A Simple Explanation of the Sobolev Gradient Method

R. J. Renka

July 3, 2006

Abstract

We have observed that the term 'Sobolev gradient' is used more often than it is understood. Also, the term is often used with no apparent awareness of its origin. This short note is an attempt to correct both of these situations.

1 Introduction

John W. Neuberger coined the term 'Sobolev gradient' and began developing the method in the 1970's, but his first journal article on the subject was not published until 1985 ([4]). The primary reference for his work is now [5]. The Sobolev gradient method is a gradient descent method designed to find critical points of a (discretization of a) functional of the form $\phi(u) = \int_\Omega F(Du)$, where D is a differential operator and u is an element of a Sobolev space $\mathbf{H}$ for which $F(Du)$ is integrable. The functional represents a variational formulation of a differential equation, and any differential equation $G(Du) = 0$ can be treated as a least squares variational problem by taking $F(Du) = G(Du)^2/2$ if no better alternative is available. While the primary application is numerical solution of nonlinear partial differential equations, the method has been extended to geometric modeling problems ([6], [7]), inverse problems ([3]), optimal control problems ([1]), and image processing problems ([8]).

Section 2 presents a brief outline of the key ideas in the context of Sobolev spaces, and Section 3 treats the discretized problem as one of numerical optimization. As a numerical optimization method, the key idea is preconditioned gradient descent, and the efficacy of the method is clear without any consideration of function spaces. Unlike other methods for preconditioning, however, the Sobolev gradient method arises in a very natural and elegant manner from the Sobolev space setting. Refer to [2] for an extensive discussion of preconditioning.

2 Function Space Setting

In order to make the ideas concrete, we apply the method to a simple example problem — Poisson's equation: $u_{xx} + u_{yy} = f$ on a two-dimensional

domain Ω with Dirichlet boundary conditions. Taking $u \in \mathbf{H} \equiv \mathbf{H}^{1,2}(\Omega)$ and $Du = (u, u_x, u_y) \in \mathbf{L}_2(\Omega)^3$, we have $F(Du) = (u_x^2 + u_y^2)/2 + fu$ integrable for $f \in \mathbf{L}_2(\Omega)$. Then Poisson's equation is the Euler-Lagrange equation associated with $\phi(u) = \int_\Omega F(Du)$. We show below that it is obtained by setting the L_2 gradient of ϕ to zero. Note, however, that this involves an implicit assumption that ϕ has a gradient in $\mathbf{L}_2(\Omega)$ or, equivalently, that $u \in \mathbf{H}^{2,2}(\Omega)$. Since the variational formulation imposes a weaker restriction on u, it is referred to as a weak formulation. In order to treat the boundary conditions, we use an initial estimate that satisfies the conditions and restrict perturbations of u to the subspace $\mathbf{H}_0$ of functions that satisfy homogeneous boundary conditions.

The Fréchet derivative $\phi'(u)$, when it exists, is the bounded linear functional defined by

$$\phi'(u)h = \lim_{\alpha \to 0} \frac{1}{\alpha}[\phi(u + \alpha h) - \phi(u)]$$

for $h \in \mathbf{H}_0$. For the example problem, we have

$$\begin{aligned}
\phi'(u)h &= \lim_{\alpha \to 0} \frac{1}{2\alpha} \int_\Omega [(u_x + \alpha h_x)^2 + (u_y + \alpha h_y)^2 + 2f(u + \alpha h) - \\
&\qquad\qquad (u_x^2 + u_y^2 + 2fu)] \\
&= \lim_{\alpha \to 0} \frac{1}{2\alpha} \int_\Omega [2\alpha(u_x h_x + u_y h_y + fh) + O(\alpha^2)] \\
&= \int_\Omega [u_x h_x + u_y h_y + fh] \qquad \forall h \in \mathbf{H}_0. \qquad (1)
\end{aligned}$$

The Sobolev gradient $\nabla_S \phi(u)$ is defined to be the unique element of $\mathbf{H}_0$ that represents the bounded linear functional:

$$\phi'(u)h = \langle \nabla_S \phi(u), h \rangle_{\mathbf{H}} \qquad \forall h \in \mathbf{H}_0.$$

Note that boundedness of $\phi'(u)$ is related to existence of $\nabla_S \phi(u) \in \mathbf{H}_0$ by

$$|\phi'(u)h| \le \|\nabla_S \phi(u)\|_{\mathbf{H}} \|h\|_{\mathbf{H}} \qquad \forall h \in \mathbf{H}_0.$$

