International Scholarly Research Network
ISRN Applied Mathematics
Volume 2011, Article ID 164564, 11 pages
doi:10.5402/2011/164564

Research Article

Numerical Differentiation of Noisy, Nonsmooth Data

Rick Chartrand

Theoretical Division, MS B284, Los Alamos National Laboratory, Los Alamos, NM 87545, USA

Correspondence should be addressed to Rick Chartrand, rickc@lanl.gov

Received 8 March 2011; Accepted 4 April 2011

Academic Editors: L. Marin and D. Xiao

Copyright © 2011 Rick Chartrand. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We consider the problem of differentiating a function specified by noisy data. Regularizing the differentiation process avoids the noise amplification of finite-difference methods. We use total-variation regularization, which allows for discontinuous solutions. The resulting simple algorithm accurately differentiates noisy functions, including those which have a discontinuous derivative.

1. Introduction

In many scientific applications, it is necessary to compute the derivative of functions specified by data. Conventional finite-difference approximations will greatly amplify any noise present in the data. Denoising the data before or after differentiating does not generally give satisfactory results (see an example in Section 4).

A method which does give good results is to regularize the differentiation process itself. This guarantees that the computed derivative will have some degree of regularity, to an extent that is often under control by adjusting parameters. A common framework for this is Tikhonov regularization [1] of the corresponding inverse problem. That is, the derivative of a function f, say on $[0, L]$, is the minimizer of the functional

$$F(u) = \alpha R(u) + DF(Au - f), \tag{1.1}$$

where $R(u)$ is a regularization or penalty term that penalizes irregularity in u, $Au(x) = \int_0^x u$ is the operator of antidifferentiation, $DF(Au - f)$ is a data fidelity term that penalizes discrepancy between Au and f, and α is a regularization parameter that controls the balance between the two terms.

The data fidelity term DF is most commonly the square of the L^2 norm, $DF = \int_0^L |\cdot|^2$, as is appropriate if f has additive, white Gaussian noise [2]. Other data fidelity terms can be used if the noise has a different, known distribution; see [3] for an alternative in the case of Poisson noise, and [4] for a more general approach.

In [1], the regularization term is the squared L^2 norm; this controls the size of u, without forcing minimizers to be regular. Tikhonov regularization was first applied to numerical differentiation by Cullum [5], where the regularization is the squared H^1 norm, $R(u) = \int_0^L |u'|^2$. This forces minimizers to be continuous, as is required for the H^1 norm to be finite. This prevents the accurate differentiation of functions with singular points.

Other variational methods have the same drawback of forcing smoothness. An approach that penalizes the L^2 norm of u'' forces the minimizer to be a cubic spline (see [6–8]). The variational approach of Knowles and Wallace [9] does not fall into the category of Tikhonov regularization, but explicitly assumes that u is smooth.

2. Total-Variation Regularization

We propose to use total-variation regularization in (1.1). We will thus compute the derivative of f on $[0, L]$ as the minimizer of the functional

$$F(u) = \alpha \int_0^L |u'| + \frac{1}{2} \int_0^L |Au - f|^2. \tag{2.1}$$

We assume $f \in L^2$ (an empty assumption in the discrete case), and for convenience that $f(0) = 0$ (in practice we simply subtract $f(0)$ from f). The functional F is defined on $BV[0, L]$, the space of functions of bounded variation. It is in fact continuous on BV, as BV is continuously embedded in L^2, and A is continuous on L^2 (being an integral operator with bounded kernel). Existence of a minimizer for F follows from the compactness of BV in L^2 [10, page 152] and the lower semicontinuity of the BV seminorm [10, page 120]. This and the strict convexity of F are sufficient to guarantee that F has a unique minimizer u^*.

Use of total variation accomplishes two things. It suppresses noise, as a noisy function will have a large total variation. It also does not suppress jump discontinuities, unlike typical regularizations. This allows for the computation of discontinuous derivatives, and the detection of corners and edges in noisy data.

Total-variation regularization is due to Rudin et al. in [11]. It has since found many applications in image processing. Replacing A in the two-dimensional analog of (2.1) with the identity operator gives a method for denoising an image f; see [12, 13] for an example where A is the Abel transform, giving a method for regularizing Abel inversion.

