 8, 18, 21, 22, 25, 36, 35, 38], as well as more recent applications such as image compression [12]. Recall that the problem of image diffusion consists of designing a suitable differential operator A on a suitable function space H of images (e.g., $L^2(\Omega)$, where $\Omega \subset \mathbb{R}^2$ denotes the image domain), and of considering the one-parameter family for which A is the infinitesimal generator; in other words, the following functional differential equation is considered:

$$u(0) = u_0, \qquad \frac{du}{dt} = Au, \quad t > 0, \tag{1.1}$$

where u_0, the initial condition, is the image to be processed. Depending on the choice of the operator A, the one-parameter family $\{u(t)\}_{t\geq 0}$ thus obtained may correspond to successively enhanced versions of the initial image u_0.

A classical choice for A, first proposed by [35, 36], is the Laplace operator, which leads to

*Received by the editors September 16, 2009; accepted for publication (in revised form) September 20, 2010; published electronically December 2, 2010.
http://www.siam.org/journals/siims/3-4/77126.html
†Department of Mathematics and Statistics, Queen's University, Kingston, ON K7L 3N6, Canada (jcalder@mast.queensu.ca, mansouri@mast.queensu.ca). The second author's work was partially supported by a research grant from the Natural Sciences and Engineering Research Council of Canada (NSERC).
‡School of Electrical and Computer Engineering, Georgia Tech, Atlanta, GA 30332 (Anthony.yezzi@ece.gatech.edu). This author was supported in part by the National Science Foundation (NSF) grant CCF-0728911.

the heat equation

$$u(0) = u_0, \qquad \frac{du}{dt} = \Delta u, \quad t > 0. \tag{1.2}$$

The one-parameter family obtained in this way (in $L^2(\mathbb{R}^2)$) corresponds to successively blurred versions of the initial image u_0, where the blurring achieved is equivalent to convolving the image with Gaussian kernels of successively increasing variance. The key drawback of the heat equation for denoising purposes is that it blurs out important image structure as well; this is already suggested by the rotational invariance of the Laplace operator, which leads to *isotropic* smoothing.

To remedy the isotropic nature of the heat equation, Perona and Malik [22] proposed the PDE

$$u(0) = u_0, \qquad \frac{du}{dt} = \nabla \cdot (c(\|\nabla u\|)\nabla u), \quad t > 0, \tag{1.3}$$

where $c > 0$ is a smooth monotonically decreasing function typically chosen so that $c(s) \to 0$ as $s \to \infty$ and $c(s) \to 1$ as $s \to 0$; the key idea here is that diffusion should be inhibited in the presence of strong edges (as measured by c) and should proceed in the presence of weak edges. It can be shown that the diffusion induced by this PDE is such that it takes place along edges and not across edges, and hence is *anisotropic*. It is important to note that choosing c to be the constant function with value 1 yields the heat equation.

The isotropic diffusion (heat) equation can be interpreted as a gradient descent equation on the functional

$$u \mapsto \int_\Omega \|\nabla u\|^2,$$

and similarly the anisotropic diffusion equation proposed by [22] can be interpreted as a gradient descent equation on a functional of the form

$$u \mapsto \int_\Omega g(\|\nabla u\|^2)$$

for some suitable function g. It is important to note, however, that, in both cases, the gradient is meant with respect to the L^2 metric. Natural questions at this point include: What if the metric is modified from L^2 to some other metric? How does the corresponding gradient descent equation (on the very same functional) behave? What properties does the solution (if it exists) have?

In terms of Riemannian geometry, modifying the Riemannian metric implicitly changes the notion of distance on a manifold. In the context of image processing, modifying the metric on the space of images (from L^2 to another metric) yields a new distance function between images. It has long been suggested that the L^2 distance function is a poor approximation to our perceptual notion of the distance between images. Therefore, it is certainly of interest to study the gradient descent PDE (for various classical functionals) with respect to other metrics which may be more aligned with our visual perception.

In this paper, we study, for the case of isotropic diffusion, the gradient flows obtained by modifying the metric on the space of images from L^2 to a Sobolev metric. We prove existence

and uniqueness and a number of maximum principles for the corresponding gradient descent differential equation. In particular, we show that the new gradient flow is well-posed both in the forward and backward directions, allowing us to reverse the flow direction for image sharpening effects. We further exploit this same well-posed behavior in order to formulate new constrained gradient flows on other energy functionals which preserve this original energy (via orthogonal projection) for interesting combined smoothing and sharpening effects. We compare our Sobolev sharpening algorithm to the well-known shock filter [21] and show that, aside from being well-posed, the Sobolev sharpening PDE generates natural-looking images devoid of the "staircasing" artifacts which plague the shock filter.

2. Previous and related works.

2.1. Sobolev gradient flows. The work presented here is directly inspired by works on active contours [28, 27, 29], where the gradient descent PDEs for segmentation and tracking energies are recast under a Sobolev metric. The resulting tracking and segmentation algorithms, referred to as Sobolev active contours, are found to be far more robust with respect to noise and more global in nature. Charpiat et al. [9] study the case for active contours in more generality and show how to design inner products in order to achieve a desired amount of spatial coherency in the gradient flow. They also show how computing the gradient with respect to an arbitrary inner product can be interpreted as solving a minimization problem, and they use this as motivation for defining the notion of an "extended gradient."

Richardson [24] studied the use of Sobolev gradients for image decomposition tasks. Such tasks include decomposing an image into its constituent smooth, texture, and noise parts. These decompositions are typically obtained via the minimization of a least squares functional, and by using Sobolev gradient flows, one can obtain faster convergence rates and avoid irrelevant local minima. Richardson [23] also studied the use of high order Sobolev gradients in the context of numerically solving nonlinear differential equations. For further reading on Sobolev gradients and differential equations, see the book by Neuberger [20].

As can be seen, much of the previous and related work on Sobolev gradients in the context of image processing and computer vision is aimed at recasting a gradient descent algorithm under a Sobolev inner product to achieve better convergence results and less susceptibility to local minima. All that is important in these approaches is (1) which local minimum is reached by the gradient descent procedure and (2) how quickly one reaches this local minimum. We study Sobolev gradients for an entirely different purpose. In image processing, it is not the limit point of the gradient descent process that matters; instead it is the family of images generated by the gradient descent procedure that we are interested in. As such, we are far more interested in the properties of the solutions to these gradient flows and much less interested in their convergence properties as $t \to \infty$. Also, we will be interested in both the gradient descent and ascent flows, as the first leads to image smoothing while the second yields image sharpening.

2.2. Image smoothing via PDEs. Many researchers have explored image restoration and smoothing via extensions and modifications of the Perona–Malik functional. Weickert [32, 33] proposed a generalization of the Perona–Malik equation where the diffusion coefficient $c(|\nabla u|)$ depends not only on the magnitude of ∇u but also on local variations in its orientation, and has established rigorous existence and uniqueness results for the resulting PDE. Bai and Feng [4]

proposed a class of fractional order anisotropic diffusion PDEs, which are the Euler–Lagrange equations for a Perona–Malik-style functional which depends on fractional order derivatives of the image. Guidotti and coworkers [15, 16] introduced a variant on the Perona–Malik equation where the diffusion coefficient $c(|\nabla u|)$ depends on the fractional gradient $\nabla^\varepsilon u$, and established existence and uniqueness in the weak setting. Sapiro and Ringach [26] proposed anisotropic diffusion for multivalued images, and Martin-Herrero [19] extended this work to hyperspectral images. Lysaker, Lundervold, and Tai [18] proposed a class of fourth order PDEs for noise removal. Tschumperlé [30] and Tschumperlé and Deriche [31] proposed anisotropic diffusion for multivalued images and a unifying framework for vector-valued image regularizing PDEs. Yezzi [37] proposed viewing an image as a surface embedded in $\mathbb{R}^3$ and performed denoising by curvature evolution equations. Zhang and Hancock [39] proposed a discrete framework for anisotropic diffusion based on the heat equation on a graph.

The above summary is by no means exhaustive, but it is important to note that all of these restoration algorithms are performing image diffusion and smoothing. We now present a summary of the relevant sharpening algorithms in the literature.

2.3. Image sharpening via PDEs. Although there is an enormous amount of literature on image diffusion via PDEs, there is relatively little on image sharpening and enhancement. This is perhaps due to the inherently ill-posed nature of such PDEs. One of the oldest and most familiar image sharpening PDEs is the shock filter proposed by Osher and Rudin [21]. The general form for the shock filter is

$$u(0) = u_0, \qquad \frac{du}{dt} = -|\nabla u|\mathcal{L}(u), \quad t > 0, \tag{2.1}$$

where $\mathcal{L}$ is an edge detecting operator. Typical choices for $\mathcal{L}$ are

$$\mathcal{L}(u) = \frac{\Delta u}{1 + |\Delta u|} \quad \text{and} \quad \mathcal{L}(u) = \frac{u_{\eta\eta}}{1 + |u_{\eta\eta}|},$$

where $u_{\eta\eta} = \sum_{i,j} u_{x_i} u_{x_j} u_{x_i x_j}$ is the second derivative in the direction of the gradient ∇u. From a mathematical point of view, the shock filter looks severely ill-posed; however, Osher and Rudin [21] devised a sophisticated numerical scheme which gives satisfying results. They have conjectured that, for continuous initial data $u_0(x)$, the one-dimensional shock filter has a unique solution which is continuous everywhere except at a finite number of points that correspond to the inflection points of u_0. Despite its apparent ill-posedness, the shock filter can be combined with anisotropic diffusion in a well-posed sharpening and smoothing framework [2].

More recently, Gilboa, Sochen, and Zeevi [14] proposed a forward and backward diffusion model based on the Perona–Malik model (1.3) with diffusion coefficient

$$c(s^2) = \frac{1}{1 + (s/k_f)^n} - \frac{\alpha}{1 + ((s - k_b)/w)^{2m}}, \tag{2.2}$$

where k_f, k_b, w, m, n are all constants. Such a model behaves like the smoothing Perona–Malik model for weak gradients but becomes a reverse diffusion process for large gradients. They give a stability condition for the one-dimensional case, but the two-dimensional model has observed instabilities [34]. A sophisticated numerical method has been proposed for these

forward/backward diffusion PDEs that is provably stable [34]; however, these stability results are discrete and rely heavily on the relative amounts of smoothing and sharpening. Thus, these PDEs cannot be used for pure sharpening.

3. Preliminaries.

3.1. Gâteaux-differentials and gradients.

Let H be a (possibly infinite-dimensional) Hilbert space, with inner product $\langle \cdot, \cdot \rangle_H$. H is canonically endowed with the structure of a Hilbert manifold, and the tangent space to H at every $u \in H$, denoted by $T_u H$, can be canonically identified with H itself. The dual space $T_u^\star H$ to $T_u H$ is defined to be the space of continuous linear functionals on $T_u H$, and is therefore identified with the dual space to H. The inner product $\langle \cdot, \cdot \rangle_H$ on H also defines (by virtue of identifying $T_u H$ with H) an inner product on $T_u H$ for all $u \in H$ and, as a result, a Riemannian structure on H. We now recall some preliminary definitions regarding Gâteaux-differentials and gradients. In the following, let $E : H \to \mathbb{R}$ be a functional on H and $g : H \times H \to \mathbb{R}$ a bilinear, symmetric, and positive-definite form on H.

Definition 3.1. *We say that E is Gâteaux-differentiable at $u \in H$ if there exists a continuous linear functional $dE|_u : H \to \mathbb{R}$ such that*

$$\lim_{t \to 0} \frac{E(u + tv) - E(u)}{t} = dE|_u(v) \qquad \forall v \in H.$$

$dE|_u$ is called the Gâteaux-differential *of E at u.*

Definition 3.2. *Suppose that $E : H \to \mathbb{R}$ is Gâteaux-differentiable at some $u \in H$. Then if there exists a unique $w \in H$ such that*

$$g(w, v) = dE|_u(v) \qquad \forall v \in H,$$

then we call w the gradient of E at u *with respect to the Riemannian structure induced by g. In this case, we will denote w by $\nabla_g E|_u$.*

Remark 1. If $E : H \to \mathbb{R}$ is Gâteaux-differentiable at $u \in H$ and H is a Hilbert space (i.e., complete) with respect to the inner product induced by g, then the existence and uniqueness of $\nabla_g E|_u$ follows from the Riesz representation theorem. If H endowed with g is not a Hilbert space, then E being Gâteaux-differentiable need not imply the existence of a gradient.

Definition 3.3. *The gradient descent differential equation for E, with respect to the Riemannian structure induced by g, is defined to be the differential equation*

$$u(0) = u_0, \qquad \frac{du}{dt}(t) = -\nabla_g E|_{u(t)}, \quad t > 0, \tag{3.1}$$

where $u_0 \in H$ is a given initial condition.

We define the gradient ascent equation of E in the same way, except for removing the minus sign in (3.1). We interpret solutions to (3.1) in the strong sense; i.e., a solution is a function $u \in C^1([0, T]; H)$ for some $T > 0$ such that $\nabla_g E|_{u(t)}$ exists for all $t \in [0, T]$ and u satisfies the differential equation (3.1). As a remark, note that if u is a solution to (3.1),

then $E \circ u$ is a monotonically decreasing function of t. In fact, given $u_0 \in H$ and $\xi < E[u_0]$, solutions of (3.1) are precisely the shortest paths[1] from u_0 to the level-set $E = \xi$.

From an image processing perspective, the goal is to identify a suitable Hilbert space H of images and a suitable functional E on H; the solution $t \mapsto u(t) \in H$ to the gradient descent equation on E (if it exists) with initial condition $u(0) = u_0$ is then interpreted as a family of images which correspond, with increasing t, to successively "denoised" or "enhanced" versions of the initial image u_0. Again, from the same perspective, what is important is not so much the minimizer of the functional E but rather the one-parameter family defined by the gradient descent equation on E.

3.2. Sobolev spaces of images. Let $\Omega \subset \mathbb{R}^n$ be an open subset of $\mathbb{R}^n$, with smooth boundary $\partial\Omega$, and with coordinate functions $(x_1, \dots, x_n)$; we consider Ω as the domain of our image functions. We define an image (or image function) to be a mapping $u : \Omega \to \mathbb{R}$. For each multi-index $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{N}_0^n$,[2] we define $|\alpha| = \alpha_1 + \cdots + \alpha_n$ and $D^\alpha = D_{x_1}^{\alpha_1} \cdots D_{x_n}^{\alpha_n}$, where D_{x_i} denotes the distributional derivative with respect to x_i. For each $k \in \mathbb{N}_0$, the Sobolev space $H^k(\Omega)$ is defined as

$$H^k(\Omega) = \{u \in L^2(\Omega) \mid D^\alpha u \in L^2(\Omega), \forall \alpha \in \mathbb{N}_0^n \text{ such that } |\alpha| \leq k\}$$

together with the Hilbert space structure induced by the inner product

$$(u, v) \mapsto \langle u, v \rangle_{H^k} = \sum_{|\alpha| \leq k} \langle D^\alpha u, D^\alpha v \rangle_{L^2},$$

where $\langle \cdot, \cdot \rangle_{L^2}$ denotes the usual L^2 inner product. Note that $H^0(\Omega)$ is nothing other than the usual Hilbert space $L^2(\Omega)$.

