Variational and Topological Methods: Theory, Applications, Numerical Simulations, and Open Problems (2012). *Electronic Journal of Differential Equations*, Conference 21 (2014), pp. 173–181. ISSN: 1072-6691. http://ejde.math.txstate.edu, http://ejde.math.unt.edu ftp ejde.math.txstate.edu

TIKHONOV REGULARIZATION USING SOBOLEV METRICS

PARIMAH KAZEMI, ROBERT J. RENKA

Abstract. Given an ill-posed linear operator equation $Au = f$ in a Hilbert space, we formulate a variational problem using Tikhonov regularization with a Sobolev norm of u, and we treat the variational problem by a Sobolev gradient flow. We show that the gradient system has a unique global solution for which the asymptotic limit exists with convergence in the strong sense using the Sobolev norm, and that the variational problem therefore has a unique global solution. We present results of numerical experiments that demonstrates the benefits of using a Sobolev norm for the regularizing term.

1. Introduction

Consider the following operator equation in which A is a linear mapping from a Hilbert space L to another Hilbert space K:

$$Au = f. \tag{1.1}$$

The equation is said to be well-posed if it satisfies the properties of existence, uniqueness, and stability; i.e., a solution u exists, it is unique, and it depends continuously on the data f (A has bounded inverse). Failure to satisfy any of the three properties characterizes the problem as ill-posed. These terms originated in the context of differential equations [5], but have been applied to problems, both linear and nonlinear, in almost every area of mathematics [7]. Ill-posed operator equations with A unbounded are relatively rare but have received some attention recently in [11] and [6]. Our focus will be on problems in which A is injective and compact.

To treat (1.1) when it is ill-posed, we can seek a minimizer of the least squares functional

$$\phi(u) = \frac{1}{2}\|Au - f\|_K^2.$$

However, if (1.1) has a nonunique solution so that A has a nontrivial kernel, then the minimization problem will not have a unique solution. We may choose the minimum-norm solution in this case. If A is invertible but A^{-1} is not continuous, then any noise present in f can lead to an arbitrarily large change in u. In this case we must regularize the problem in order to obtain a stable solution. One method

2000 *Mathematics Subject Classification.* 47A52, 65D25, 65F22.
Key words and phrases. Gradient system; Ill-posed problem; least squares; Sobolev gradient; Tikhonov regularization.

Published February 10, 2014.

is Tikhonov regularization in which we balance the size of the residual $\phi(u)$ against the norm of the solution u by minimizing

$$\phi_\alpha(u) = \frac{1}{2}\|Au - f\|^2 + \frac{\alpha}{2}\|u\|^2 \tag{1.2}$$

for positive regularization parameter α [13, 14, 15]. We thus have a convex minimization problem. Using the L^2 norm in (1.2), the solution is $u = (\alpha I + A^*A)^{-1}A^*f$ when A is bounded and f is in the domain of A^*.

Tikhonov regularization with A compact and injective has been extensively studied and is well understood, but the regularization term is almost always formulated in terms of the L^2 norm or the Euclidean norm in the finite dimensional case. The primary focus of this work is to demonstrate the advantage of using a discretized Sobolev norm for u. This has been referred to as higher-order Tikhonov regularization in [2]. Our contribution is an experiment that demonstrates the effectiveness of the Sobolev norm approach for numerical differentiation, and an analysis that includes convergence of the gradient flow and expressions for the gradient and solution of the minimization problem. The regularity properties of the solution evince the appropriateness of regularization with a Sobolev norm in place of the L^2 norm.

In Sections 2, 3, and 4 we develop the analysis for the problem of minimizing ϕ_α using a Sobolev gradient flow. We show that the gradient system has a global solution, the solution is unique, it has an asymptotic limit, and convergence to the limit is in the strong sense using the Sobolev norm. We derive an expression for the regularizing operator and for the solution u_α of the minimization problem. In Section 5 we present results of a numerical test that demonstrates the superiority of the Sobolev-norm based regularization strategy when u is expected to be smooth.