3. Numerical Implementation

A simple approach to minimizing (2.1) is gradient descent. This amounts to evolving to stationarity the PDE obtained from the Euler-Lagrange equation:

$$u_t = \alpha \frac{d}{dx} \frac{u'}{|u'|} - A^T(Au - f), \tag{3.1}$$

where $A^T v(x) = \int_x^L v$ is the L^2-adjoint of A. Replacing the $|u'|$ in the denominator with $\sqrt{(u')^2 + \epsilon}$ for some small $\epsilon > 0$ avoids division by zero. Typically, (3.1) is implemented with explicit time marching, with u_t discretized as $(u_{n+1} - u_n)/\Delta t$ for some fixed Δt.

The problem with (3.1) is that convergence is slow. A faster algorithm is the lagged diffusivity method of Vogel and Oman [14]. The idea is to replace at each iteration of (3.1) the nonlinear differential operator $u \mapsto (d/dx)(u'/|u'|)$ with the linear operator $u \mapsto (d/dx)(u'/|u'_n|)$. The algorithm has been proven to converge to the minimizer of F [15].

We consider two discrete implementations of the lagged diffusivity algorithm. The first uses explicit matrix constructions, and is faster for smaller problems, but becomes impractical for data of more than a few thousand points. We assume that u is defined on a uniform grid $\{x_i\}_0^L = \{0, \Delta x, 2\Delta x, \ldots, L\}$. Derivatives of u are computed halfway between grid points as centered differences, $Du(x_i + \Delta x/2) = u(x_{i+1}) - u(x_i)$. This defines our $L \times (L+1)$ differentiation matrix D. This approach avoids the consideration of boundary conditions for differentiation, and we find it gives better algorithmic results at the boundary. Integrals of u are likewise computed halfway between grid points, using the trapezoid rule to define an $L \times (L+1)$ matrix A. Let E_n be the diagonal matrix whose ith entry is $((u_n(x_i) - u_n(x_{i-1}))^2 + \epsilon)^{-1/2}$. Let $L_n = \Delta x D^T E_n D$, $H_n = K^T K + \alpha L_n$. The matrix H_n is an approximation of the Hessian of F at u_n. The update $s_n = u_{n+1} - u_n$ is the solution to $H_n s_n = -g_n$, where $g_n = K^T(Ku_n - f) + \alpha L_n u_n$, and we solve the equation using MATLAB's backslash operator. For further algorithm details, see [16].

For larger problems, we use a modified version that avoids explicit matrices and uses simpler numerics. The differentiation matrix D is constructed as the square, sparse matrix for simple forward differences, with periodic boundary conditions. We use a function handle for A, simply using MATLAB's cumulative sum operator, as well as for A^T. It is important that the discretization of A^T be consistent with the discretization of A, in the sense that if one were to use the function handles to construct matrices, then they would be transposes of each other. In other respects, the algorithm is as above, except the definition of E_n now uses periodic boundary conditions, and the equation $H_n s_n = -g_n$ is solved using preconditioned conjugate gradient. For the preconditioner, we perform incomplete-Cholesky factorization on the sum of αL_n and the diagonal matrix whose entries are the row sums of $A^T A$, these sums being computable analytically.

Less straightforward is the choice of the regularization parameter α. One approach is to use the discrepancy principle: the mean-squared difference between Au^* and f should equal the variance of the noise in f. This has the effect of choosing the most regular solution of the ill-posed inverse problem $Au = f$ that is consistent with the data f. In practice, the noise in f is not generally known, but the noise variance can be estimated by comparing f with a smoothed version of f. The other approach is simply to use trial and error, adjusting α to obtain the desired balance between suppression of oscillations and fidelity to the data. In the next section, we will see an example showing the effect of the value of α.

4. Examples

4.1. A Simple Nonsmooth Example

Let $f_0(x) = |x - 1/2|$, defined at 100 evenly spaced points in $[0, 1]$. We obtain f by adding Gaussian noise of standard deviation 0.05. Figure 1 shows the resulting f. First, we show in Figure 2 the result of computing f' by simple centered differencing. The noise has been greatly amplified, so much that denoising the result is hopeless.