We denote by $\mathcal{D}(\Omega)$ the vector space of infinitely differentiable compactly supported $\mathbb{R}$-valued functions on Ω. $\mathcal{D}(\Omega)$ is a vector subspace of $H^k(\Omega)$ for every $k \geq 1$, and the closure of $\mathcal{D}(\Omega)$ in $H^k(\Omega)$ is denoted by $H_0^k(\Omega)$. The inner product on $H_0^k(\Omega)$ is taken to be the same as the inner product on $H^k(\Omega)$.

We now consider two distinct Riemannian structures on $H_0^1(\Omega)$:

(a) Consider first the mapping $g_0 : H_0^1(\Omega) \times H_0^1(\Omega) \to \mathbb{R}$ defined by

$$(v, w) \mapsto g_0(v, w) = \langle v, w \rangle_{L^2},$$

where $\langle \cdot, \cdot \rangle_{L^2}$ denotes the L^2 inner product on $H_0^1(\Omega)$. Note that $H_0^1(\Omega)$ is not a Hilbert space (as it is not complete) for the inner product g_0. We shall refer to the metric g_0 as the L^2 metric on $H_0^1(\Omega)$.

(b) Consider now, for each $\lambda > 0$, the mapping $g_\lambda : H_0^1(\Omega) \times H_0^1(\Omega) \to \mathbb{R}$ defined by

$$(v, w) \mapsto g_\lambda(v, w) = (1 - \lambda)\langle v, w \rangle_{L^2} + \lambda \langle v, w \rangle_{H^1}.$$

[1] Recall that the length of a piecewise C^1 curve $\gamma : \mathbb{R} \supset I \to H$ with respect to the Riemannian structure induced by g is defined by

$$\mathcal{L}(\gamma) = \int_I (g(\dot\gamma(s), \dot\gamma(s)))^{1/2} ds.$$

[2] $\mathbb{N}_0$ denotes the nonnegative integers; i.e., $\mathbb{N}_0 = \{0, 1, 2, 3, \dots\}$.

For any $\lambda > 0$, the norm defined on $H_0^1(\Omega)$ by $u \mapsto (g_\lambda(u,u))^{1/2}$ is equivalent to the standard norm on $H_0^1(\Omega)$, and hence the Hilbert space structure induced by g_λ on $H_0^1(\Omega)$ is isomorphic to the standard Hilbert space structure of $H_0^1(\Omega)$. We shall refer to the metric g_λ as the Sobolev metric on $H_0^1(\Omega)$.

Consider now a functional

$$E : H_0^1(\Omega) \to \mathbb{R}$$

and assume it is Gâteaux-differentiable at every $u \in H_0^1(\Omega)$. The gradient of E at $u \in H_0^1(\Omega)$ with respect to the metric g_λ must satisfy

$$g_\lambda(\nabla_{g_\lambda} E|_u, v) = dE|_u(v) \qquad \forall v \in H_0^1(\Omega).$$

Since $H_0^1(\Omega)$ equipped with the inner product g_λ is a Hilbert space, it follows from the Riesz representation theorem that the gradient $\nabla_{g_\lambda} E|_u$ exists and is unique for every $u \in H_0^1(\Omega)$; we shall refer to this gradient as the Sobolev gradient of E. It is important to note that the gradient of E with respect to the L^2 metric g_0 is not defined at all points of $H_0^1(\Omega)$ since $H_0^1(\Omega)$ is not a Hilbert space under the inner product defined by g_0. If for some $u \in H_0^1(\Omega)$ there exists $w \in H_0^1(\Omega)$ such that

$$g_0(w, v) = dE|_u(v) \qquad \forall v \in H_0^1(\Omega),$$

then $\nabla_{g_0} E|_u = w$. Note that, by the definiteness of g_0, the L^2 gradient of E at a point is unique if it exists.

3.3. Hölder spaces. In order to refine certain regularity results, it will be useful to consider Hölder spaces. Let $\Omega \subset \mathbb{R}^n$ be open. For $k \in \mathbb{N}_0$ and $\gamma \in \mathbb{R}$, $0 < \gamma \leq 1$, the Hölder space $C^{k,\gamma}(\bar{\Omega})$ consists of all functions $u \in C^k(\bar{\Omega})$ for which the norm

$$\|u\|_{C^{k,\gamma}(\bar{\Omega})} := \sum_{|\alpha| \leq k} \sup_{x \in \Omega} |D^\alpha u(x)| + \sum_{|\alpha| = k} \sup_{\substack{x,y \in \Omega \\ x \neq y}} \left\{ \frac{|D^\alpha u(x) - D^\alpha u(y)|}{|x - y|^\gamma} \right\}$$

is finite. It is a classical result that, under the above norm, $C^{k,\gamma}(\bar{\Omega})$ is a Banach space. We will also use the notation $BC^k(\bar{\Omega})$ to denote the Banach space of k-times continuously differentiable functions on $\bar{\Omega}$ under the norm

$$\|u\|_{BC^k(\bar{\Omega})} = \sum_{|\alpha| \leq k} \sup_{x \in \Omega} |D^\alpha u(x)|.$$

4. Isotropic diffusion under the L^2 and Sobolev metrics. We consider the functional $E : H_0^1(\Omega) \to \mathbb{R}$ defined by

$$E(u) = \frac{1}{2} \int_\Omega \|\nabla u\|^2 = \frac{1}{2} \int_\Omega (|D_{x_1} u|^2 + \cdots + |D_{x_n} u|^2).$$

E is Gâteaux-differentiable at every $u \in H_0^1(\Omega)$, and $dE|_u$ is the continuous linear functional on $H_0^1(\Omega)$ defined by

$$v \mapsto dE|_u(v) = \int_\Omega \langle \nabla u, \nabla v \rangle.$$

We wish to investigate the differences between gradient descent on the functional E with respect to the L^2 metric and gradient descent on E with respect to the Sobolev metric g_λ.

4.1. Gradient descent on E in the L^2 metric. We recall here the classical result on gradient descent on E with respect to the L^2 metric. The L^2 gradient of E is defined not on all of $H_0^1(\Omega)$ but on the vector subspace $H_0^1(\Omega) \cap H^2(\Omega)$; indeed, for all $u \in H_0^1(\Omega) \cap H^2(\Omega)$ we can write

$$dE|_u(v) = \int_\Omega \langle \nabla u, \nabla v \rangle = \int_\Omega (-\Delta u) v \qquad \forall v \in H_0^1(\Omega). \tag{4.1}$$

The L^2 gradient descent equation on E is the familiar heat equation

$$\begin{cases} \frac{du}{dt} = \Delta u & \text{in } \Omega \times]0, \infty[, \\ u = 0 & \text{on } \partial\Omega \times]0, \infty[, \\ u(x, 0) = u_0(x) & \text{in } \Omega, \end{cases} \tag{4.2}$$

where u_0 denotes the initial condition. Associated with this gradient descent equation, we have the following classical regularity result [6, 11].

Theorem 4.1. *For all $u_0 \in H_0^1(\Omega)$, the solution u to the L^2 gradient descent equation on E is C^∞ on $\bar{\Omega} \times [\varepsilon, \infty[$ for all $\varepsilon > 0$.*

Of course the heat equation is one of the oldest and most familiar PDEs in image processing [3, 36, 35]. The interpretation of this regularity result from an image processing perspective is that the heat equation diffuses the image "instantaneously" and in such a way that all "nonregularities" (e.g., points of discontinuity of the function or its derivatives) are instantaneously removed. It is precisely this regularity result which renders the heat equation ill-posed in the reverse direction [3]. If it were not for this ill-posedness, one could attempt to construct a very simple sharpening algorithm by just running the heat equation backward. This is the premise of Laplacian sharpening (i.e., subtracting a fraction of the Laplacian from an image), but such algorithms introduce unstable oscillations into the image after only a few iterations. As we shall see, the instability present in the reverse heat equation is *not* present when one considers gradient descent/ascent on E with respect to Sobolev metrics.

4.2. Gradient descent on E in the Sobolev metric. The Sobolev gradient descent/ascent equation on E is given by

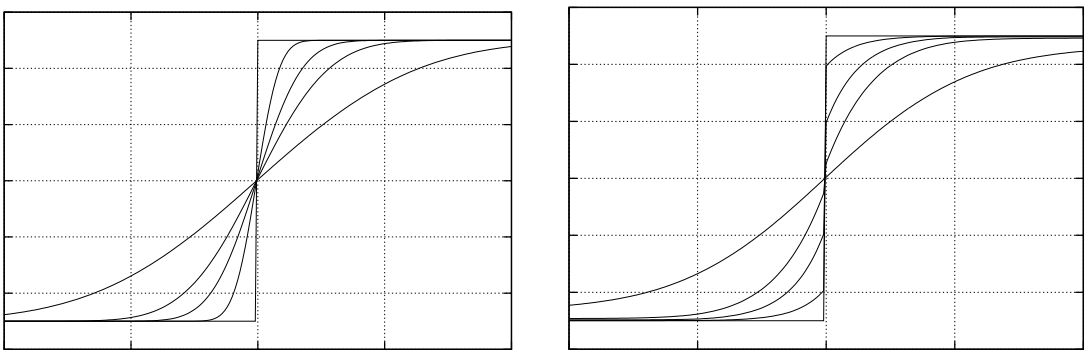
$$u(0) = u_0, \qquad \frac{du}{dt} = -\xi \nabla_{g_\lambda} E|_{u(t)}, \quad t > 0, \tag{4.3}$$

where the Sobolev gradient $\nabla_{g_\lambda} E|_u$ at u satisfies

$$g_\lambda(\nabla_{g_\lambda} E|_u, v) = \int_\Omega \langle \nabla u, \nabla v \rangle \qquad \forall v \in H_0^1(\Omega) \tag{4.4}$$

and $\xi = \pm 1$. By selecting $\xi = 1$ (resp., $\xi = -1$), we obtain the Sobolev gradient descent (resp., ascent) equation on E. Associated with this differential equation, we have the following existence and uniqueness theorem.

Theorem 4.2. *Let $\lambda > 0$, and let $\Omega \subset \mathbb{R}^n$ be an open set. Then for all $u_0 \in H_0^1(\Omega)$ there exists a unique $u \in C^1([0, \infty[; H_0^1(\Omega))$ solving (4.3).*



(a) L^2 diffusion. (b) Sobolev (g_λ) diffusion.

Figure 1. *Sobolev versus L^2 diffusion of a 1D step function. In* (a) *we see that the L^2 equations instantly smooth the discontinuity, whereas in* (b) *the discontinuity remains for finite period of time. This helps to explain why edges persist a little bit longer in 2D Sobolev diffusion.*

Proof. Let $Au := \xi \nabla_{g_\lambda} E|_u$. By the Cauchy–Schwarz inequality and the definition of g_λ, we have

$$\|Au\|_{H_0^1}^2 \leq \frac{1}{\lambda} g_\lambda(Au, Au) = \frac{1}{\lambda} \int_\Omega \langle \nabla u, \nabla(Au) \rangle \leq \frac{1}{\lambda} \|Au\|_{H_0^1} \|u\|_{H_0^1}. \tag{4.5}$$

Hence the mapping defined by $u \mapsto Au$ is a bounded linear operator on $H_0^1(\Omega)$, and the result follows from the classical theory of ODEs [6]. ∎

As a corollary of this theorem, we note that (1) the Sobolev gradient descent equation is an ODE in $H_0^1(\Omega)$, and hence we should not expect the same type of regularity result as for the heat equation (i.e., we should not expect the image to be instantaneously blurred) and, more importantly, (2) the Sobolev gradient ascent equation on E is well-posed, suggesting that it can be used for image sharpening tasks. See Figures 1, 2, and 3 for illustrative examples of the differences between L^2 and Sobolev diffusion on a step function, and some standard test images.

We now intend to explore further properties of (4.3), but first we need a more explicit form for the gradient $\nabla_{g_\lambda} E|_u$.

Proposition 4.3. *For all $u \in H_0^1(\Omega)$, the Sobolev gradient $\nabla_{g_\lambda} E|_u$ is given by*

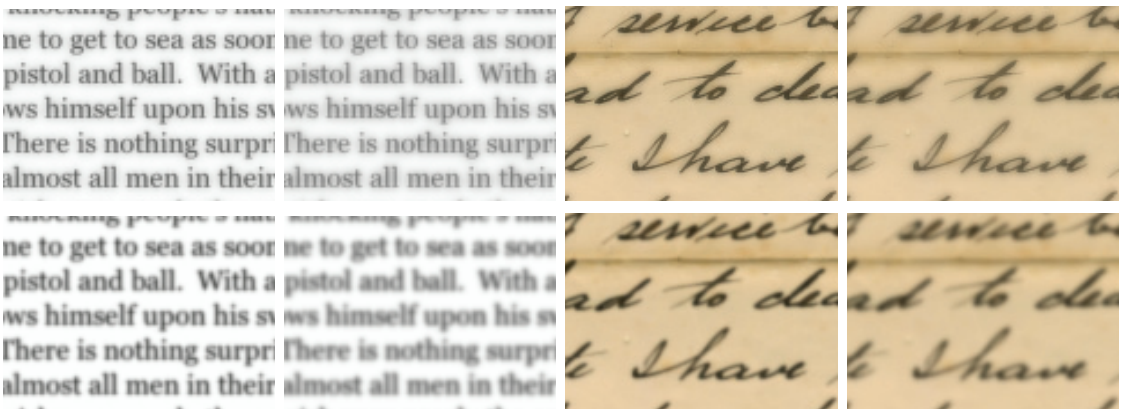
$$\nabla_{g_\lambda} E|_u = -\Delta(I - \lambda\Delta)^{-1} u = \frac{1}{\lambda} \left(I - (I - \lambda\Delta)^{-1} \right) u, \tag{4.6}$$

where I denotes the identity operator and $w = (I - \lambda\Delta)^{-1} u \in H_0^1(\Omega)$ is the unique solution of the PDE

$$\begin{cases} w - \lambda \Delta w = u & \text{in } \Omega, \\ w = 0 & \text{on } \partial\Omega. \end{cases} \tag{4.7}$$

(a) 75%. (b) 90%. (c) 75%. (d) 90%.

Figure 2. *Sobolev (top) versus L^2 (bottom) diffusion on Lena and Baboon test images shown when $u \mapsto \int \|\nabla u\|^2$ has decreased by 75% and 90%.*



(a) 75%. (b) 90%. (c) 75%. (d) 90%.