2. Tikhonov regularization in Hilbert space

Suppose that L and K are L^2 spaces, H is a Sobolev space compactly and densely embedded in L, and $A : L \to K$ is a bounded linear operator. We wish to minimize the functional $\phi_\alpha : H \to \mathbb{R}$ defined by

$$\phi_\alpha(u) = \frac{1}{2}\|Au - f\|_K^2 + \frac{\alpha}{2}\|u\|_H^2,$$

where $\alpha > 0$ and we have used the stronger H norm in place of the more commonly used L norm. The first and second Fréchet derivatives are

$$\phi'_\alpha(u)h = \langle Ah, Au - f\rangle_K + \alpha\langle h, u\rangle_H \tag{2.1}$$

and

$$\phi''_\alpha(u)(h, h) = \|Ah\|_K^2 + \alpha\|h\|_H^2$$

for $h \in H$. Note that ϕ_α is an everywhere defined C^2 function defined on the subspace H, and satisfies the following definition of convexity.

Definition 2.1. Suppose F is a C^2 function defined on a Hilbert space H. F is said to be convex if there exists a positive number ϵ so that for all $h \in H$,

$$F''(u)(h, h) \geq \epsilon\|h\|_H^2.$$

3. Minimization using Sobolev gradients

To obtain a minimum of ϕ_α, we compute a gradient with respect to the metric

$$\langle v, w\rangle_{\alpha,H,A} = \langle v, w\rangle_H + \frac{1}{\alpha}\langle Av, Aw\rangle_K.$$

The following definition of a Sobolev gradient is taken from [10].

Definition 3.1. Suppose that F is a real or complex valued Fréchet differentiable function that is everywhere defined on a Hilbert space H. Then for each $u \in H$, there exists a unique element $\nabla_H F(u) \in H$ such that

$$F'(u)h = \langle h, \nabla_H F(u)\rangle_H$$

for all $h \in H$. The gradient system associated with F is

$$z(0) = u_0 \text{ and } z'(t) = -\nabla_H F(z(t)) \tag{3.1}$$

for $u_0 \in H$.

Now suppose that H is a Sobolev space of order $k \geq 1$. Then

$$\langle v, w\rangle_H = \left\langle \begin{pmatrix} v \\ Dv \end{pmatrix}, \begin{pmatrix} w \\ Dw \end{pmatrix} \right\rangle_{\bar{L}}$$

where D is a differentiable operator involving partial derivatives of order 1 to k and $\bar{L} = L \times L^{n_k}$ with n_k denoting the number of multi-indices whose order is between 1 and k.

We recall the definition of the Sobolev space $H^{k,p}(\Omega)$ as given in [1].

Definition 3.2. Suppose Ω is an open subset of $\mathbb{R}^n$. $H^{k,p}(\Omega)$ is the completion of $C^k(\Omega)$ with respect to the norm $\|u\|_{k,p} = \left(\sum_{|\alpha|\leq k} \|D^\alpha u\|^p_{L^p(\Omega)}\right)^{(1/p)}$, where α is a multi-index of order less than or equal to k and D^α is a partial derivative operator.

It follows from this definition that D is a closed and densely defined operator on L. By the assumption that A is bounded on L and hence on H, $\|\cdot\|_{\alpha,H,A}$ induces an equivalent norm on H for each α. Hence the operator

$$T_\alpha = \begin{pmatrix} D \\ \frac{1}{\sqrt{\alpha}}A \end{pmatrix}$$

is also a closed densely defined operator on L. Thus there exists an orthogonal projection P_α onto the graph of $T_\alpha \subset \bar{L} \times K = L \times L^{n_k} \times K$.

From (2.1) we have

$$\phi_\alpha'(u)h = \alpha\left\langle \begin{pmatrix} h \\ T_\alpha h \end{pmatrix}, \begin{pmatrix} u \\ T_\alpha u \end{pmatrix} \right\rangle_{\bar{L}\times K} - \left\langle \begin{pmatrix} h \\ T_\alpha h \end{pmatrix}, \begin{pmatrix} 0 \\ \binom{0}{\sqrt{\alpha} f} \end{pmatrix} \right\rangle_{\bar{L}\times K}.$$