To show that our example functional is C^1 (Fréchet differentiable for all u), and the Sobolev gradient exists, denote by P the (self-adjoint) orthogonal projection of $\mathbf{L}_2(\Omega)^3$ onto $\left\{ Dk = \begin{pmatrix} k \\ k_x \\ k_y \end{pmatrix} : k \in \mathbf{H}_0 \right\}$, and define $\pi \begin{pmatrix} k_1 \\ k_2 \\ k_3 \end{pmatrix} = k_1$ for $k_1, k_2, k_3 \in \mathbf{L}_2(\Omega)$. Then, from (1),

$$\begin{aligned}
\phi'(u)h &= \left\langle \begin{pmatrix} h \\ h_x \\ h_y \end{pmatrix}, \begin{pmatrix} f \\ u_x \\ u_y \end{pmatrix} \right\rangle_{\mathbf{L}_2^3} = \left\langle PDh, \begin{pmatrix} f \\ u_x \\ u_y \end{pmatrix} \right\rangle_{\mathbf{L}_2^3} \\
&= \left\langle Dh, P \begin{pmatrix} f \\ u_x \\ u_y \end{pmatrix} \right\rangle_{\mathbf{L}_2^3} = \left\langle h, \pi P \begin{pmatrix} f \\ u_x \\ u_y \end{pmatrix} \right\rangle_{\mathbf{H}} \qquad \forall h \in \mathbf{H}_0.
\end{aligned}$$

Hence

$$\nabla_S \phi(u) = \pi P \begin{pmatrix} f \\ u_x \\ u_y \end{pmatrix} \in \mathbf{H}_0.$$

(Note that D was chosen so that $\langle u, v \rangle_{\mathbf{H}} \equiv \langle Du, Dv \rangle_{\mathbf{L}_2^3}$.) A more computationally amenable expression for $\nabla_S \phi(u)$ will be derived in the next section. Our purpose here is to show that the Sobolev gradient exists when the $\mathbf{L}_2$ gradient may not. In order for $\phi'(u)$ to be bounded in the $\mathbf{L}_2$ norm, u must be in $\mathbf{H}^{2,2}(\Omega)$ so that, integrating by parts (or applying Green's theorem) in (1), we obtain

$$\begin{aligned} \phi'(u)h &= -\int_\Omega (u_{xx} + u_{yy} - f)h \\ &= \langle -(u_{xx} + u_{yy} + f), h \rangle_{\mathbf{L}_2} \\ &= \langle \nabla\phi(u), h \rangle_{\mathbf{L}_2} \quad \forall h \in \mathbf{H}_0, \end{aligned}$$

giving the $\mathbf{L}_2$ gradient $\nabla\phi(u) = -(u_{xx} + u_{yy}) + f$.

3 Numerical Setting

Consider a gradient descent method for minimizing a functional ϕ defined on a Hilbert space $\mathbf{H} = \mathbf{H}^{1,2}(\Omega)$. The method of steepest descent is defined by the iteration

$$u_{k+1} = u_k - \alpha_k \nabla\phi(u_k) \tag{2}$$

for $u_0 \in \mathbf{H}$ and some means of choosing step-lengths α_k. In order that the iterates remain in $\mathbf{H}$, the gradients $\nabla\phi(u_k)$ must be elements of $\mathbf{H}$; i.e., derivatives must be compatible with the metric (inner product and corresponding norm) on $\mathbf{H}$. More precisely, the directional derivative of ϕ at u in the direction h defines the Sobolev gradient $\nabla_S \phi(u)$ by $\phi'(u)h = \langle \nabla_S \phi(u), h \rangle_{\mathbf{H}} \ \ \forall h \in \mathbf{H}$. Stated differently, the negative gradient of ϕ at u is the direction of steepest descent because it maximizes the decrease in ϕ per unit change in u. The norm used to measure units in the domain of ϕ strongly influences the gradient direction.

Now consider a finite difference discretization of the problem. Although variational problems are usually discretized by a finite element method, finite differencing is simpler and, in some cases, much more efficient ([7]) than finite elements. We retain the above notation but now have $u_k \in \mathbf{R}^N$ for N grid points in Ω (or N/m grid points if u has m components associated with each grid point). Let $\mathbf{H}_0$ be the subspace of vectors that satisfy homogeneous boundary conditions, so that by restricting gradients to $\mathbf{H}_0$, the iterates u_k satisfy the same boundary conditions satisfied by u_0.