Figure 3. *Sobolev (top) versus L^2 (bottom) diffusion on book and handwriting test images shown when $u \mapsto \int \|\nabla u\|^2$ has decreased by 75% and 90%.*

Proof. It is a classical result [11] that for all $u \in H_0^1(\Omega)$ there exists a unique $w \in H_0^1(\Omega) \cap H^3(\Omega)$ solving (4.7). Furthermore, for any $v \in H_0^1(\Omega)$, integration by parts yields

$$\int_\Omega uv = \int_\Omega (w - \lambda \Delta w)\, v = \int_\Omega wv + \lambda \int_\Omega \langle \nabla w, \nabla v \rangle.$$

Thus for all $v \in H_0^1(\Omega)$ we have

$$g_\lambda \left(\frac{1}{\lambda}(u - w), v \right) = \int_\Omega \frac{1}{\lambda}(u - w)v + \int_\Omega \langle \nabla(u - w), \nabla v \rangle = dE|_u(v). \tag{4.8}$$

Furthermore, for any $u \in H_0^1(\Omega)$ we have

$$\begin{aligned} -\lambda\Delta(I-\lambda\Delta)^{-1}u &= -\lambda\Delta(I-\lambda\Delta)^{-1}u + (I-\lambda\Delta)^{-1}u - (I-\lambda\Delta)^{-1}u \\ &= (I-\lambda\Delta)(I-\lambda\Delta)^{-1}u - (I-\lambda\Delta)^{-1}u \\ &= u - (I-\lambda\Delta)^{-1}u. \qquad \blacksquare \end{aligned}$$

Interestingly, $\nabla_{g_\lambda} E|_u = \Delta(I-\lambda\Delta)^{-1}u$ is well defined for all $u \in L^2(\Omega)$; thus we can consider extending the Sobolev gradient descent PDE (4.3) to initial conditions $u_0 \in L^2(\Omega)$. We consider the PDE

$$u(0) = u_0, \qquad \frac{du}{dt} = \xi\Delta(I-\lambda\Delta)^{-1}u(t), \quad t > 0, \tag{4.9}$$

where again we take $\xi = \pm 1$. We should note that, in considering the Sobolev gradient descent PDE with initial data in $H_0^1(\Omega)$, we have implicitly enforced the boundary condition $u = 0$ on $\partial\Omega \times [0, \infty[$. In generalizing the PDE from $H_0^1(\Omega)$ to $L^2(\Omega)$, we lose any notion of boundary conditions for the PDE (because $\partial\Omega$ has zero Lebesgue measure). We also lose the notion of the PDE representing gradient descent on the functional E (which is not defined for a general $u \in L^2(\Omega)$). However, from an image processing perspective, it is certainly interesting to consider $L^2(\Omega)$ since $H_0^1(\Omega)$ excludes many discontinuous functions which may represent images with sharp edges. For $u_0 \in L^2(\Omega)$ we have the following existence and uniqueness result.

Theorem 4.4. *Let $\lambda > 0$, and let $\Omega \subset \mathbb{R}^n$ be an open set. Then for all $u_0 \in L^2(\Omega)$ there exists a unique $u \in C^1([0,\infty[; L^2(\Omega))$ solving (4.9).*

Proof. Let $D(A) = H_0^1(\Omega) \cap H^2(\Omega)$, and define the operator $A : D(A) \to L^2(\Omega)$ by $Av = -\Delta v$ for every $v \in D(A)$. It is a classical result that A is a maximal monotone operator on the Hilbert space $L^2(\Omega)$ (we refer the interested reader to [5, 6] for an excellent exposition on this topic). It follows from the theory of maximal monotone operators that for all $\lambda > 0$ the operator $\pm\Delta(I-\lambda\Delta)^{-1}$ is a bounded linear operator on $L^2(\Omega)$ with operator norm $\leq \frac{1}{\lambda}$ (the Yosida regularization of $-\Delta$); hence the result follows directly from the classical theory of ODEs [6]. $\blacksquare$

It is worth noting that, since the Sobolev gradient is the Yosida regularization of $-\Delta$, it follows [6] that as $\lambda \to 0$, the solution to the Sobolev diffusion equation (4.9) with $\xi = 1$ converges uniformly on any compact interval $[0, T]$ to the solution of the heat equation (4.2). This suggests that the gradient ascent equation (with $\xi = -1$) may prove to be a useful regularization of the ill-posed reverse heat equation for deblurring and sharpening images.

The previous two theorems show that the Sobolev gradient descent/ascent PDE on E satisfies a regularity preservation property. If $u_0 \in H_0^1(\Omega)$ (resp., $L^2(\Omega)$), then $u(t) \in H_0^1(\Omega)$ (resp., $L^2(\Omega)$) for all $t > 0$. The natural question to ask is whether this property can be extended to initial data with more regularity. Thus we will consider initial data $u_0 \in C^{k,\gamma}(\bar{\Omega})$. For bounded Ω, $C^{k,\gamma}(\bar{\Omega}) \subset L^2(\Omega)$, so Theorem 4.4 gives us an existence and uniqueness result for such initial data; thus the question becomes whether $u(t) \in C^{k,\gamma}(\bar{\Omega})$ for all $t > 0$.

Theorem 4.5. *Let $\lambda > 0$, and let $\Omega \subset \mathbb{R}^n$ be a bounded open set with a smooth boundary $\partial\Omega$. Let $k \in \mathbb{N}_0$, $0 < \gamma < 1$. Then for all $u_0 \in C^{k,\gamma}(\bar{\Omega})$ there exists a unique $u \in C^1([0,\infty[; C^{k,\gamma}(\bar{\Omega}))$ solving (4.9).*

Proof. Let $f \in C^{k,\gamma}(\bar{\Omega})$. It is a classical result from the theory of elliptic PDEs [13] that there exists a unique $u \in C^{k+2,\gamma}(\bar{\Omega})$ satisfying

$$\begin{cases} u - \lambda \Delta u = f & \text{in } \Omega, \\ u = 0 & \text{on } \partial\Omega, \end{cases} \tag{4.10}$$

and

$$\|u\|_{C^{k+2,\gamma}(\bar{\Omega})} \leq C \left(\sup_{x \in \Omega} |u(x)| + \|f\|_{C^{k,\gamma}(\bar{\Omega})} \right),$$

where $C > 0$ can be chosen independently of f. It is also a classical result [6, 13] that (4.10) satisfies the maximum principle $\sup_{x\in\Omega} |u(x)| \leq \sup_{x\in\Omega} |f(x)|$. It follows immediately that $(I - \lambda\Delta)^{-1} : C^{k,\gamma}(\bar{\Omega}) \to C^{k+2,\gamma}(\bar{\Omega})$ is a bounded linear operator. Hence the Sobolev gradient

$$\Delta(I - \lambda\Delta)^{-1} = \frac{1}{\lambda}((I - \lambda\Delta)^{-1} - I) : C^{k,\gamma}(\bar{\Omega}) \to C^{k,\gamma}(\bar{\Omega})$$

is a bounded linear operator. ∎

It follows from Theorem 4.5 that if the initial data, u_0, is continuous, then applying the reverse Sobolev diffusion equations ((4.9) with $\xi = -1$) will *not* lead to discontinuities (or shocks) in the image, contrary to other enhancement PDEs such as shock filters [21].

We now wish to extend these results to unbounded domains, specifically $\Omega = \mathbb{R}^n$. In this case, we can find an explicit form for the operator $(I - \lambda\Delta)^{-1}$.

4.3. Fundamental solution for $\Omega = \mathbb{R}^n$. We now address the problem of finding an explicit form for the operator $(I - \lambda\Delta)^{-1}$ for the case $\Omega = \mathbb{R}^n$. The fundamental solution S_λ of the PDE $u - \lambda\Delta u = f$ in the domain $\mathbb{R}^n$ is a rescaled Bessel potential (see section 9.1 of the appendix),

$$S_\lambda(x) = \frac{1}{(4\lambda\pi)^{n/2}} \int_0^\infty \frac{e^{-t - \frac{|x|^2}{4t\lambda}}}{t^{n/2}} dt. \tag{4.11}$$

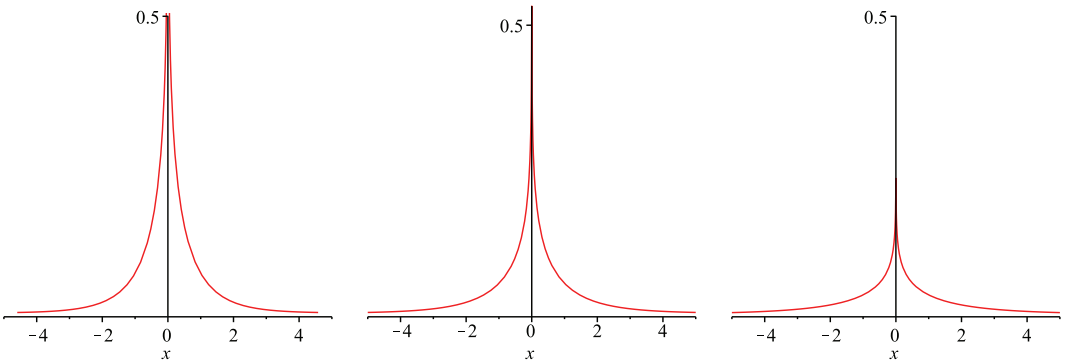
Figure 4 depicts graphs of S_λ for various values of λ. Thus we can express $(I - \lambda\Delta)^{-1} f$ as

$$(I - \lambda\Delta)^{-1} f = \frac{1}{(4\lambda\pi)^{n/2}} \int_{\mathbb{R}^n} \left(\int_0^\infty \frac{e^{-t - \frac{|x|^2}{4t\lambda}}}{t^{n/2}} dt \right) f(y - x) dx. \tag{4.12}$$

In the special case of $n = 1$, S_λ simplifies to

$$S_\lambda(x) = \frac{1}{\sqrt{4\lambda}} e^{-\frac{|x|}{\sqrt{\lambda}}}. \tag{4.13}$$

It is immediate that S_λ is a positive averaging kernel (i.e., $S_\lambda > 0$ and $\int_{\mathbb{R}^n} S_\lambda = 1$).



(a) $\lambda = 1$. (b) $\lambda = 2$. (c) $\lambda = 5$.

Figure 4. *Fundamental solution S_λ in $\mathbb{R}^2$ restricted to one axis for various values of λ. Note the singularity at the origin.*

4.4. Existence and uniqueness for $\Omega = \mathbb{R}^n$. Now that we have a more explicit form for the operator $(I - \lambda\Delta)^{-1}$, we can extend the existence and uniqueness theorems to the special case $\Omega = \mathbb{R}^n$. Also, since $S_\lambda \in L^1(\mathbb{R}^n)$, the convolution form $(I - \lambda\Delta)^{-1}u = S_\lambda * u$ is valid for any $u \in L^p(\mathbb{R}^n)$; thus we can extend $(I - \lambda\Delta)^{-1}$ to a linear operator on $L^p(\mathbb{R}^n)$. For this case, we have the following existence and uniqueness result.

Theorem 4.6. *Let $\lambda > 0$, $p \le 1 \le \infty$, and $k \in \mathbb{N}_0$. Then the following hold:*

(i) *For all $u_0 \in BC^k(\mathbb{R}^n)$ there exists a unique $u \in C^1([0,\infty[; BC^k(\mathbb{R}^n))$ solving (4.9).*

(ii) *For all $u_0 \in L^p(\mathbb{R}^n)$ there exists a unique $u \in C^1([0,\infty[; L^p(\mathbb{R}^n))$ solving (4.9).*

Proof. In each case, it is enough to show that $A_\lambda := \frac{1}{\lambda}((I - \lambda\Delta)^{-1} - I)$ is a bounded linear operator on the appropriate Banach space.

(i) Let $u \in BC^k(\mathbb{R}^n)$. Using the convolution form for the operator $(I - \lambda\Delta)^{-1}$, we see that

$$A_\lambda u = \frac{1}{\lambda} S_\lambda * u - \frac{1}{\lambda} u. \tag{4.14}$$

Since $S_\lambda \in L^1(\mathbb{R}^n)$ and $u \in BC^k(\mathbb{R}^n)$ we see that $S_\lambda * u \in BC^k(\mathbb{R}^n)$ and therefore $A_\lambda : BC^k(\mathbb{R}^n) \to BC^k(\mathbb{R}^n)$ is well defined. Now, for any multi-index α, $|\alpha| \le k$, we have

$$D^\alpha(A_\lambda u) = \frac{1}{\lambda} S_\lambda * D^\alpha u - \frac{1}{\lambda} D^\alpha u.$$

Since $\int_{\mathbb{R}^n} S_\lambda = 1$ we have

$$\|A_\lambda u\|_{BC^k(\mathbb{R}^n)} = \sum_{|\alpha| \le k} \sup |D^\alpha(A_\lambda u)| \le \frac{2}{\lambda} \|u\|_{BC^k(\mathbb{R}^n)}. \tag{4.15}$$

(ii) Let $u \in L^p(\mathbb{R}^n)$. It is a standard result that since $S_\lambda \in L^1(\mathbb{R}^n)$ and $u \in L^p(\mathbb{R}^n)$ we have $S_\lambda * u \in L^p(\mathbb{R}^n)$ and $\|S_\lambda * u\|_{L^p(\mathbb{R}^n)} \le \|S_\lambda\|_{L^1(\mathbb{R}^n)} \|u\|_{L^p(\mathbb{R}^n)}$; therefore $A_\lambda : L^p(\mathbb{R}^n) \to L^p(\mathbb{R}^n)$ is well defined, and since $\int_{\mathbb{R}^n} S_\lambda = 1$ we have

$$\|A_\lambda u\|_{L^p(\mathbb{R}^n)} = \frac{1}{\lambda} \|S_\lambda * u - u\|_{L^p(\mathbb{R}^n)} \le \frac{2}{\lambda} \|u\|_{L^p(\mathbb{R}^n)}. \tag{4.16}$$

■

4.5. Maximum principles. We now investigate maximum principles for the Sobolev diffusion PDE (4.9). Since both the L^2 and Sobolev gradient descent PDE are minimizing the same functional, it is reasonable to suspect they satisfy the same type of maximum principle.

4.5.1. Maximum principles for bounded Ω. We begin with a maximum principle for bounded domains $\Omega \subset \mathbb{R}^n$.

Theorem 4.7. *Let $\Omega \subset \mathbb{R}^n$ be a bounded open set and $0 < \gamma < 1$. Let $u_0 \in C^{0,\gamma}(\bar{\Omega}) \cap H_0^1(\Omega)$, and let $u \in C^1([0,\infty[; C^{0,\gamma}(\bar{\Omega})) \cap C^1([0,\infty[; H_0^1(\Omega))$ be the unique solution to (4.9). Then for all $(x,t) \in \Omega \times [0,\infty[$*

$$\min_{y\in\Omega} u_0(y) \leq u(x,t) \leq \max_{y\in\Omega} u_0(y).$$

Proof. Fix $T > 0$, and define $g(t) = \max_{x\in\bar{\Omega}} u(x,t)$ for $t \in [0,T]$. Since $u \in C^1([0,\infty[; C^{0,\gamma}(\bar{\Omega})) \subset Lip([0,T]; BC^0(\bar{\Omega}))$, it follows that g is Lipschitz continuous on $[0,T]$ and hence differentiable a.e. Let $t \in [0,T]$ be a point of differentiability of g, and let $x_0 \in \bar{\Omega}$ be such that $u(x_0,t) = g(t)$. It follows from the maximum principles for elliptic operators [13, 6] that

$$\frac{\partial u}{\partial t}(x_0,t) = \frac{1}{\lambda}\left((I-\lambda\Delta)^{-1}u - u\right)(x_0,t) \leq 0.$$

Thus, we have

$$g'(t) = \lim_{h\to 0^+} \frac{1}{h}\left(\max_{x\in\bar{\Omega}} u(x,t) - \max_{x\in\bar{\Omega}} u(x,t-h)\right) \leq \lim_{h\to 0^+} \frac{u(x_0,t) - u(x_0,t-h)}{h} \leq 0.$$

Therefore $g' \leq 0$ a.e., and hence $g(t) - g(0) = \int_0^t g'(s)ds \leq 0$ for all $t \in [0,T]$. To show that $u(x,t) \geq \min_{y\in\Omega} u_0(y)$, we can apply the same reasoning to the Sobolev diffusion PDE with initial condition $-u_0$. ∎

The previous theorem also implies that the reverse Sobolev diffusion equation (with $\xi = -1$) cannot satisfy any maximum principle. In fact, in this case, the maximum of $u(x,t)$ over all $x \in \Omega$ will be a nondecreasing function of t, contrary again to other enhancement algorithms such as the shock filter [21], whose discretization satisfies a maximum principle. We now use Theorem 4.7 in order to extend the maximum principle to the more general case of initial conditions in $L^\infty(\Omega)$.