Hence

$$\begin{aligned}
\phi_\alpha'(u)h &= \alpha\langle h, u\rangle_{\alpha,H,A} - \left\langle P_\alpha \begin{pmatrix} h \\ T_\alpha h \end{pmatrix}, \begin{pmatrix} 0 \\ \binom{0}{\sqrt{\alpha} f} \end{pmatrix} \right\rangle_{\bar{L}\times K} \\
&= \alpha\langle h, u\rangle_{\alpha,H,A} - \left\langle \begin{pmatrix} h \\ T_\alpha h \end{pmatrix}, P_\alpha \begin{pmatrix} 0 \\ \binom{0}{\sqrt{\alpha} f} \end{pmatrix} \right\rangle_{\bar{L}\times K}
\end{aligned}$$

$$= \alpha\langle h, u\rangle_{\alpha,H,A} - \left\langle h, \Pi P_\alpha\begin{pmatrix} 0 \\ \binom{0}{\sqrt{\alpha}f}\end{pmatrix}\right\rangle_{\alpha,H,A},$$

where $\Pi(x, y) = x$. Thus the gradient is

$$\nabla_{\alpha,H,A}\phi_\alpha(u) = \alpha u - \Pi P_\alpha\begin{pmatrix} 0 \\ \binom{0}{\sqrt{\alpha}f}\end{pmatrix}.$$

To obtain a more useful expression for the gradient, we employ an expression for P_α due to von Neumann ([16]) and used for the development of the Sobolev gradient theory in [10]:

$$P_\alpha = \begin{pmatrix} (I + T_\alpha^* T_\alpha)^{-1} & T_\alpha^*(I + T_\alpha T_\alpha^*)^{-1} \\ T_\alpha(I + T_\alpha^* T_\alpha)^{-1} & T_\alpha T_\alpha^*(I + T_\alpha T_\alpha^*)^{-1}\end{pmatrix}$$

where T_α^* is the adjoint of T_α as a closed and densely defined operator. Using this expression, the final form of the gradient is

$$\nabla_{\alpha,H,A}\phi_\alpha(u) = \alpha u - T_\alpha^*(I + T_\alpha T_\alpha^*)^{-1}\sqrt{\alpha}\begin{pmatrix}0\\ f\end{pmatrix}.$$

In [8] it is shown that

$$T_\alpha^*(I + T_\alpha T_\alpha^*)^{-1}\sqrt{\alpha}\begin{pmatrix}0\\ f\end{pmatrix} = \alpha MA^*(\alpha I + AMA^*)^{-1}f$$

where A^* is the adjoint of A when viewed as a closed densely defined operator on L, and M is the embedding operator for H and L; that is, $\langle Mx, y\rangle_H = \langle x, y\rangle_L$ for all $x \in L$ and $y \in H$. Thus

$$\nabla_{\alpha,H,A}\phi_\alpha(u) = \alpha u - \alpha MA^*(\alpha I + AMA^*)^{-1}f,$$

and, the gradient is zero for

$$u_\alpha = MA^*(\alpha I + AMA^*)^{-1}f. \tag{3.2}$$

Suppose that f is in the range of A and we seek a minimum-norm solution $u \in H$ to (1.1) as the limit of a sequence u_α as α approaches zero. If the L^2 metric were used for regularization, the limit of the sequence with convergence in the L^2 norm would not necessarily have the required smoothness of an element of H. With the Sobolev metric on the other hand, it follows from (3.2) that, for each α, u_α is in the range of M which is a subspace of H. Thus with convergence defined in the H norm, the limit of the sequence satisfies the desired regularity requirement.

We now make some remarks regarding properties of the operator $MA^*(\alpha I + AMA^*)^{-1}$.

Proposition 3.3. *Let $S_\alpha = (\alpha I + AMA^*)^{-1}$. Then the following statements hold:*

(1) *S_α is everywhere defined on K.*
(2) *S_α is a bounded linear operator from K to K with norm less than or equal to $(1/\alpha)$.*
(3) *MA^*S_α is a bounded linear operator from K to H with norm less than or equal to $(1/\sqrt{\alpha})$.*
(4) *$u_\alpha = MA^*(\alpha I + AMA^*)^{-1}f = (\alpha M^{-1} + A^*A)^{-1}A^*f$.*