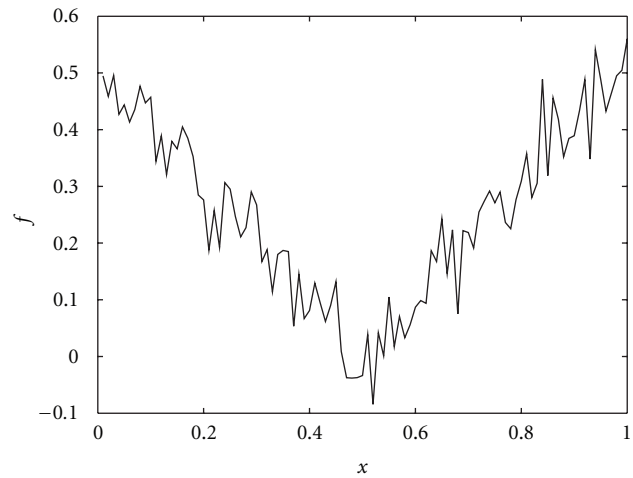

Figure 1: The function f, obtained from $|x - 1/2|$ by adding Gaussian noise of standard deviation 0.05.

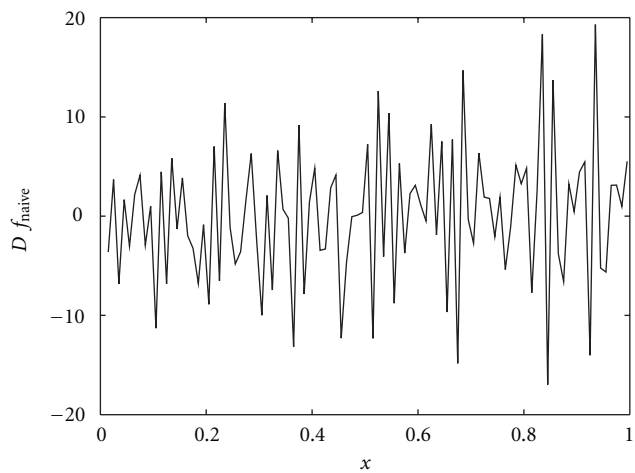

Figure 2: Computing f' with finite differences greatly amplifies noise.

We compare this with the result in Figure 3 of denoising f before computing f' by differencing. The denoising is done by H^1 regularization, minimizing

$$\alpha \int_0^L |u'|^2 + \frac{1}{2} \int_0^L |u - f|^2, \tag{4.1}$$

an appropriate denoising mechanism for continuous functions. We use $\alpha = 3.5 \times 10^{-3}$, using the discrepancy principle, as this results in the L^2 norm of $u^* - f$ being 0.5, the expected value

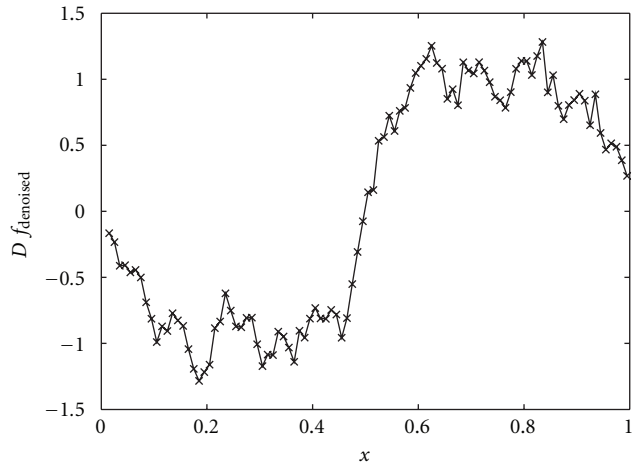

Figure 3: The function f is denoised, then differentiated with finite differences. The result is noisy and inaccurate.

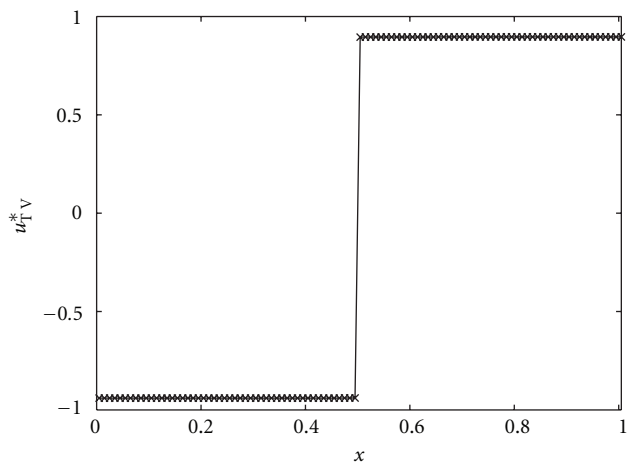

Figure 4: Regularizing the differentiation process with total-variation produces a noiseless derivative with a correctly located, sharp jump. The discrepancy of the values from ±1 are due to contrast loss, an artifact of total variation methods in the presence of noise.

of the L^2 norm of the noise vector $f - f_0$. The residual noise in the denoised f is still amplified enough by the differentiation process to give an unsatisfactory result.