Theorem 4.8. *Let $\Omega \subset \mathbb{R}^n$ be a bounded open subset with a smooth boundary $\partial\Omega$. Let $u_0 \in L^\infty(\Omega) \subset L^2(\Omega)$ such that*

$$\lim_{\varepsilon\to 0^+} \|u\|_{L^\infty(\Omega\setminus\Omega_\varepsilon)} = 0, \tag{4.17}$$

where $\Omega_\varepsilon := \{x \in \Omega \mid \operatorname{dist}(x,\partial\Omega) > \varepsilon\}$, and let $u \in C^1([0,\infty[; L^2(\Omega))$ be the unique solution to (4.9) with initial condition u_0. Then for each $t > 0$

$$\operatorname*{ess\,inf}_{y\in\Omega} u_0(y) \leq u(x,t) \leq \operatorname*{ess\,sup}_{y\in\Omega} u_0(y) \quad \text{a.e. in } \Omega.$$

Proof. Let $\varepsilon > 0$, and define the cutoff function $\zeta_\varepsilon \in C_c^\infty(\mathbb{R}^n)$ such that $0 \leq \zeta_\varepsilon \leq 1$, $\zeta_\varepsilon(x) = 1$ for all $x \in \Omega_\varepsilon$, and $\zeta_\varepsilon(x) = 0$ for all $x \notin \Omega$. Let $u_0^\varepsilon = \zeta_\varepsilon(x)\,(\eta_\varepsilon * \tilde{u}_0)\,(x)$ be the

truncated mollification (see appendix 9.2) of u_0, where $\tilde{u}_0(x) = u_0(x)$ for $x \in \Omega$, and $\tilde{u}_0(x) = 0$ for $x \notin \Omega$. Now, for any $x \in \Omega_\varepsilon$, we have

$$u_0^\varepsilon(x) = \int_{\mathbb{R}^n} \eta_\varepsilon(x-y)\tilde{u}_0(y)dy \leq \operatorname*{ess\,sup}_{y\in\mathbb{R}^n} \tilde{u}_0(y) = \operatorname*{ess\,sup}_{y\in\Omega} u_0(y),$$

and for any $x \in \Omega\backslash\Omega_\varepsilon$ we have

$$|u_0^\varepsilon(x)| \leq |\zeta_\varepsilon(x)| \int_{\Omega\backslash\Omega_{2\varepsilon}} \eta_\varepsilon(x-y)|\tilde{u}_0(y)|dy \leq \|u_0\|_{L^\infty(\Omega\backslash\Omega_{2\varepsilon})}.$$

Thus, for any $x \in \Omega$, we have the bound

$$u_0^\varepsilon(x) \leq \max\left(\|u\|_{L^\infty(\Omega\backslash\Omega_{2\varepsilon})}, \operatorname*{ess\,sup}_{y\in\Omega} u_0(y)\right). \tag{4.18}$$

Let $u^\varepsilon \in C^1([0,\infty[; C^{0,\gamma}(\bar{\Omega}))$ be the solution to the Sobolev PDE (4.9) with initial condition $u_0^\varepsilon \in H_0^1(\Omega) \cap C^\infty(\Omega)$. As $\varepsilon \to 0^+$, we have that $u_0^\varepsilon \to u_0$ in L^2, and that $u^\varepsilon(t) \to u(t)$ a.e. in Ω. Thus, we can employ (4.18) and Theorem 4.7 to obtain

$$u(x,t) = \lim_{\varepsilon\to 0^+} u^\varepsilon(x,t) \leq \lim_{\varepsilon\to 0^+} \sup_{y\in\Omega} u_0^\varepsilon(y) \leq \operatorname*{ess\,sup}_{y\in\Omega} u_0(y)$$

for almost every $x \in \Omega$. To obtain the opposite direction, we can apply the same reasoning to the Sobolev PDE with initial condition $-u_0$. ∎

Remark 2. Alternatively, one can prove Theorems 4.7 and 4.8 by considering the fact that the Sobolev gradient $\Delta(I-\lambda\Delta)^{-1}$ is the L^2 gradient of the Moreau–Yosida regularization [10] of $E(u) = \int \|\nabla u\|^2$,

$$E_\lambda(u) = \inf_{v\in H_0^1(\Omega)} \int_\Omega \|\nabla v\|^2\, dx + \frac{1}{\lambda}\|u-v\|^2_{L^2(\Omega)},$$

and using the fact that E_λ decreases by positive truncation (i.e., $E_\lambda(\min(u,k)) \leq E_\lambda(u)$ for $k > 0$). This approach can also be used to show that the Sobolev flows are semigroup contractions on $L^p(\Omega)$.

4.5.2. Maximum principles for $\Omega = \mathbb{R}^n$. We now extend Theorems 4.7 and 4.8 to the case where $\Omega = \mathbb{R}^n$.

Theorem 4.9. *Let $u_0 \in BC^0(\mathbb{R}^n)$, and let $u \in C^1([0,\infty[; BC^0(\mathbb{R}^n))$ be the unique solution to (4.9). Then for all $(x,t) \in \mathbb{R}^n \times [0,\infty[$*

$$\inf_{y\in\mathbb{R}^n} u_0(y) \leq u(x,t) \leq \sup_{y\in\mathbb{R}^n} u_0(y).$$

Proof. Let $g(t) = \sup_{x\in\mathbb{R}^n} u(x,t)$. Then, as in Theorem 4.7, we have that

$$\frac{\partial u}{\partial t}(x,t) = \frac{1}{\lambda}\left((I-\lambda\Delta)^{-1}u - u\right)(x,t) \leq \frac{1}{\lambda}(g(t) - u(x,t)) \tag{4.19}$$

for any $(x,t) \in \Omega \times [0,\infty[$. Now fix $t > 0$, $\varepsilon > 0$, and let $x_k \in \mathbb{R}^n$ be a sequence satisfying $g(t) < u(x_k,t) + \frac{\varepsilon}{k}$ for all $k \geq 1$. In particular, this yields the estimates $\frac{\partial u}{\partial t}(x_k,t) < \frac{\varepsilon}{k\lambda}$ and

$$\frac{g(t)-g(t-h)}{h} = \frac{1}{h}\left(g(t) - \sup_{x\in\mathbb{R}^n} u(x,t-h)\right) \leq \frac{u(x_k,t)-u(x_k,t-h)}{h} + \frac{\varepsilon}{kh}.$$

Now, since $u \in C^1([0,\infty[; BC^0(\Omega))$, we may choose $\delta > 0$ so that for all $0 < h < \delta$ and $x \in \mathbb{R}^n$ we have

$$\frac{u(x,t)-u(x,t-h)}{h} \leq \frac{\partial u}{\partial t}(x,t) + \varepsilon.$$

Taking the limit as $k \to \infty$, we obtain

$$\frac{g(t)-g(t-h)}{h} \leq \liminf_{k\to\infty}\left(\frac{\partial u}{\partial t}(x_k,t) + \frac{\varepsilon}{kh} + \varepsilon\right) = \varepsilon$$

for all $0 < h < \delta$. Taking $h \to 0^+$, we obtain that $g'(t) \leq 0$ at any point t where g is differentiable. As in Theorem 4.7, since g is Lipschitz, this implies that $g(t) \leq g(0)$ for all $t > 0$. To show that $\inf_{y\in\mathbb{R}^n} u_0(y) \leq u(x,t)$, we can apply the same reasoning to the Sobolev diffusion PDE with initial condition $-u_0$. ∎

The previous theorem can be extended to initial data $u_0 \in L^\infty(\mathbb{R}^n)$.

Theorem 4.10. *Let $u_0 \in L^\infty(\mathbb{R}^n)$, and let $u \in C^1([0,\infty[; L^\infty(\mathbb{R}^n))$ be the unique solution to (4.9). Then for each $t > 0$*

$$\operatorname*{ess\,inf}_{y\in\mathbb{R}^n} u_0(y) \leq u(x,t) \leq \operatorname*{ess\,sup}_{y\in\mathbb{R}^n} u_0(y) \quad \text{a.e. in } \mathbb{R}^n.$$

Proof. Let $\varepsilon > 0$, and let $u_0^\varepsilon = \eta_\varepsilon * u_0$ be the mollification of u_0 (see appendix 9.2). Then, for any $x \in \mathbb{R}^n$ we have

$$u_0^\varepsilon(x) = \int_{\mathbb{R}^n} \eta_\varepsilon(x-y)u_0(y)dy \leq \operatorname*{ess\,sup}_{y\in\mathbb{R}^n} u_0(y).$$

Let $u^\varepsilon \in C^1([0,\infty[; BC^0(\mathbb{R}^n))$ be the solution to (4.9) with initial condition u_0^ε, and fix $t > 0$. Then, as in Theorem 4.8, as $\varepsilon \to 0^+$, we have $u^\varepsilon(t) \to u(t)$ a.e., and so it follows from Theorem 4.9 that

$$u(x,t) = \lim_{\varepsilon\to 0^+} u^\varepsilon(x,t) \leq \lim_{\varepsilon\to 0^+} \sup_{y\in\mathbb{R}^n} u_0^\varepsilon(y) \leq \operatorname*{ess\,sup}_{y\in\mathbb{R}^n} u_0(y)$$

for almost every $x \in \mathbb{R}^n$. To obtain that $\operatorname{ess\,inf}_{y\in\mathbb{R}^n} u_0(y) \leq u(x,t)$, we can apply the same reasoning to the Sobolev diffusion PDE with initial condition $-u_0$. ∎

5. Applications of Sobolev diffusion.

5.1. Image sharpening. Due to the reversibility of the Sobolev diffusion equations, we can explore new techniques for image sharpening that were previously impossible with the L^2 diffusion equations. The most basic approach would be to run the Sobolev diffusion equations in the backward direction on a given image u_0, and at some stopping time, say $\tau > 0$, to declare $u(\tau)$ to be the sharpened version of u_0. However, it is not clear in the approach how

one would appropriately select the stopping time $\tau > 0$, as it would necessarily vary between images and application domains.

We will consider a slightly more sophisticated approach in which the stopping time is encoded into the functional. Since $\int_\Omega \|\nabla u\|^2$ is a measure of the "sharpness" of an image u, it is natural to use this term as a stopping condition. Hence consider the functional

$$u \mapsto E[u; u_0] = \frac{1}{4}\left(\int_\Omega ||\nabla u_0||^2\right)\left(\frac{\int_\Omega ||\nabla u||^2}{\int_\Omega ||\nabla u_0||^2} - \alpha\right)^2, \tag{5.1}$$

where $\alpha \in \mathbb{R}$ is the sharpness factor. For $\alpha < 1$ we get blurring, and for $\alpha > 1$ we get sharpening. The gradient descent PDE for the above functional with respect to the Sobolev metric is

$$u(0) = u_0, \qquad \frac{du}{dt} = \left(\frac{\int_\Omega ||\nabla u(t)||^2}{\int_\Omega ||\nabla u_0||^2} - \alpha\right)\Delta(I - \Delta)^{-1}u(t), \quad t > 0. \tag{5.2}$$

Thus we have forward or reverse diffusion until the ratio of $\int_\Omega ||\nabla u||^2$ to $\int_\Omega ||\nabla u_0||^2$ is α, so this is really equivalent to selecting a stopping time for the forward or reverse Sobolev diffusion equation, where this stopping time is encoded into the sharpness factor α. We should note that for $\alpha > 1$ the L^2 gradient descent equation on (5.1) is ill-posed, so we can perform this kind of sharpening *only* with our proposed Sobolev gradient descent framework.

For any value of α this functional admits stable minima, so there is no opportunity for unbounded gradient flow. We should note that there are infinitely many u satisfying $\int_\Omega ||\nabla u||^2 / \int_\Omega ||\nabla u_0||^2 = \alpha$, many of which we would not consider as sharpened versions of u_0. However, since we are considering a gradient descent flow, we will find the *closest* local minimizer to u_0 (in the H^1 norm) of (5.1). Thus, just by using a gradient descent flow, we are implicitly incorporating a data fidelity term into the sharpening process.

Next, we consider adding a data fidelity term explicitly to the heat equation functional. The most natural data fidelity term is the L^2 distance, which yields

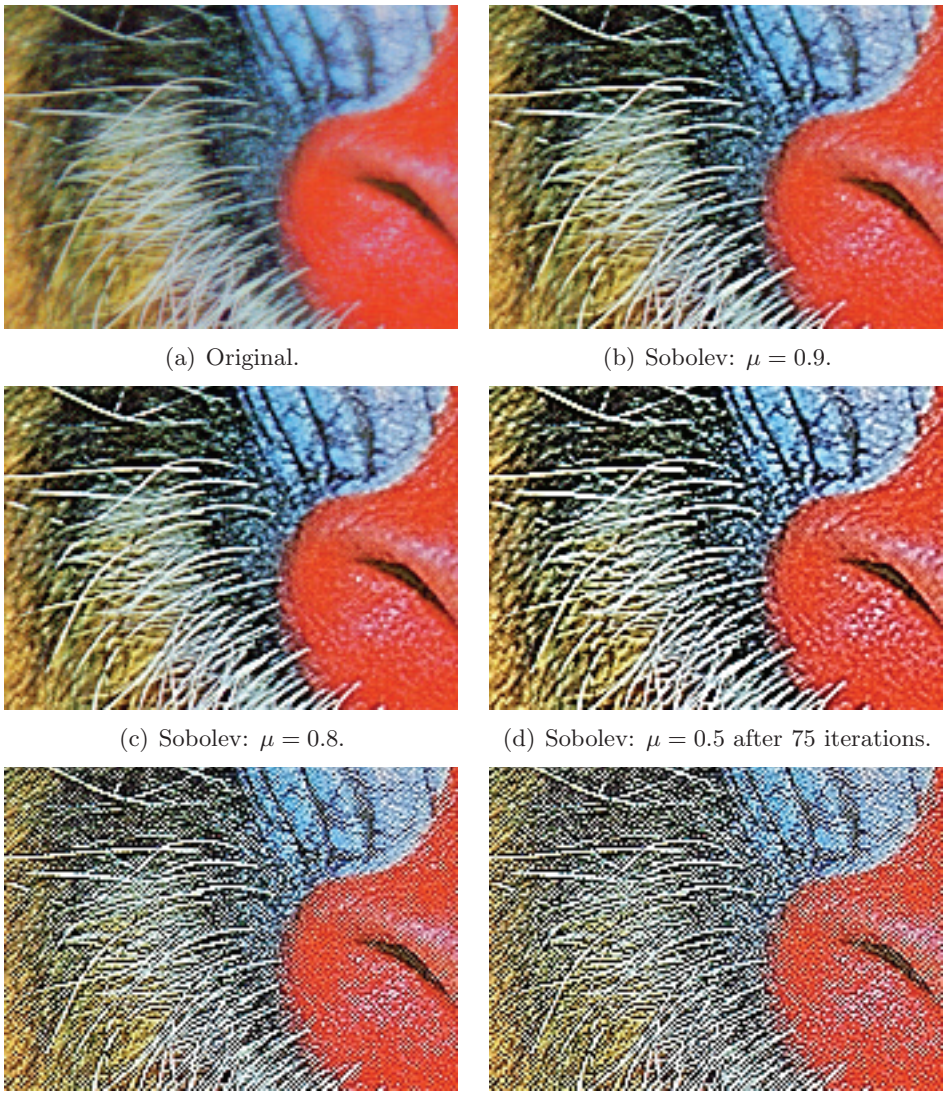
$$u \mapsto E[u; u_0] = -\frac{1}{2}\int_\Omega ||\nabla u||^2 dx + \frac{\mu}{2}||u - u_0||_{L^2}^2, \tag{5.3}$$

where u_0 is the initial image. However, (5.3) is not coercive for any $\mu > 0$, and hence we cannot expect it to admit a stable minimizer. Hence we must consider the H^1 distance for the data fidelity term, which yields the functional

$$u \mapsto E[u; u_0] = -\frac{1}{2}\int_\Omega ||\nabla u||^2 dx + \frac{\mu}{2}||u - u_0||_{H^1}^2. \tag{5.4}$$