Proof. Property 1 is proved in [8, Theorem 3]. To prove property 2, simply note that $\langle S_\alpha^{-1} f, f\rangle_K \geq \alpha\|f\|_K^2$ for all f in the domain of S_α^{-1}. For case 3, note that for a linear operator J and c a real number, $J(cI+J)^{-1} = I - c(cI+J)^{-1}$. Thus $AMA^*S_\alpha = I - \alpha S_\alpha$. Let $f \in K$. Then

$$\begin{aligned}\|MA^*S_\alpha f\|_H^2 &= \langle MA^*S_\alpha f, MA^*S_\alpha f\rangle_H = \langle A^*S_\alpha f, MA^*S_\alpha f\rangle_L \\ &= \langle S_\alpha f, AMA^*S_\alpha f\rangle_K = \langle S_\alpha f, f - \alpha S_\alpha f\rangle_K \\ &\leq \langle S_\alpha f, f\rangle_K \leq (1/\alpha)\|f\|_K^2.\end{aligned}$$

Thus $\|MA^*S_\alpha f\|_H \leq (1/\sqrt{\alpha})\|f\|_K$ and the assertion follows. Finally, to prove property 4, first note that since A is bounded on L by assumption, A^* is everywhere defined on K. Thus

$$A^*f = A^*S_\alpha^{-1}S_\alpha f = (\alpha M^{-1} + A^*A)MA^*(\alpha I + AMA^*)^{-1}f.$$

Applying $(\alpha M^{-1}+A^*A)^{-1}$ to both sides of the equation gives the desired result. $\square$

4. Convergence of the gradient flow

We consider now the gradient system

$$z(0) = u_0 \in H \text{ and } z'(t) = -\nabla_{\alpha,H,A}\phi_\alpha(z(t)). \tag{4.1}$$

We restate [10, theorems 4.1 and 7.1].

Theorem 4.1. *Suppose F is a C^1 real-valued function defined on a Hilbert space H, bounded below, and has a locally Lipschitzian derivative. Then for any initial state u_0, the gradient system* (3.1) *has a unique global solution.*

Theorem 4.2. *Suppose F is a nonnegative C^2 function on a Hilbert space and is convex in the sense of 2.1. Then the gradient system* (3.1) *has a unique global solution, and there exists $u \in H$ such that*

$$u = \lim_{t\to\infty} z(t) \quad \textit{and} \quad \nabla_H F(u) = 0,$$

where convergence is in the strong sense using the H norm. Further, the rate of convergence is exponential.

These two theorems establish that for any initial state u_0 the flow (4.1) for ϕ_α has a unique global solution and that the flow converges strongly in the H norm with an exponential convergence rate. Since ϕ_α is convex, this minimum is unique. We denote by u_α the unique minimizer of the energy obtained as the asymptotic limit of the gradient system. The expression in (3.2) shows that u_α is in the range of M and hence in H.

5. An example

Consider the problem of computing an approximation u to the first derivative f' of a smooth function $f : [a,b] \to \mathbb{R}$ given only $f(a)$ and a set of m discrete noise-contaminated data points $\{(x_i, y_i)\}$ with

$$a \leq x_1 \leq x_2 \leq \ldots \leq x_m \leq b,$$

and

$$y_i = f(x_i) + \eta_i, \quad (i = 1, \ldots, m)$$

for independent identically distributed zero-mean noise values η_i. This problem is the inverse of the problem of computing integrals and is ill-posed because the solution f' does not depend continuously on the data f. If we use divided difference approximations to derivative values, then small relative perturbations in the data can lead to arbitrarily large relative changes in the solution, and the discretized problem is ill-conditioned. We therefore require some form of regularization in order to avoid overfitting. Tikhonov regularization was first applied to the numerical differentiation problem in [3].

Let

$$\hat{f}(x) = f(x) - f(a)$$

for $x \in [a, b]$ so that $\hat{f}'(x) = f'(x)$ and $\hat{f}(a) = 0$. Then for $k = 0$, 1, or 2 and $f \in H^{k+1}(a, b)$ the problem is to find $u \in H^k(a, b)$ such that

$$Au(x) = \int_a^x u(t)\,dt = \hat{f}(x), \quad x \in [a, b]. \tag{5.1}$$

Note that A is a bounded operator from $L^2(a, b)$ to $L^2(a, b)$ since

$$\|Au\|_{L^2}^2 = \int_a^b \left(\int_a^x u(t)\,dt \right)^2 dx \le (b-a)^2 \int_a^b u^2.$$