Simply replacing the Sobolev gradient by the ordinary gradient (vector of partial derivative values) does not lead to an effective method. The ordinary gradient is associated with the Euclidean inner product which (up to a scale factor on a regular grid) is the discretization of the $\mathbf{L}_2$ inner product. Hence, the ordinary gradient is essentially a discretization of the $\mathbf{L}_2$ gradient, and it lacks

the required smoothness of a discretized Sobolev gradient. The undiscretized functional ϕ involves derivatives of u and is therefore not well-defined on $\mathbf{L}_2(\Omega)$. As the mesh width decreases, the discretization approaches an ill-defined problem. In short, standard descent methods are completely lacking in integrity for minimizing functionals that involve derivatives.

In order to properly discretize the Sobolev gradient, we must discretize the Sobolev inner product. To this end, denote by D the M by N matrix representing finite difference approximations to the appropriate derivatives, along with values obtained by averaging if necessary. Assume that D has linearly independent columns so that D^TD is positive definite as well as symmetric. In practice it may be necessary to restrict D^TD to $\mathbf{H}_0$ and to follow its application by projection onto $\mathbf{H}_0$. We now have a discrete Sobolev inner product:

$$\langle g, h\rangle_{\mathbf{H}} = (Dg)^T(Dh) = g^TD^TDh.$$

We can use the two inner products to relate the two gradients:

$$\phi'(u)h = \nabla\phi(u)^Th = h^T\nabla\phi(u)$$

and

$$\phi'(u)h = \langle\nabla_S\phi(u), h\rangle_{\mathbf{H}} = \nabla_S\phi(u)^TD^TDh = h^TD^TD\nabla_S\phi(u)$$

for all $h \in \mathbf{H}_0$, implying that

$$\nabla_S\phi(u) = (D^TD)^{-1}\nabla\phi(u). \tag{3}$$

The Sobolev gradient is thus computed by solving the linear system with the sparse symmetric positive definite matrix D^TD and the ordinary gradient as right hand side. Note that D^TD is a discretized second derivative operator (typically of the form $I - \Delta$ for identity I and Laplacian Δ), and its inverse is therefore a smoothing operator, where smoothness is measured by the number of continuous derivatives. It is shown in [5] that for a very general pair of compatible inner products, the corresponding gradients are related by an operator that has the properties of a Laplacian.

We now discuss an alternative perspective — the discretized Laplacian as a preconditioner. Assuming ϕ is twice continuously differentiable in a neighborhood of u, we have the Taylor series

$$\phi(u + \alpha h) = \phi(u) + \alpha h^T\nabla\phi(u) + \frac{\alpha^2}{2}h^TG(u)h + O(\alpha^3), \tag{4}$$

where $G(u)$ denotes the Hessian of ϕ at u. Let L be a symmetric positive definite order-N matrix so that $\langle g, h\rangle_L = g^TLh$ defines an inner product. Then there exists a corresponding gradient $\nabla_L\phi(u)$ and Hessian $G_L(u)$ such that

$$\begin{aligned}\phi(u + \alpha h) &= \phi(u) + \alpha\langle h, \nabla_L\phi(u)\rangle_L + \frac{\alpha^2}{2}\langle h, G_L(u)h\rangle_L + O(\alpha^3)\\ &= \phi(u) + \alpha h^TL\nabla_L\phi(u) + \frac{\alpha^2}{2}h^TLG_L(u)h + O(\alpha^3)\end{aligned} \tag{5}$$

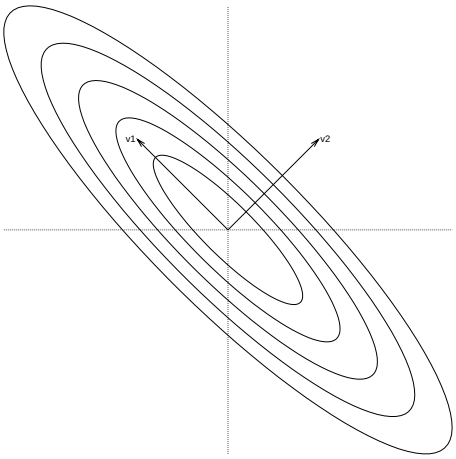

Figure 1: Ellipses $\{h \in \mathbf{R}^2 : \phi(u^*) + \frac{1}{2}h^T G(u^*)h = k\}$

for all $\alpha \in \mathbf{R}, h \in \mathbf{R}^N$. Hence, equating terms in (4) and (5),

$$\nabla_L \phi(u) = L^{-1}\nabla\phi(u) \quad \text{and} \quad G_L(u) = L^{-1}G(u).$$