For $\mu > 1$, (5.4) is coercive and hence admits a minimizer. Any minimizer of (5.4) represents a sharpened version of the initial image u_0. If we were to seek this minimizer via an L^2 gradient descent PDE, we would obtain

$$u(0) = u_0, \qquad \frac{du}{dt} = (\mu - 1)\Delta u(t) - \mu\Delta u_0 + \mu(u_0 - u(t)), \quad t > 0. \tag{5.5}$$



(a) Original. (b) Sobolev: $\mu = 0.9$.

(c) Sobolev: $\mu = 0.8$. (d) Sobolev: $\mu = 0.5$ after 75 iterations.

(e) L^2: $\mu = 0.9$. (f) L^2: $\mu = 0.8$ after 30 iterations.

Figure 5. *Stable Sobolev sharpening vs. unstable L^2 sharpening.* (b), (c) *For $\mu = 0.9, 0.8$ the discretized gradient descent PDE* (5.6) *converges to a local minimum (though the true continuous PDE would not).* (d) *For $\mu = 0.5$ the PDE does not converge but does yield a family of sequentially sharper images. In* (e) *and* (f), *we see that the L^2 gradient descent PDE very quickly exhibits instabilities and oscillatory behavior (diffusion was performed on the luminance component only).*

However, as seen in Figure 5, (5.5) immediately introduces unstable oscillations into the image and hence cannot be used for well-posed image sharpening.

Such instability can be avoided by considering gradient descent on (5.4) with respect to the Sobolev metric, which yields

$$u(0) = u_0, \qquad \frac{du}{dt} = -\Delta(I - \lambda\Delta)^{-1}u(t) + \mu(u_0 - u(t)), \quad t > 0. \tag{5.6}$$

Again, we expect a stable minimum for $\mu > 1$. For $\mu < 1$ there is indeed no stable minimum; however, the resulting gradient descent is a well-posed PDE which results in a family of images that are sequentially sharper. In this way, the Sobolev gradient offers much more than an improvement on the L^2 gradient, opening up options for image sharpening that were not even possible under the L^2 gradient. See section 7 for image sharpening experiments.

5.2. Connections between Sobolev sharpening and removing Gaussian blur. It is well known that evolving the heat equation in the forward direction for time t is equivalent to convolution with a Gaussian kernel of variance $2t$. It is also well known that this process is ill-posed in the reverse direction, meaning that we cannot deblur Gaussian blur by simply running the heat equation backward. However, the Sobolev diffusion equations are well-posed in the reverse direction, so the natural question to ask is: Is reverse Sobolev diffusion good at deblurring Gaussian blur?

To answer this question, we examine Sobolev diffusion in the Fourier domain. Suppose that we have an image u_0 which is then convolved with a Gaussian kernel $G_\sigma(x) := (2\pi\sigma^2)^{-n/2}\exp(-|x|^2/2\sigma^2)$. The convolution $u_b = G_\sigma * u_0$ is represented in the Fourier domain via the convolution theorem as

$$\hat{u}_b(\omega) = \exp\left(-\frac{\sigma^2}{2}|\omega|^2\right)\hat{u}_0(\omega). \tag{5.7}$$

Now set u_b as the initial condition for the reverse Sobolev diffusion PDE. Then by taking Fourier transforms of (4.9), we have

$$\hat{u}(0) = \hat{u}_b, \qquad \frac{d\hat{u}}{dt} = \frac{|\omega|^2}{1+\lambda|\omega|^2}\hat{u}(t), \quad t > 0.$$

Solving this ODE yields

$$\hat{u}(t) = \exp\left(-\frac{\sigma^2}{2}|\omega|^2 + \frac{|\omega|^2}{1+\lambda|\omega|^2}t\right)\hat{u}_0. \tag{5.8}$$

If we select $t = \sigma^2/2$ as our stopping condition, then we obtain

$$\hat{u}\left(\frac{\sigma^2}{2}\right) = \hat{u}_0\exp\left(-\frac{\lambda\sigma^2|\omega|^4}{2(1+\lambda|\omega|^2)}\right). \tag{5.9}$$

So define

$$F_{\sigma,\lambda}(\omega) \triangleq \exp\left(-\frac{\lambda\sigma^2|\omega|^4}{2(1+\lambda|\omega|^2)}\right). \tag{5.10}$$

Then running reverse Sobolev diffusion on a Gaussian blurred image is equivalent to filtering the original image with $F_{\sigma,\lambda}$. Figure 6 shows the profile of $F_{\sigma,\lambda}$ for various values of λ. We can see that $F_{\sigma,\lambda}$ has unity gain at low frequencies, but then it tails off for high frequencies. This means that reverse Sobolev diffusion behaves as an ideal inverse filter for Gaussian blurring at low frequencies while tailing off (for stability) at higher frequencies. Furthermore, we can see that the choice of λ selects the cutoff frequency of $F_{\sigma,\lambda}$. In the results section we compare Sobolev diffusion against the shock filter on Gaussian blurred images.

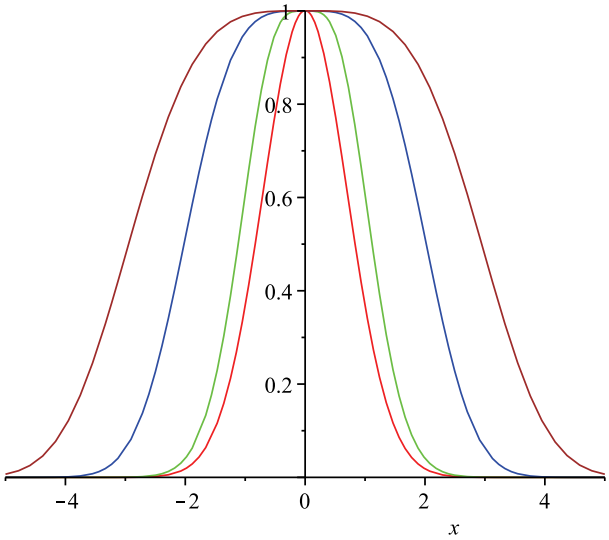


Figure 6. *Plot of the deblurring filter $F_{\sigma,\lambda}$ for $\sigma^2 = 2$ and (from inside to outside) $\lambda = 1000, 1, 0.05, 0.01$.*

5.3. Other possibilities. The well-posedness of the Sobolev gradient both forward and backward to energies involving only first-order image derivatives considered thus far may be further exploited in conjunction with higher order image derivative energies for combinations of simultaneous sharpening and smoothing. In this section we explore a few interesting applications, but many more are possible. The key point is that none of these options are available under the L^2 topology as they all yield unstable gradient flows.

Consider, for example, the functional

$$u \mapsto E[u] = \frac{1}{2}\int_\Omega (\Delta u)^2. \tag{5.11}$$

In order to justify the use of such a functional, note that if $u \in H_0^1(\Omega)$ and $\Omega \subset \mathbb{R}^n$, then we can integrate by parts twice to obtain

$$\int_\Omega (\Delta u)^2 dx = \sum_{i,j=1}^n \int_\Omega u_{x_i x_i} u_{x_j x_j} dx = -\sum_{i,j=1}^n \int_\Omega u_{x_i x_i x_j} u_{x_j} dx = \sum_{i,j=1}^n \int_\Omega (u_{x_i x_j})^2 dx.$$

Indeed, this functional is quite well justified, as it is a measure of the L^2 norm of all second order partial derivatives. The L^2 gradient, $\nabla_{g_0} E$, of the above functional is

$$\nabla_{g_0} E = \Delta^2 u. \tag{5.12}$$

Thus the Sobolev gradient $\nabla_{g_\lambda} E$ is (in a similar spirit to Proposition 4.3)

$$\nabla_{g_\lambda} E = \Delta^2 (I - \lambda\Delta)^{-1} u = \frac{1}{\lambda}(-\Delta u + \Delta(I - \lambda\Delta)^{-1} u).$$

5.3.1. Second order smoothing + sharpening (first order) constraints. Now consider the constrained minimization (second order smoothing) problem

$$\min_{\substack{u\in H_0^1(\Omega):\\ \int_\Omega ||\nabla u||^2=C}} \frac{1}{2}\int_\Omega (\Delta u)^2 \tag{5.13}$$

for some constant $C > 0$. We would like to minimize $E_1[u] := 1/2\int_\Omega(\Delta u)^2$ while keeping $E_2[u] := \int_\Omega ||\nabla u||^2$ constant. We can achieve this by performing gradient descent on E_1 while removing any component of $\nabla_{H^1}E_1$ that is in the direction of $\nabla_{H^1}E_2$, thereby remaining on a level set of E_2. This process yields the PDE

$$u(0) = u_0, \qquad \frac{du}{dt} = -\left(\nabla_{g_\lambda}E_1|_u - \frac{g_\lambda\left(\nabla_{g_\lambda}E_1|_u, \nabla_{g_\lambda}E_2|_u\right)}{g_\lambda\left(\nabla_{g_\lambda}E_2|_u, \nabla_{g_\lambda}E_2|_u\right)}\nabla_{g_\lambda}E_2|_u\right), \quad t > 0. \tag{5.14}$$

Such a constrained minimization would be ill-posed under the L^2 metric; hence this is possible only with a Sobolev gradient descent PDE. As always, one could further temper the second order smoothing by adding a data fidelity term to E_1 so that

$$E_1[u; u_0] = \frac{1}{2}\int_\Omega (\Delta u)^2 dx + \frac{\mu}{2}||u - u_0||^2_{H^1}.$$

We now address issues of numerical implementation and then present experimental results.

6. Numerical implementation. In this section we will describe in detail the numerical implementation methods used to compute the Sobolev gradient $\Delta(I - \lambda\Delta)^{-1}u$. For the choice of the Sobolev inner product g_λ, the image diffusion equations involve a convolution with a kernel S_λ which was determined in the previous sections. For the case $n = 1$, S_λ is a continuous, bounded, and integrable function, and so the usual discrete convolution scheme (derived from the Riemann sum) can be used to approximate $S_\lambda * f$. However, for $n \geq 2$, S_λ has a singularity at $x = 0$; thus one needs to be careful about how this convolution is implemented in order to obtain accurate numerical results without the need to perform any complicated numerical integration during the diffusion process. The idea is to obtain a similar discrete convolution scheme by precomputing the integral of S_λ over a sufficiently fine disjoint covering of the image domain. Such a scheme can then be applied to a convolution kernel S_λ which is merely integrable. Although most applications will be in dimensions $n = 1, 2, 3$, we will describe our numerical scheme for any $n \geq 1$ for the sake of generality.

Lemma 6.1. *Let $u : \mathbb{R}^n \to \mathbb{R}$ be continuous and bounded, and let $g \in L^1(\mathbb{R}^n)$. Suppose that for any $\varepsilon > 0$ we have a disjoint covering $\{A_n^\varepsilon\}_{n\in\mathbb{N}}$ of $\mathbb{R}^n$ by measurable sets satisfying*

$$\sup_{x,y\in A_n^\varepsilon} \|x - y\| < \varepsilon \qquad \forall n \in \mathbb{N}. \tag{6.1}$$

Furthermore, suppose that for each $\varepsilon > 0$ we have a sequence $\{x_n^\varepsilon\}_{n\in\mathbb{N}}$ in $\mathbb{R}^n$ such that $x_n^\varepsilon \in A_n^\varepsilon$ for each $n \in \mathbb{N}$. Then

$$\lim_{\varepsilon\to 0}\sum_{n\in\mathbb{N}} u(x_n^\varepsilon)\int_{A_n^\varepsilon} g(x)dx = \int_{\mathbb{R}^n} u(x)g(x)dx. \tag{6.2}$$

Proof. Let $u_\varepsilon : \mathbb{R}^n \to \mathbb{R}$ be the piecewise constant approximation to u defined by

$$u_\varepsilon(x) = \sum_{n\in\mathbb{N}} \chi_{A_n^\varepsilon}(x) u(x_n^\varepsilon),$$

where χ_A denotes the indicator (or characteristic) function of the set A. We claim that for each $x \in \mathbb{R}^n$, $u_\varepsilon(x) \to u(x)$ as $\varepsilon \to 0$. To see this, fix $x \in \mathbb{R}^n$. Then for each $\varepsilon > 0$ there exists a unique $A^\varepsilon_{n(\varepsilon)}$ with $x \in A^\varepsilon_n(\varepsilon)$, and hence $u_\varepsilon(x) = u(x^\varepsilon_{n(\varepsilon)})$. Since $|x^\varepsilon_{n(\varepsilon)} - x| < \varepsilon$ and u is continuous, we have $x^\varepsilon_{n(\varepsilon)} \to x$ and hence $u_\varepsilon(x) \to u(x)$. Now we have

$$\sum_{n\in\mathbb{N}} u(x_n^\varepsilon) \int_{A_n^\varepsilon} g(x)dx - \int_{\mathbb{R}^n} u(x)g(x)dx = \int_{\mathbb{R}^n} g(x)\left(u_\varepsilon(x) - u(x)\right)dx.$$

Since u is bounded and $g \in L^1(\mathbb{R}^n)$, the dominated convergence theorem yields

$$\lim_{\varepsilon\to 0} \int_{\mathbb{R}^n} g(x)\left(u_\varepsilon(x) - u(x)\right)dx = 0. \qquad \blacksquare$$