Now, we implement our total-variation regularized differentiation, (2.1). We use the matrix-based version described above, using $\alpha = 0.2$ and $\epsilon = 10^{-6}$, initializing with the naive derivative (specifically `[0; diff( f ); 0]`, to obtain a vector of the appropriate size). Although convergence is nearly complete after 100 iterations, the points closest to the jump move much more slowly, adopting their final positions after 7000 iterations. This takes just 13.1 s, running on a conventional dual-core desktop. The result is in Figure 4. The overall shape of f_0' is captured almost perfectly. The jump is correctly located. The one inaccuracy is

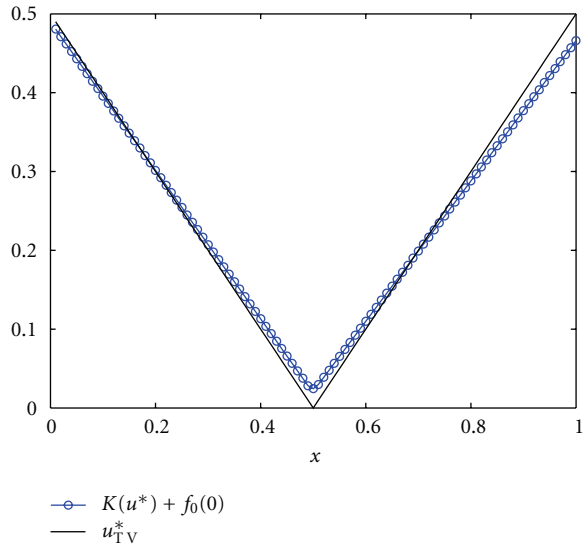

Figure 5: The function f_0 (solid line) and the antidifferentiated numerical derivative (circles). The numerically computed function is very similar to the exact one.

the size of the jump: there is a loss of contrast, which is typical of total-variation regularization in the presence of noise. Decreasing the size of the jump reduces the penalty term in (2.1), at the expense of increasing the data-fidelity term.

We also show the result of applying the antidifferentiation operator to the computed f', and compare with f_0 in Figure 5. The corner is sharp and the lines are straight, though a little too flat.

4.2. Respirometry Data

Now, we consider data obtained from a whole-room calorimeter, courtesy of Edward L. Melanson of the University of Colorado Denver. The metabolic rate of a subject within the calorimeter can be determined via respirometry, the measurement of oxygen consumption, and carbon dioxide production within the room. The raw data traces need to undergo response corrections in order to be useful, which involves differentiation. Quoting [17]:

> In essence, the first derivative of the trace is calculated, multiplied by a constant derived from the volume of the room and the volumetric flow rate, and added to the original data. Because of the long time constant of the room (5 h), the multiplicative constant is very large. Consequently, any significant noise in the derivatized data will overwhelm the original trace.

Thus, we see the need to regularize the differentiation process for this application.

In Figure 6 there is an example of the raw data for the oxygen consumption rate, which needs to be differentiated for the purposes of response correction. The data consists of samples taken every second for most of a day, for a total of 82,799 samples. In Figure 7

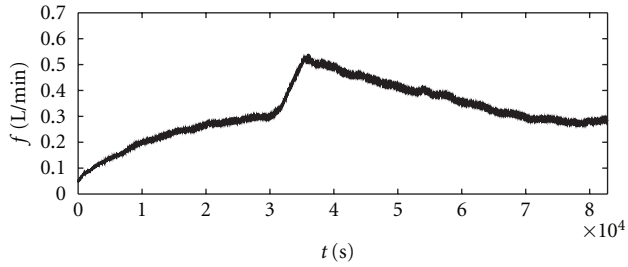

Figure 6: The trace of raw oxygen consumption data, consisting of 82,799 samples, one per second.