Let $u = u^*$ be a critical point at which $\nabla\phi = 0$ so that

$$\phi(u^* + \alpha h) = \phi(u^*) + \frac{\alpha^2}{2}h^T G(u^*)h + O(\alpha^3).$$

For small α, the level sets defined by $\phi(u^* + \alpha h) = k$, k a constant, are approximated by the conic sections

$$\{u^* + h : \phi(u^*) + \frac{1}{2}h^T G(u^*)h = k\}.$$

The behavior of ϕ in the vicinity of u^* is governed by the eigenvalues of $G(u^*)$. If they are all positive, the conic sections are hyperellipsoids, and u^* is a local minimum. This is depicted for the case $N = 2$ in Figure 1. The curves have been translated by $-u^*$, and the (orthonormal) eigenvectors v_1 and v_2 associated with $\lambda_1 = 1$ and $\lambda_2 = 4$ are displayed.

As the condition number of the Hessian $\lambda_{\max}/\lambda_{\min}$ increases, the location of u^* becomes more sensitive to perturbations of the data, and more difficult to compute accurately. As the condition number approaches 1, on the other hand, the level sets become circular (or approach hyperspherical in general), and the negative gradient direction at every point is toward the minimizer u^*. Note, however, that unless ϕ is quadratic, so that $G(u)$ is constant, the level sets deviate from ellipsoidal with distance from the critical point. The purpose of the preconditioner L is to reduce the condition number of the Hessian. Since

each descent step requires solution of the linear system $L\nabla_L\phi(u_k) = \nabla\phi(u_k)$, it is advantageous to choose preconditioners for which the systems are easily solved. We therefore have a tradeoff between cheaply solved systems but with many descent steps due to ineffective preconditioning, and more expensive but fewer descent steps. The two extremes are $L = I$, I denoting the identity (no preconditioning), and $L = G(u_k)$ at step k, in which case $G_L = I$ is perfectly conditioned, and the descent iteration is

$$\begin{aligned} u_{k+1} &= u_k - \alpha_k \nabla_L \phi(u_k) \\ &= u_k - \alpha_k G(u_k)^{-1} \nabla\phi(u_k). \end{aligned}$$

With the optimal step-length $\alpha_k = 1$, this is a Newton iteration. Newton's method is clearly the optimal choice when it is viable. However, $G(u_k)$ may not be positive definite except in a small neighborhood of a local minimum. Also, for large-scale problems such as the discretized partial differential equations of interest here, the Hessian is often unavailable or is too expensive to compute or to store. The preconditioner used by the Sobolev gradient method $L = D^T D$ on the other hand can always be made positive definite and, with the exception of extremely complex or high-order differential operators, is generally well-conditioned, making the linear systems inexpensive to solve. Due to its regular structure, $D^T D$ need not be stored at all in most cases. For many problems the best strategy is to start with a Sobolev gradient method and then, if high accuracy is required, switch to a Newton iteration when sufficiently close to a local minimum.

References

[1] A. Jameson and S. Kim, *Reduction of the adjoint gradient formula in the continuous limit*, AIAA paper, AIAA Aerospace Sciences meeting, Reno, NV, January 2003.

[2] Faragó, I., Karátson, J., *Numerical solution of nonlinear elliptic problems via preconditioning operators: Theory and applications.* Advances in Computation, Volume 11, NOVA Science Publishers, New York, 2002.

[3] I. Knowles, *Variational methods for ill-posed problems*, Contemporary Mathematics, 357 (2004), 187–199.
http://www.math.uab.edu/knowles/pubs.html.

[4] J. W. Neuberger, *Steepest descent and differential equations*, J. Math. Soc. Japan 37 (1985), 187–195.

[5] J. W. Neuberger, *Sobolev Gradients and Differential Equations*, Springer Lecture Notes in Mathematics #1670, 1997.

[6] R. J. Renka and J. W. Neuberger, *Minimal surfaces and Sobolev gradients*, SIAM. J. Sci. Comput. 16 (1995), 1412–1427.

[7] R. J. Renka, *Constructing fair curves and surfaces with a Sobolev gradient method*, CAGD 21 (2004), 137–149.

[8] G. Sundaramoorthi, A. Yezzi, and A. Mennucci, *Sobolev Active Contours*, IEEE Workshop on Variational and Level Set Methods, 2005.