In particular, for any continuous and bounded image u and any $y \in \mathbb{R}^n$, we can apply Lemma 6.1 to the translation of u to obtain

$$\lim_{\varepsilon\to 0} \sum_{n\in\mathbb{N}} u(y - x_n^\varepsilon) \int_{A_n^\varepsilon} S_\lambda(x)dx = (u * S_\lambda)(y) = ((I - \lambda\Delta)^{-1}u)(y). \tag{6.3}$$

6.1. Discretization of the convolution kernel S_λ. We now describe the discrete convolution scheme. Take h to denote the sampling period, and assume that all images $u : \mathbb{R}^n \to \mathbb{R}$ are bounded such that $0 \le u(x) \le M$ for some $M \in \mathbb{R}$. The previous lemma allows us to accurately compute the convolution of a continuous function with an L^1 function. However, the discrete sum may have an infinite number of terms since S_λ does not have compact support; thus the need to truncate the support of S_λ. Since $S_\lambda \in L^1(\mathbb{R}^n)$, for any $\varepsilon > 0$ there exists a $W \in \mathbb{N}$ such that

$$0 < \int_{\mathbb{R}^n \setminus U_W} S_\lambda(x)dx < \varepsilon, \tag{6.4}$$

where $U_W := \{x \in \mathbb{R}^n \mid -W - h/2 \le x_i < W + h/2,\ i = 1, \dots, n\}$. Thus we will approximate the convolution $S_\lambda * u$ by

$$(S_\lambda * u)(y) \approx \int_{U_W} S_\lambda(x)u(y - x)dx$$

for some $W > h$ to be chosen later. Now let $A := \{x \in \mathbb{R}^n \mid -h/2 \le x_i < h/2,\ i = 1, \dots, n\}$ be the hypercube of side length h, centered at the origin, and for $y = (y_1, \dots, y_n) \in \mathbb{R}^n$ let

$$A_y := y + A = \{y + x \mid x \in A\}$$

be the translated hypercube. Define

$$V_W := \{z = (z_1, \ldots, z_n) \in \mathbb{Z}^n \mid h|z_i| \leq W + 1,\ i = 1, \ldots, n\}.$$

The sets $\{A_{hz}\}_{z \in V_W}$ form a disjoint covering of U_W, and we have the bound

$$\sup_{x,y \in A_{hz}} \|x - y\| = \sqrt{n}h \qquad \forall z \in V_W.$$

Thus the sets $\{A_{hz}\}_{z \in V_W}$ satisfy the hypothesis of the lemma, and we therefore approximate the convolution $S_\lambda * u$ at a point $y \in \mathbb{R}^n$ by

$$(S_\lambda * u)(y) \approx \sum_{z \in V_W} u(y - hz) \int_{A_{hz}} S_\lambda(x)dx.$$

It follows from (6.4) and Lemma 6.1 that

$$\left| \lim_{h \to 0^+} \sum_{z \in V_W} u(y - hz) \int_{A_{hz}} S_\lambda(x)dx - (S_\lambda * u)(y) \right| < \varepsilon \|u\|_{L^\infty};$$

hence, for bounded images, the discretization error of this scheme can be made arbitrarily small by selecting $h > 0$ small enough and the window of the filter W large enough. However, one of the key properties of S_λ is that it is an averaging kernel (i.e., $\int_{\mathbb{R}^n} S_\lambda = 1$). Since we have truncated the support of S_λ, the coefficients of our discrete convolution kernel will sum to slightly less than 1. For longer simulations (as $t \to \infty$), this introduces a shift in the image as the truncation errors build up iteration after iteration. Therefore, we have decided to include a normalization factor into the kernel. Let $C = \int_{U_W} S_\lambda$, and define

$$K_\lambda(z) = \frac{1}{C} \int_{A_{hz}} S_\lambda(x)dx. \tag{6.5}$$

We obtain the discrete convolution approximation

$$(S_\lambda * u)(hy) \approx \sum_{z \in V_W} u(hy - hz)K_\lambda(z) \quad (y \in \mathbb{Z}^n).$$

For the special case where $n = 2$ and the image sampling period is taken to be $h = 1$, we have

$$K_\lambda(i, j) \triangleq \frac{1}{C} \int_{j-1/2}^{j+1/2} \int_{i-1/2}^{i+1/2} S_\lambda(x, y)dxdy, \quad -W \leq i, j \leq W, \tag{6.6}$$

and the discrete convolution is given by

$$(S_\lambda * u)(k, \ell) = \sum_{j=-W}^{W} \sum_{i=-W}^{W} u(k - i, \ell - j)K_\lambda(i, j) \quad ((k, \ell) \in \mathbb{Z}^2). \tag{6.7}$$

In our experiments, we take $h = 1$, $n = 1, 2$, and $W = 5$. This yields $C = \int_{U_W} S_\lambda > 0.99$. The coefficients $K(i, j)$ can be computed offline and are provided in Figure 7 to facilitate implementation of our method. Equation (6.7) is used to compute the Sobolev gradient in all the experimental results of the next section. Algorithm 1 shows the explicit discretization scheme of the diffusion PDE. Note that we could have used either $\frac{\xi}{\lambda}((I - \lambda\Delta)^{-1}u - u)$ or $\xi\Delta(I - \lambda\Delta)^{-1}u$ for the Sobolev gradient. The numerical advantage of the first form is that it does not involve computing discrete derivatives. At the boundaries of the image, the convolutions are performed by reflecting the image about its boundary. Note that for forward diffusion, a constant time step of $\delta t = 0.1$ may be chosen. For reverse diffusion, since the gradients of the image increase with time, for any prolonged simulations the time step δt should be constrained according to line 7 of Algorithm 1 for numerical stability.

$$10^{-3} \times \begin{bmatrix} 0.61700 & 0.55367 & 0.40523 & 0.24979 & 0.13474 & 0.06585 \\ 1.86521 & 1.62596 & 1.10380 & 0.61740 & 0.30170 & \\ 5.80472 & 4.81757 & 2.90906 & 1.42427 & & \\ 19.01402 & 14.29655 & 7.09452 & & & \\ 69.80074 & 40.13533 & & & & \\ 201.79638 & & & & & \end{bmatrix}$$

Figure 7. *Coefficients of $K_\lambda(i, j)$ for $0 \le i \le j \le 5$ with $\lambda = 1$.* Note that these coefficients are the results of local integrations of the kernel S_λ as per (6.6).

Algorithm 1. Sobolev diffusion in $\mathbb{R}^2$: One iteration.

1: $u_0 :=$ Initial Image
2: $\delta t = 0.1$, $\xi = \pm 1$, $\lambda > 0$, $M = 0$
3: **for** $(k, \ell) \in \Omega$ **do**
4: Compute: $\rho(k, \ell) = \left(\sum_{j=-W}^{W} \sum_{i=-W}^{W} u_n(k - i, \ell - j)K_\lambda(i, j)\right) - u_n(k, \ell)$
5: Set: $M = \max(M, |\rho(k, \ell)|)$
6: **end for**
7: Set: $\delta t = \min(0.1, 2.5/M)$
8: **for** $(k, \ell) \in \Omega$ **do**
9: Update step: $u_{n+1}(k, \ell) = u_n(k, \ell) + \delta t \frac{\xi}{\lambda}\rho(k, \ell)$
10: **end for**

6.2. Error bounds on discretization. Lemma 6.1 is a type of consistency result; it says that as the sampling period tends to zero, the discretization error of our numerical scheme also tends to zero. However, since we must choose some small, but fixed, sampling period, it is useful to obtain explicit error bounds on our discretization scheme as a function of the sampling period. Before we do this, let us be more explicit about the relationship between the image sampling period, denoted by h, and the parameter λ. In image processing, we are given an image $\bar{u} : \mathbb{Z}^n \to \mathbb{R}$, usually defined on a lattice, which is assumed to be a sampled representation of some image $u : \mathbb{R}^n \to \mathbb{R}$. The sampling period is often taken to be $h = 1$, but we are free to alter this, so suppose the sampling period was $h \neq 1$. But since sampling with period $h \neq 1$ must yield the same sampled image $\bar{u}$ that we have been given, we should

view $\bar{u}$ as the image $u_h(x) \triangleq u(x/h)$ sampled with period h. Then it is immediate that

$$\begin{aligned}(S_\lambda * u_h)(y) &= \int_{\mathbb{R}^n} S_\lambda(x) u\left(\frac{y-x}{h}\right) dx \\ &= \int_{\mathbb{R}^n} h^n S_\lambda(hx) u\left(\frac{y}{h} - x\right) dx \\ &= (S_{\lambda/h^2} * u)\left(\frac{y}{h}\right).\end{aligned}$$

This tells us that changing the sampling period from $1 \to h$ is equivalent to changing the parameter of S_λ from $\lambda \to \lambda/h^2$.

Hence we will fix the sampling period at $h = 1$ and study the effect of changing λ on the discretization error. Since S_λ is a continuous function away from the origin, the case of interest is the discretization cell which contains the singularity of S_λ at the origin. Let $B := \{(x, y) \in \mathbb{R}^n \mid -0.5 \le x, y \le 0.5\}$ be the discretization cell containing the origin (with respect to a sampling period $h = 1$). Then, given an image $u : \mathbb{R}^n \to \mathbb{R}$, the discretization error introduced into the computation of the convolution $S_\lambda * u$ by replacing S_λ by its integral over B is given by

$$\left| \int_B S_\lambda(x) u(x) dx - \int_B u(x) \int_B S_\lambda(z) dz dx \right| \le \|u\|_{L^\infty} \int_B \left| S_\lambda(x) - \int_B S_\lambda(z) dz \right| dx.$$

If we assume that all images of interest are bounded uniformly (i.e., $\|u\|_{L^\infty} \le M$ for some $M > 0$), then we should study the error term

$$V(\lambda) \triangleq \int_B \left| S_\lambda(x) - \int_B S_\lambda(z) dz \right| dx. \tag{6.8}$$

$V(\lambda)$ is difficult to work with analytically, so we have numerically computed $V(\lambda)$ for various values of λ, and a plot is shown in Figure 8. Since $\lambda \to \infty$ is equivalent to $h \to 0^+$, Lemma 6.1 guarantees that $V(\lambda) \to 0$ as $\lambda \to \infty$, so this property of Figure 8 is not surprising. What is useful is the rate of convergence, as it can give us information about how to choose the parameter λ. We can see that a choice of $\lambda = 1$ yields a very accurate numerical scheme while keeping the essential support of S_λ small enough (an 11×11 window) that the numerical method is fast. Although we can increase λ to reduce the discretization error $V(\lambda)$ even further, we are at the point on the graph of $V(\lambda)$ where we have a diminishing rate of return, so each additional percentage point of improvement in $V(\lambda)$ comes at an increasing computational cost.

7. Experimental results and applications to image sharpening.

7.1. Reversibility of the Sobolev diffusion equations. It is well known that the reverse L^2 diffusion equation is ill-posed; thus this is one of the *key* differences between the Sobolev diffusion equations and the L^2 diffusion equations. As we have already noted, the reverse Sobolev diffusion equation is well-posed; thus we can perform forward Sobolev diffusion followed by reverse Sobolev diffusion and theoretically arrive back at our original image! Figure 9 experimentally validates this assertion; given an initial image, we can run the Sobolev diffusion

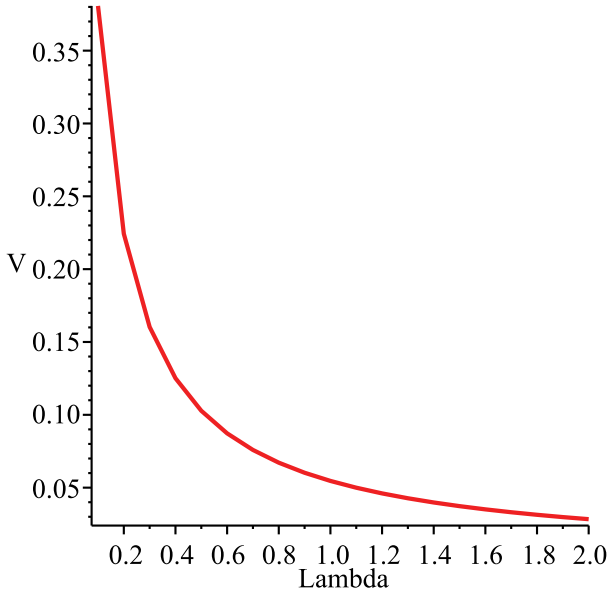


Figure 8. *Plot of λ versus the discretization error introduced by averaging the kernel S_λ about its singularity. Note that $V(1) = 0.0546$.*

(a) Original image.

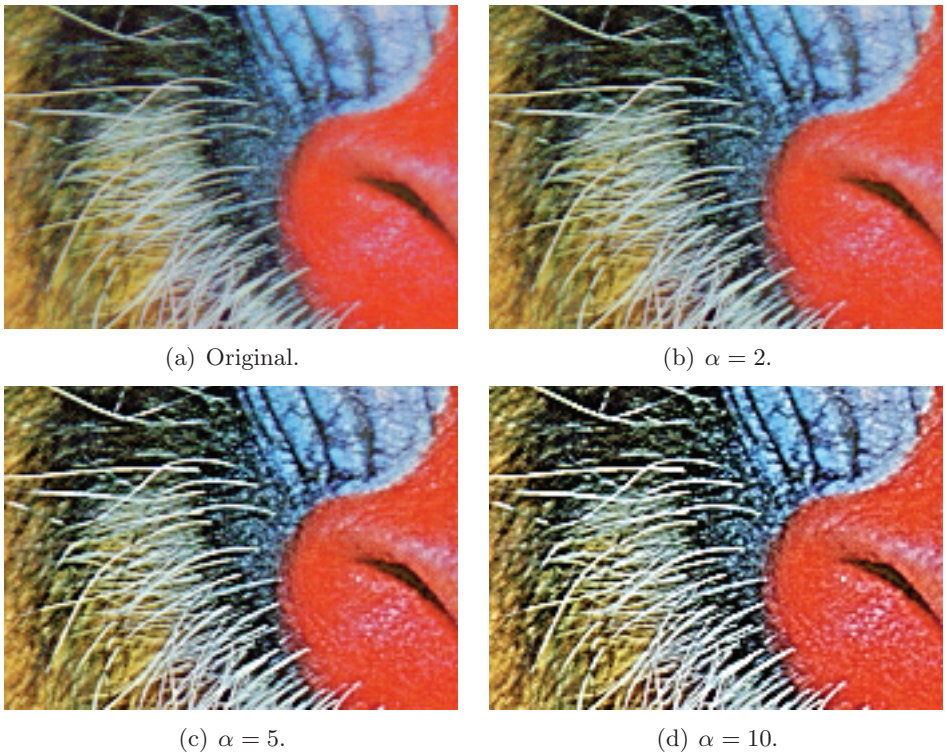
(b) Forward Sobolev diffusion: 40 steps.

(c) Reversed Sobolev diffusion: 40 steps.

Figure 9. *Reversibility of the Sobolev diffusion equations.* (a) *Original image,* (b) *results after* 40 *iterations of forward Sobolev diffusion, and* (c) *results after* 40 *follow-up iterations of reverse Sobolev diffusion (diffusions performed on each color component, Y, Cb, Cr, with $\delta t = 0.1$).*

equations forward (for 40 iterations) to introduce a certain amount of smoothing, and then we can reverse the equations (for 40 more iterations) to recover the initial image. Although it is theoretically possible to do this for an arbitrary number of iterations, this is not possible in practice with a fixed discretization (and quantization) step size. Note that running the L^2 equations backward would exhibit clearly noticeable instabilities immediately (just one or two iterations with the same time step $\delta t = 0.1$). As such a procedure is not justified mathematically, we have omitted such a meaningless comparison with L^2 in this figure.