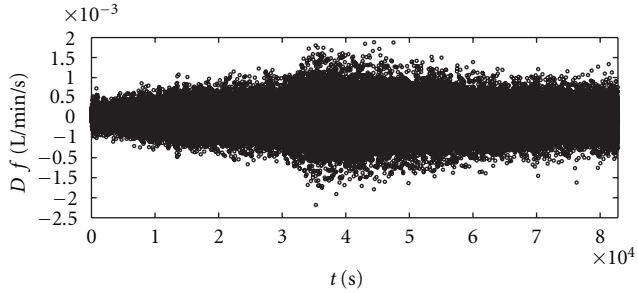

Figure 7: Computing the derivative with finite differences gives a useless result.

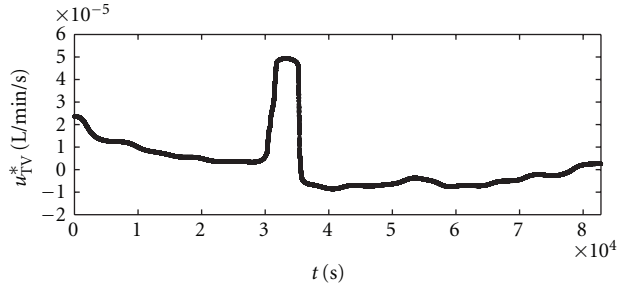

Figure 8: The derivative computed with strong TV regularization. Despite the heavy smoothing, the rapid rise and fall in the derivative is captured well.

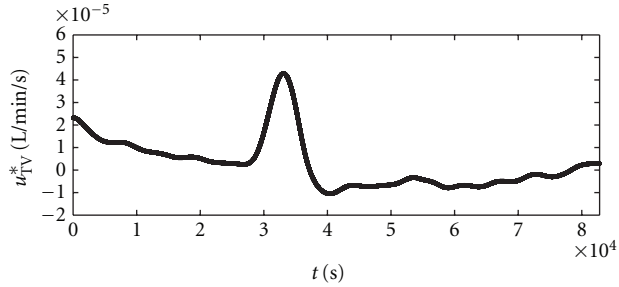

Figure 9: Using H^1 regularization, the jumps in the derivative are smoothed away.

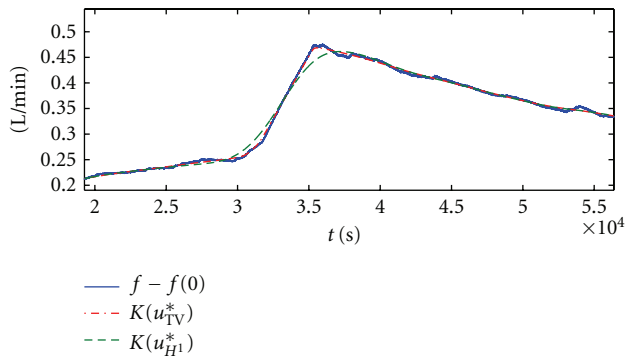

Figure 10: Integrating the derivatives and comparing with the original function, we find the TV result follows the curve more closely near the peak. Both curves follow the curve away from the peak, but ignore small fluctuations, to a degree determined by α.

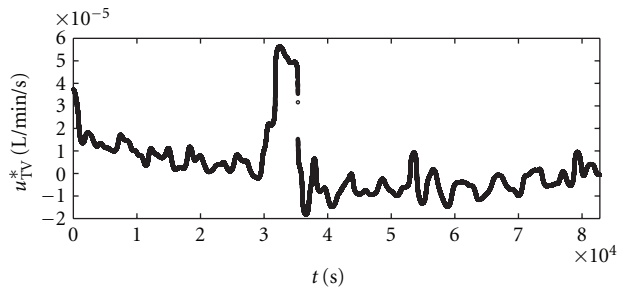

Figure 11: The derivative computed with lesser TV regularization, preserving more structure, including a discontinuity in the derivative.

there is the result of computing the unregularized, finite-difference derivative. If restricted to the same vertical range as the following plots, the plot would be a solid black rectangle.