We discretize the problem by partitioning the domain $[a, b]$ into n subintervals of length $\Delta t = (b - a)/n$, and representing the solution u by the n-vector of midpoint values

$$u_j = u(t_j + \Delta t/2), \ \ (j = 1, \dots, n)$$

for

$$t_j = a + (j-1)\Delta t, \ \ (j = 1, \dots, n+1).$$

The discretized system is then $A\mathbf{u} = \hat{\mathbf{y}}$, where $\hat{y}_i = y_i - f(a)$ and A_{ij} is the length of $[a, x_i] \cap [t_j, t_{j+1}]$:

$$A_{ij} = \begin{cases} 0 & \text{if } x_i \le t_j \\ x_i - t_j & \text{if } t_j < x_i < t_{j+1} \\ \Delta t & \text{if } t_{j+1} \le x_i. \end{cases}$$

This linear system may be underdetermined or overdetermined and is likely to be ill-conditioned. We therefore use a least squares formulation with Tikhonov regularization. We minimize the convex functional

$$\phi(\mathbf{u}) = \|A\mathbf{u} - \hat{\mathbf{y}}\|^2 + \alpha \|D\mathbf{u}\|^2, \tag{5.2}$$

where α is a nonnegative regularization parameter, $\|\cdot\|$ denotes the Euclidean norm, and D is a differential operator of order k defining a discretization of the H^k Sobolev norm:

$$D^t = \begin{cases} I & \text{if } k = 0 \\ (I \ D_1^t) & \text{if } k = 1 \\ (I \ D_1^t \ D_2^t) & \text{if } k = 2, \end{cases}$$

where I denotes the identity matrix, and D_1 and D_2 are first and second difference operators. We use second-order central differencing so that D_1 maps midpoint values to interior grid points, and D_2 maps midpoint values to interior midpoint values. The regularization (smoothing) parameter α defines a balance between fidelity to the data on the one hand, and the size of the solution norm on the other hand. The optimal value depends on the choice of norm. Larger values of k enforce more smoothness on the solution. Setting the gradient of ϕ to zero, we obtain a linear system with an order-n symmetric positive definite matrix:

$$(A^tA + \alpha D^tD)\mathbf{u} = A^t\hat{\mathbf{y}}.$$

In the case that the error norm $\|\eta\|$ is known, a good value of α is obtained by choosing it so that the residual norm $\|A\mathbf{u} - \hat{\mathbf{y}}\|$ agrees with $\|\eta\|$. This is Morozov's discrepancy principle ([9]).

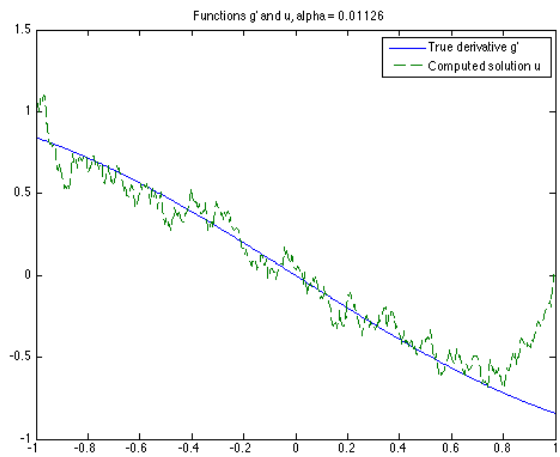

FIGURE 1. Computed approximation to $f', k = 0, f(x) = \cos(x)$

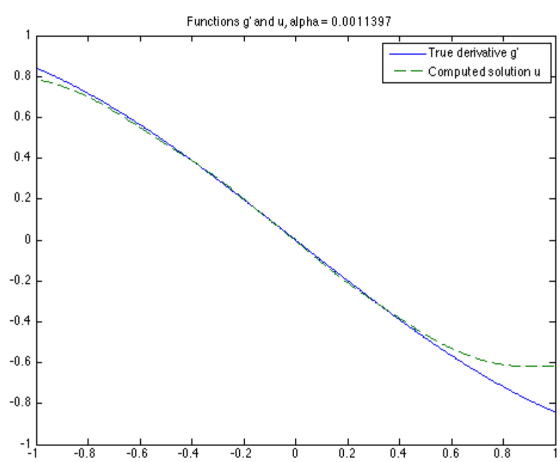

FIGURE 2. Computed approximation to $f', k = 1, f(x) = \cos(x)$

We chose the test function $f(x) = \cos(x)$ on $[-1, 1]$ and created a data set consisting of $m = 500$ points with uniformly distributed abscissae x_i and data