7.2. Well-posed sharpening. As discussed in section 5, we can apply the Sobolev diffusion equations in the reverse direction for image sharpening. We consider first the sharpening



(a) Original. (b) $\alpha = 2$. (c) $\alpha = 5$. (d) $\alpha = 10$.

Figure 10. *Stable Sobolev sharpening without stopping times. For each sharpness factor α the gradient descent PDE (5.2) converged to a local minimum in under 25 iterations using a time step $\delta t = 0.1$ (diffusion performed on the luminance component only).*

scheme (5.2) which is reproduced below:

$$u(0) = u_0, \qquad \frac{du}{dt} = \left(\frac{\int_\Omega \|\nabla u(t)\|^2}{\int_\Omega \|\nabla u_0\|^2} - \alpha \right) \Delta (I - \Delta)^{-1} u(t), \quad t > 0.$$

Thus we are evolving an image under reverse/forward Sobolev diffusion until the ratio of $\int_\Omega \|\nabla u\|^2$ to $\int_\Omega \|\nabla u_0\|^2$ is α. Here, α controls the level of sharpening/smoothing; for $\alpha < 1$ we get smoothing and for $\alpha > 1$ sharpening. The results of this well-posed sharpening algorithm are shown in Figure 10.

We now consider the well-posed sharpening scheme with an explicit data fidelity term (5.6), which is reproduced below:

$$u(0) = u_0, \qquad \frac{du}{dt} = -\Delta (I - \lambda \Delta)^{-1} u(t) + \mu (u_0 - u(t)), \quad t > 0.$$

The results of this well-posed sharpening algorithm are illustrated in Figure 5 without the instabilities apparent in the previous L^2 results. Note the lack of staircasing in both Figures 5 and 10. This is a direct result of the regularity properties of Theorems 4.2, 4.4, 4.5, and 4.6. We see that Sobolev sharpening tends to produce natural-looking images and does not create artifacts which were not present in the original image.

7.3. Comparison with shock filter. In this section we compare our sharpening technique with the popular L^2-based PDE technique for sharpening, known as the *shock filter* [21]. In our experiments, we found that the shock filter gave visually pleasing results when applied to original images that did not already appear blurred nor contain many smooth-looking regions. However, when we applied even slight levels of blurring to an otherwise clean-looking image (see top row of Figure 11) the shock filter would produce unnatural-looking "staircasing artifacts" in the smooth regions as the local numerical differencing schemes attempted to preserve and reinforce edges where there really were none to begin with (see bottom row of Figure 11). Such local artifacts were not created with the more global behavior of Sobolev-based sharpening. Hence, the resulting sharpened images look much more natural (see middle row of Figure 11). This same local-versus-global difference between the two sharpening filters makes the Sobolev sharpening less sensitive to changes in scale. In Figure 12 we compare the effects of running the shock filter on the baboon image and then subsampling the result versus first subsampling the baboon image and then running the shock filter on the result. It is not difficult to see, when examining the two results, that the order of these operations makes a difference. The difference image shows large positive (white) and negative (black) swings with a peak signal to noise ratio (PSNR) of 20.7 dB. The same experiment in Figure 12 with the more global Sobolev sharpening filter yields two results whose differences are much harder to perceive. The difference image shows much less drastic swings from zero (gray), with a PSNR of 29.7 dB, a difference of 9 dB. Note that we have applied exactly the same Sobolev diffusion PDE (i.e., with the same λ). We could get even better results if we scaled the Sobolev kernel (by decreasing λ) to match the image sampling period when processing the subsampled image.

7.4. Combined sharpening and smoothing. Finally, we show the results of a new combined sharpening and smoothing PDE which is derived from Sobolev diffusions. Figure 13 shows the results of (5.14), which is reproduced below:

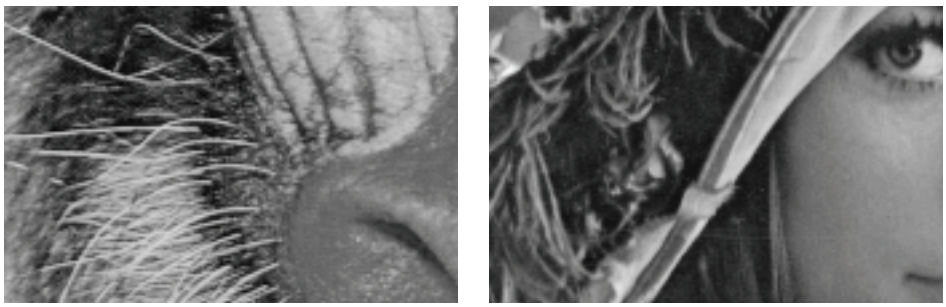
$$u(0) = u_0, \qquad \frac{du}{dt} = -\left(\nabla_{g_\lambda} E_1|_u - \frac{g_\lambda\left(\nabla_{g_\lambda} E_1|_u, \nabla_{g_\lambda} E_2|_u\right)}{g_\lambda\left(\nabla_{g_\lambda} E_2|_u, \nabla_{g_\lambda} E_2|_u\right)} \nabla_{g_\lambda} E_2|_u\right), \quad t > 0.$$

Recall that this PDE is derived from the following constrained minimization problem:

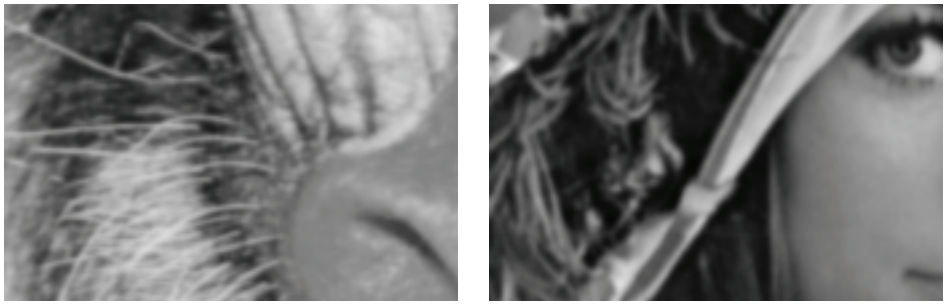
$$\min_{\substack{u \in H_0^1(\Omega): \\ \int_\Omega ||\nabla u||^2 = C}} \frac{1}{2} \int_\Omega (\Delta u)^2,$$

where $E_1[u] := 1/2 \int_\Omega (\Delta u)^2$ and $E_2[u] := \int_\Omega ||\nabla u||^2$. Since an unconstrained second order regularization will also regularize the first order image derivatives as a side effect, this first order constraint effectively forces any such local reduction in first order image derivatives to be compensated by local increases in first order derivatives elsewhere in the image. As shown in Figure 13, this has the effect of enhancing larger scale edges, while smaller scale noise and texture are smoothed away.

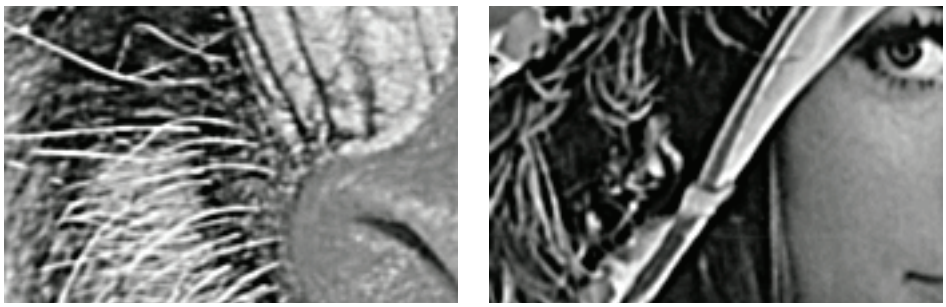
8. Conclusion and future work. In this paper we demonstrated, both theoretically and experimentally, how Sobolev gradients can be used for well-posed (and thereby more natural-looking) image sharpening. While our work was strongly motivated by some recent work



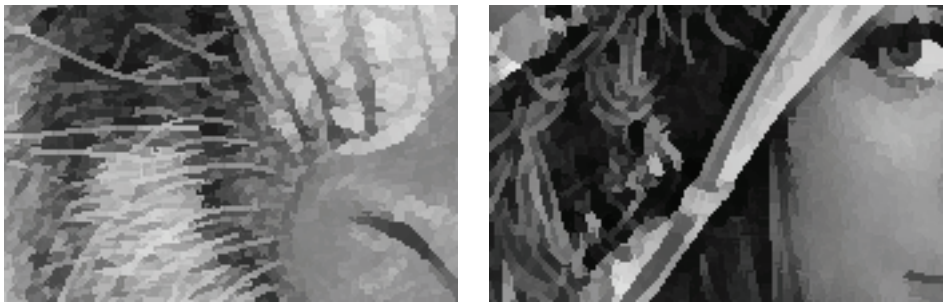
(a) Original (Baboon). (b) Original (Lena).



(c) Gaussian blurred original. (d) Gaussian blurred original.

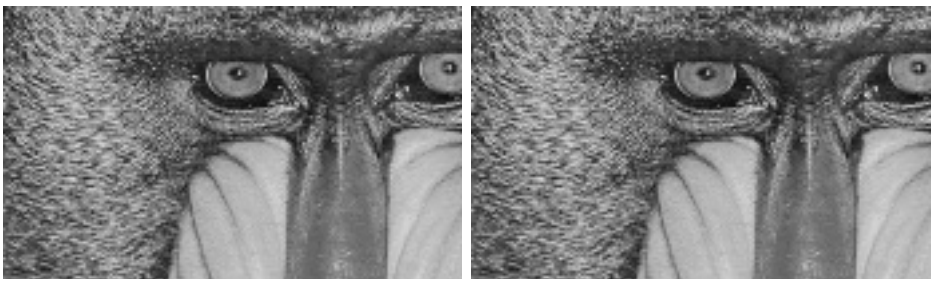


(e) Sobolev sharpening ($\alpha = 10$). (f) Sobolev sharpening ($\alpha = 5$).

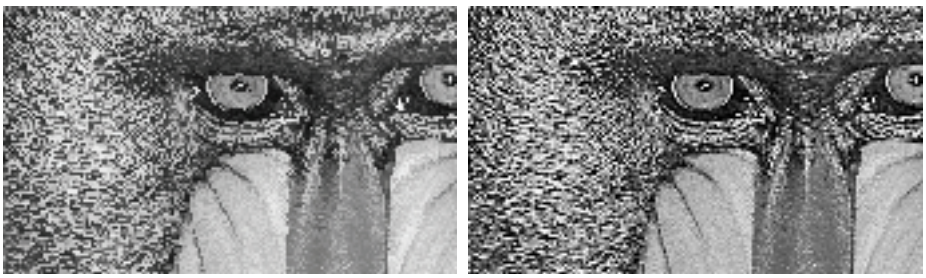


(g) Shock filter. (h) Shock filter.

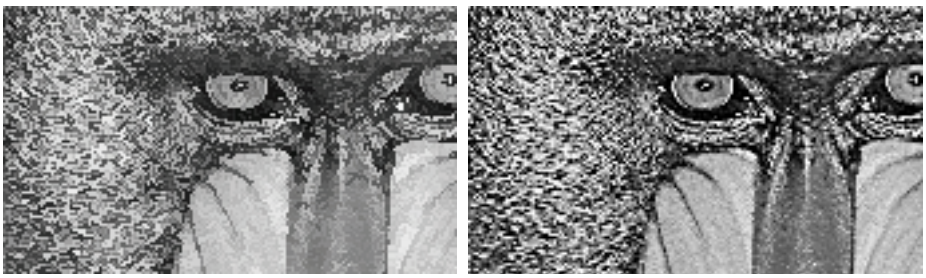
Figure 11. *Sobolev sharpening versus shock filter. The Sobolev sharpening (third row) was performed with the PDE (5.2) on Gaussian blurred versions (second row, 5 iterations of the heat equation with $\delta = 0.1$) of the Baboon and Lena images. Notice the lack of unnatural-looking "staircase artifacts" compared to the shock filter results (bottom row) for the same two images.*



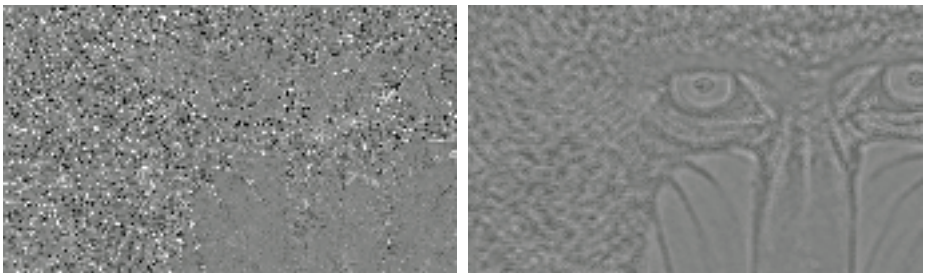
(a) Original Baboon image.



(b) Sharpened then subsampled via (left) shock filter or (right) Sobolev.



(c) Subsampled then sharpened via (left) shock filter or (right) Sobolev.



(d) Difference image: (left) Shock filter: PSNR 20.7 dB. (right) Sobolev: PSNR 29.7 dB.

Figure 12. *Scale sensitivity: Shock filter vs. Sobolev. The shock filter (left) and Sobolev sharpening (right) were each run on both the original and subsampled images. Sobolev sharpening was performed with the PDE* (5.2) *with a sharpness factor of* $\alpha = 5$. *The difference between sharpening then subsampling versus subsampling then sharpening for the shock filter is easily apparent when looking at the two results (e.g., the whites of the eyes). The PSNR of the Sobolev difference image is a full* 9 *dB higher than that of the shock filter.*

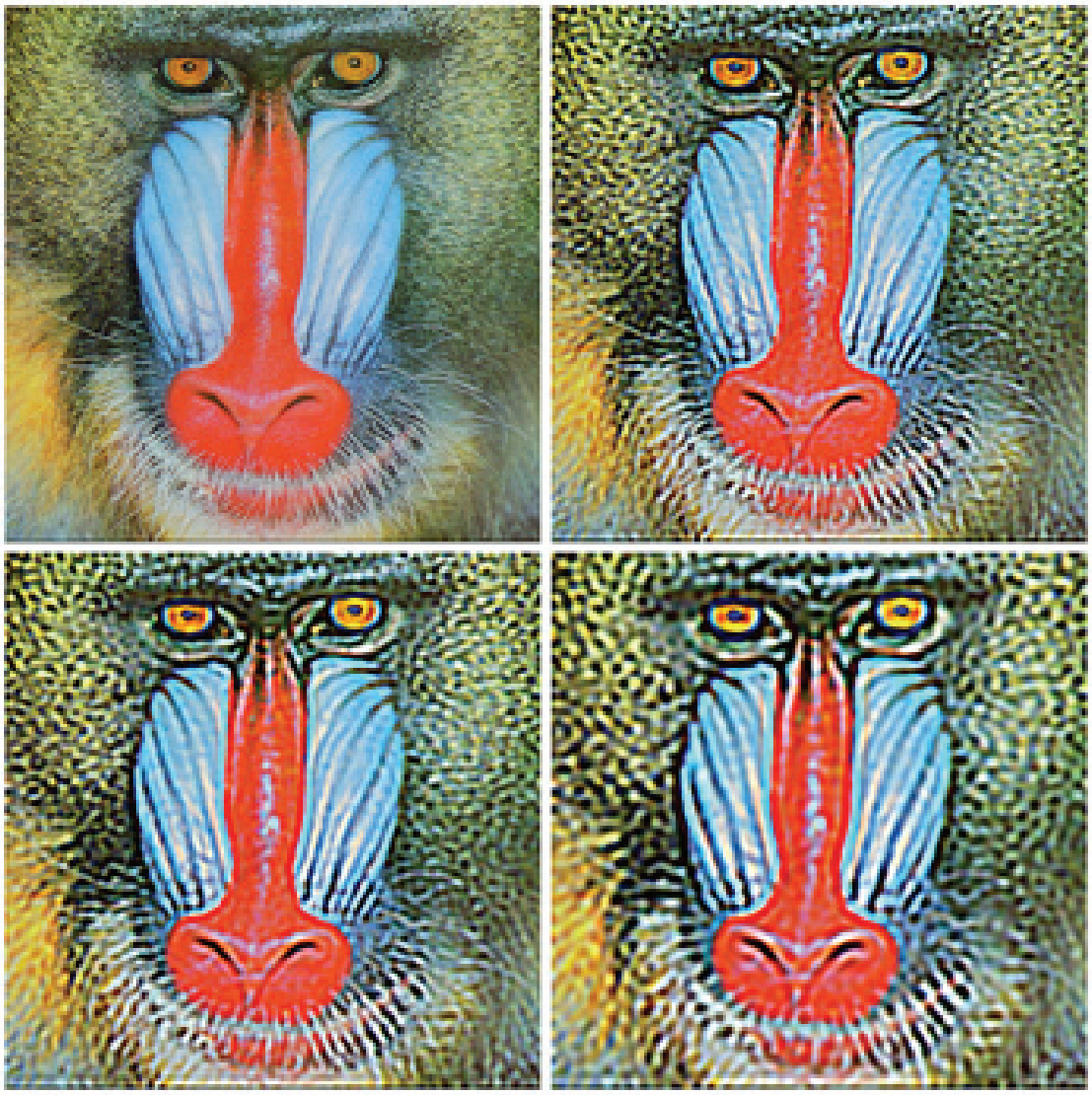


Figure 13. *Second order smoothing with sharpening constraints. Equation* (5.14) *evolved for* 0, 100, 200, 600 *iterations* ($\delta t = 0.1$) *from top left to bottom right. All three color components (Y, Cb, Cr) were processed.*

using geometric Sobolev gradients in active contour applications, we found that the benefits in image processing took on a very different twist. Sobolev active contours have been touted primarily for their robustness to noise in segmentation and tracking applications, whereas the well-posedness in the reverse direction is a key property for image processing.

We showed how this property can be exploited by demonstrating a stable Sobolev image sharpening filter which compares favorably to the well-known shock filter, yielding much more natural-looking images without exhibiting any staircase artifacts. We also demonstrated how the Sobolev sharpening PDE naturally leads to a class of interesting constrained flows on higher order energy functionals for simultaneous smoothing and sharpening effects. In this

paper we have only begun to explore the possible applications of reverse diffusion techniques; there are undoubtedly many other interesting applications.

We are currently investigating the use of H^k Sobolev metrics with $k > 1$ for the same functional and for the same image enhancement and sharpening applications that were described in this paper. We expect gradient flows on functionals with higher order derivatives to be well-posed in this context. In addition, we are also investigating the use of Sobolev metrics in image space for functionals which yield anisotropic diffusion equations under the L^2 metric. In parallel we are also investigating the use of Sobolev metrics in spaces of vector fields, with a particular eye on motion estimation applications. We expect the same smoothing properties and discontinuity-preserving behavior that were observed in the Sobolev diffusion equations to manifest themselves in this context. This could open the way for new possibilities in motion segmentation. We are also investigating the use of Sobolev metrics in other functional spaces which occur in variational problems in image processing.

9. Appendix.

9.1. Fundamental solution.

We now show how to compute the fundamental solution of $u - \lambda\Delta u = f$ in $\mathbb{R}^n$. The main computational tool that we will need for this task is the Fourier transform. For $u \in L^1(\mathbb{R}^n)$ we define its Fourier transform as

$$\hat{u}(\omega) := \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} e^{-ix\cdot\omega} u(x)dx \quad (\omega \in \mathbb{R}^n), \tag{9.1}$$

where $x \cdot \omega = \sum_{i=1}^n x_i\omega_i$, with $x = (x_1, \ldots, x_n)$ and $\omega = (\omega_1, \ldots, \omega_n)$. The inverse Fourier transform is defined as

$$\check{u}(x) := \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} e^{i\omega\cdot x} u(\omega)d\omega \quad (x \in \mathbb{R}^n). \tag{9.2}$$

It is well known that the Fourier transform can be linearly extended to an isometry from $L^2(\mathbb{R}^n)$ to $L^2(\mathbb{R}^n)$.

Now, computing $(I - \lambda\Delta)^{-1} f$ for an arbitrary $f \in L^2(\mathbb{R}^n)$ is equivalent to solving the following PDE:

$$u - \lambda\Delta u = f \quad \text{in} \quad \mathbb{R}^n \tag{9.3}$$

for the unknown function u. It is a well-known result [11] that, for every $\lambda > 0$, (9.3) has a unique weak solution $u \in H^2(\mathbb{R}^n)$. Now, take the Fourier transform of both sides of (9.3) to obtain

$$\hat{u}(\omega) = \frac{\hat{f}(\omega)}{1 + \lambda|\omega|^2} \quad (\omega \in \mathbb{R}^n).$$

Thus, by taking the inverse Fourier transform, we obtain $u = S_\lambda * f$, where S_λ is a rescaled Bessel potential [11]:

$$\hat{S}_\lambda(\omega) = \frac{1}{(2\pi)^{n/2}} \left(\frac{1}{1 + \lambda|\omega|^2} \right), \qquad S_\lambda(x) = \frac{1}{(4\lambda\pi)^{n/2}} \int_0^\infty \frac{e^{-t - \frac{|x|^2}{4t\lambda}}}{t^{n/2}} dt. \tag{9.4}$$

9.2. Mollification. We define the standard mollifier $\eta \in C^\infty(\mathbb{R}^n)$ by

$$\eta(x) := \begin{cases} C \exp\left(\dfrac{1}{|x|^2 - 1}\right) & \text{if } |x| < 1, \\ 0 & \text{if } |x| \geq 1, \end{cases}$$

where $C > 0$ is selected so that $\int_{\mathbb{R}^n} \eta dx = 1$. For each $\varepsilon > 0$ we set

$$\eta_\varepsilon(x) := \frac{1}{\varepsilon^n} \eta\left(\frac{x}{\varepsilon}\right) \quad \text{and} \quad \Omega_\varepsilon := \{x \in \Omega \mid \operatorname{dist}(x, \partial\Omega) > \varepsilon\}.$$

The functions η_ε are C^∞ and satisfy

$$\int_{\mathbb{R}^n} \eta_\varepsilon \, dx = 1 \quad \text{and} \quad \operatorname{supp}(\eta_\epsilon) \subset \{x \in \mathbb{R}^n \mid |x| \leq \varepsilon\}.$$

For any $u \in L^1_{loc}(\mathbb{R}^n)$, we define its mollification, u^ε, by $u^\varepsilon := \eta_\varepsilon * u$. It is a standard result that $u^\varepsilon \in C^\infty(\mathbb{R}^n)$ for all $\varepsilon > 0$ and $u^\varepsilon \to u_0$ a.e. as $\varepsilon \to 0$ [11].

Acknowledgments. The authors thank the anonymous referees for their detailed comments, which have improved this manuscript, especially the referee who suggested the alternative approach to the maximum principle that is sketched in Remark 2. The authors would also like to thank Mr. Matt Young for providing detailed comments on the theoretical aspects of the manuscript.

REFERENCES

[1] L. Alvarez, P.-L. Lions, and J.-M. Morel, *Image selective smoothing and edge detection by nonlinear diffusion.* II, SIAM J. Numer. Anal., 29 (1992), pp. 845–866.
[2] L. Alvarez and L. Mazorra, *Signal and image restoration using shock filters and anisotropic diffusion*, SIAM J. Numer. Anal., 31 (1994), pp. 590–605.
[3] G. Aubert and P. Kornprobst, *Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations*, Appl. Math. Sci., Springer, New York, 2006.
[4] J. Bai and X.-C. Feng, *Fractional-order anisotropic diffusion for image denoising*, IEEE Trans. Image Process., 16 (2007), pp. 2492–2502.
[5] H. Brezis, *Operateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*, North-Holland Math. Stud. 5, North–Holland, Amsterdam, 1975.
[6] H. Brezis, *Analyse fonctionnelle: Théorie et applications*, Masson, Paris, 1993.
[7] F. Catté, P.-L. Lions, J.-M. Morel, and T. Coll, *Image selective smoothing and edge detection by nonlinear diffusion*, SIAM J. Numer. Anal., 29 (1992), pp. 182–193.
[8] A. Chambolle, *Partial differential equations and image processing*, in Proceedings of the IEEE International Conference on Image Processing, Austin, TX, 1994, IEEE Press, Piscataway, NJ, pp. 16–20.
[9] G. Charpiat, P. Maurel, J.-P. Pons, R. Keriven, and O. Faugeras, *Generalized gradients: Priors on minimization flows*, Int. J. Computer Vision, 73 (2007), pp. 325–344.
[10] G. Dal Maso, *An Introduction to Γ-Convergence*, Birkhäuser Boston, Cambridge, MA, 1993.
[11] L. C. Evans, *Partial Differential Equations*, Grad. Stud. Math. 19, AMS, New York, 1998.
[12] I. Galić, J. Weickert, M. Welk, A. Bruhn, A. Belyaev, and H.-P. Seidel, *Image compression with anisotropic diffusion*, J. Math. Imaging Vision, 31 (2008), pp. 255–269.
[13] D. Gilbarg and N. S. Trudinger, *Elliptic Partial Differential Equations of Second Order*, Springer, New York, 2000.
[14] G. Gilboa, N. Sochen, and Y. Y. Zeevi, *Forward-and-backward diffusion processes for adaptive image enhancement and denoising*, IEEE Trans. Image Process., 11 (2002), pp. 689–703.

[15] P. Guidotti, *A new nonlocal nonlinear diffusion of image processing*, J. Differential Equations, 246 (2009), pp. 4731–4742.
[16] P. Guidotti and J. Lambers, *Two new nonlinear nonlocal diffusions for noise reduction*, J. Math. Imaging Vision, 33 (2009), pp. 25–37.
[17] L. H. Lieu and L. A. Vese, *Image restoration and decomposition via bounded total variation and negative Hilbert-Sobolev spaces*, Appl. Math. Optim., 58 (2008), pp. 167–193.
[18] M. Lysaker, A. Lundervold, and X.-C. Tai, *Noise removal using fourth-order partial differential equations with applications to medical magnetic resonance images in space and time*, IEEE Trans. Image Process., 12 (2003), pp. 1579–1590.
[19] J. Martin-Herrero, *Anisotropic diffusion in the hypercube*, IEEE Trans. Geosci. Remote Sensing, 45 (2007), pp. 1386–1398.
[20] J. W. Neuberger, *Sobolev Gradients and Differential Equations*, 2nd ed., Lecture Notes in Math., Springer, Berlin, Heidelberg, 2010.
[21] S. Osher and L. I. Rudin, *Feature-oriented image enhancement using shock filters*, SIAM J. Numer. Anal., 27 (1990), pp. 919–940.
[22] P. Perona and J. Malik, *Scale-space and edge detection using anisotropic diffusion*, IEEE Trans. Pattern Anal. Machine Intell., 12 (1990), pp. 629–639.
[23] W. B. Richardson, Jr., *High-order Sobolev preconditioning*, Nonlinear Anal., 63 (2005), pp. e1779–e1787.
[24] W. B. Richardson, Jr., *Sobolev gradient preconditioning for image-processing PDEs*, Comm. Numer. Methods Engrg., 24 (2008), pp. 493–504.
[25] L. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Phys. D, 60 (1992), pp. 259–268.
[26] G. Sapiro and D. L. Ringach, *Anisotropic diffusion of multivalued images with applications to color filtering*, IEEE Trans. Image Process., 5 (1996), pp. 1582–1586.
[27] G. Sundaramoorthi, J. D. Jackson, and A. Yezzi, *Tracking with Sobolev active contours*, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, New York, 2006, IEEE Computer Society Press, Piscataway, NJ, pp. 674–680.
[28] G. Sundaramoorthi, A. Yezzi, and A. C. Mennucci, *Sobolev active contours*, Int. J. Computer Vision, 73 (2005), pp. 109–120.
[29] G. Sundaramoorthi, A. Yezzi, A. C. Mennucci, and G. Sapiro, *New possibilities with Sobolev active contours*, in Proceedings of the Conference on Scale Space Variational Methods, Springer-Verlag, 2007, pp. 153–164.
[30] D. Tschumperlé, *Fast anisotropic smoothing of multi-valued images using curvature-preserving PDE's*, Int. J. Computer Vision, 68 (2006), pp. 65–82.
[31] D. Tschumperlé and R. Deriche, *Vector-valued image regularization with PDEs: A common framework for different applications*, IEEE Trans. Pattern Anal. Machine Intell., 27 (2005), pp. 506–517.
[32] J. Weickert, *Anisotropic Diffusion in Image Processing*, ECMI Series, Teubner-Verlag, Stuttgart, Germany, 1998.
[33] J. Weickert, *Coherence-enhancing diffusion filtering*, Int. J. Computer Vision, 31 (1999), pp. 111–127.
[34] M. Welk, G. Gilboa, and J. Weickert, *Theoretical foundations for discrete forward-and-backward diffusion filtering*, in Proceedings of the Second International Conference on Scale Space and Variational Methods in Computer Vision, 2009, Springer-Verlag, Berlin, pp. 527–538.
[35] A. Witkin, *Scale-space filtering*, in Proceedings of the International Joint Conference on Artificial Intelligence, Karlsruhe, West Germany, 1983, pp. 1019–1021.
[36] A. Witkin, *Scale-space filtering: A new approach to multi-scale description*, in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, IEEE Computer Society Press, Piscataway, NJ, 1984, vol. 9, pp. 153–159.
[37] A. Yezzi, *Modified curvature motion for image smoothing and enhancement*, IEEE Trans. Image Process., 7 (1998), pp. 345–352.
[38] Y. You and M. Kaveh, *Fourth-order partial differential equations for noise removal*, IEEE Trans. Image Process., 9 (2000), pp. 1723–1730.
[39] F. Zhang and E. R. Hancock, *Graph spectral image smoothing using the heat kernel*, Pattern Recognition, 41 (2008), pp. 3328–3342.