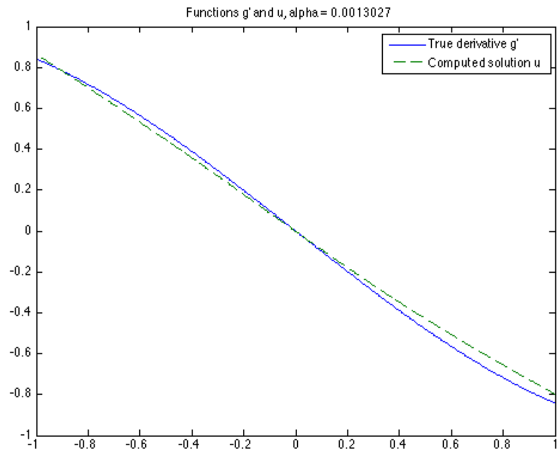

FIGURE 3. Computed approximation to $f', k = 2, f(x) = \cos(x)$

values $y_i = f(x_i) + \eta_i$, where η_i is taken from a normal distribution with mean 0 and standard deviation $\sigma = 0.1$. We used the known value of $\|\eta\|$ to compute the optimal parameter value α. The maximum relative error in the computed derivative decreased rapidly as k increased: $1.001, 0.2640$, and 0.0749 corresponding to $k = 0, k = 1$, and $k = 2$, respectively. The solutions are graphed in Figure 1, 2, and 3.

References

[1] R. A. Adams, J. F. Fournier; Sobolev Spaces, 2nd Edition, Academic Press, 2003.

[2] R. C. Aster, B. Borchers, C. H. Thurber; Parameter Estimation and Inverse Problems, second edition, Academic Press, 2013.

[3] J. Cullum; Numerical differentiation and regularization, SIAM J. Numer. Anal. 8 (1971), 254-265.

[4] H. W. Engl, M. Hanke, A. Neubauer; Regularization of Inverse Problems. Mathematics and Its Applications. Kluwer Academic Publishers, Dordrecht, 1996.

[5] J. Hadamard; Sur les problèmes aux dérivées partielles et leur signification physique (1902), 49–52.

[6] B. Hofmann, P. Mathé, H. von Weizsäcker; Regularization in Hilbert space under unbounded operators and general source condition, Inverse Problems, 25(11):115013 (15pp), 2009.

[7] S. I. Kabanikhin; Definitions and examples of inverse and ill-posed problems, J. Inv. Ill-Posed Problems 16 (2008), 317–357.

[8] P. Kazemi, R. J. Renka; A Levenberg-Marquardt method based on Sobolev gradients, Nonlinear Analysis 75 (2012) 6170–6179.

[9] V. A. Morozov; Choice of a parameter for the solution of functional equations by the regularization method, Sov. Math. Doklady 8 (1967), 1000–1003.

[10] Neuberger J. W.; Sobolev Gradients and Differential Equations, 2nd edition, Springer, 2010.

[11] A. G. Ramm; On unbounded operators and applications, Appl. Math. Lett. 21 (2008), 377–382.

[12] Riesz, Sz. Nagy; Functional Analysis, Dover Publications, Inc., 1990.

[13] A. N. Tikhonov; Regularization of incorrectly posed problems, Sov. Math. Dokl. 4 (1963), 1624-1627.

[14] A. E. Tikhonov, V. Y. Arsenin; Solutions of Ill-posed Problems, John Wiley & Sons, New York, 1977.

[15] A. N. Tikhonov, A. S. Leonov, A. G. Yagola; Nonlinear Ill-posed Problems. Number 14 in Applied Mathematics and Mathematical Computation. Chapman and Hall, London, 1998.
[16] J. von Neumann; Functional Operators II, Annls. Math. Stud., 22, 1940.

Parimah Kazemi
Department of Mathematics and Computer Science, Ripon College, P. O. Box 248, Ripon, WI 54971-0248, USA
E-mail address: parimah.kazemi@gmail.com

Robert J. Renka
Department of Computer Science & Engineering, University of North Texas, Denton, TX 76203-1366, USA
E-mail address: robert.renka@unt.edu