The data size is much too large for the explicit-matrix implementation, so we use the implicit approach. In each case, we use 60 iterations, taking about 5 minutes. We compare the total-variation regularized derivative with that computed with H^1 regularization, for two different regularization strengths. First, a stronger regularization, with a value of $\alpha = 0.1$ for the TV case. The result is in Figure 8. We then adjust the parameter for H^1 regularization until the curve matches the TV result away from the large bump, namely, $\alpha = 500$; see Figure 9. A value of $\epsilon = 10^{-8}$ was used in both cases. The TV regularization is much more tolerant of rapid rises and falls while the H^1 result smooths away this information. We also compare the results of antidifferentiating the derivative and comparing with the original trace, with Figure 10 displaying a zoomed-in portion. The H^1 curve is unable to conform to the peak, as the H^1 regularization term heavily penalizes what would be the curvature of the curve in this figure. Away from the peak, the integrated derivatives follow the original trace, but not too closely, ignoring small-scale fluctuations. This is the effect of the regularization, with the choice of α serving to determine the scale of fluctuation that is considered insignificant.

The above result would be appropriate if the rapid rise and fall in the derivative corresponded to the only feature of interest. Now, we examine the result of weaker

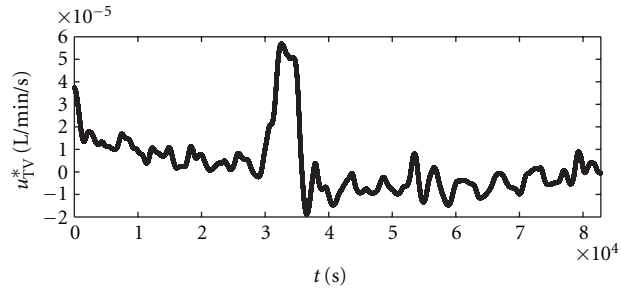

Figure 12: Using H^1 regularization, the discontinuity in the derivative is smoothed away.

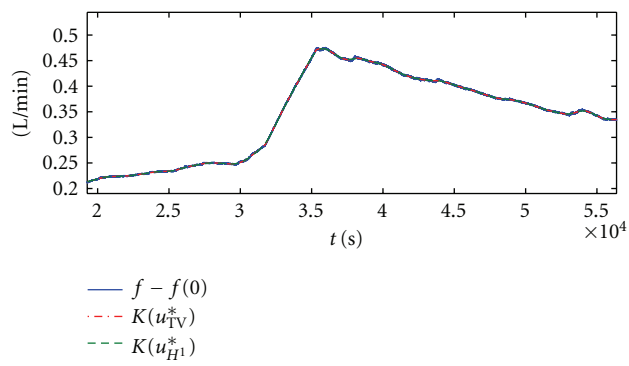

Figure 13: Integrating the derivatives and comparing with the original function, we find that with the weaker regularization, the integrals follow the curve closely, effectively considering much lesser fluctuations to be significant.

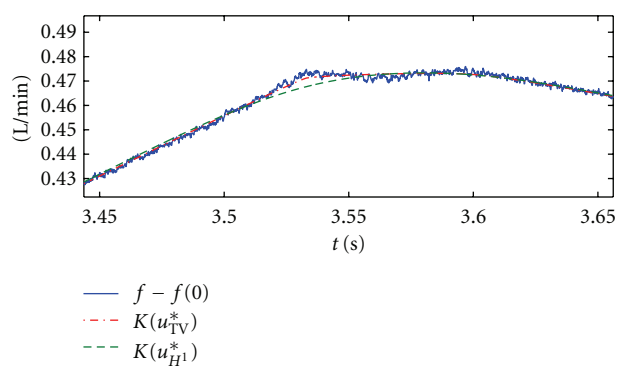

Figure 14: A closer look reveals that unlike the TV result, the H^1 curve cannot follow the high curvature at the corner, preventing the computed derivative from dropping discontinuously.

regularization, so as to preserve smaller-scale features. In this instance, we use $\alpha = 10^{-3}$ for the TV regularization. As before, we choose α for the H^1 regularization so that the result matches the TV result over most of the time period, in this case $\alpha = 1$. This time a slightly larger value of $\epsilon = 10^{-6}$ was required, in order to offset the poorer conditioning of the linear system solved

in the lagged diffusivity algorithm when α is smaller (in general, there is a tradeoff between better conditioning with larger values of ϵ and greater accuracy with smaller values).

Figures 11 and 12 show the results. Both derivatives capture more oscillations, including more structure in the rapid rise and fall. But the TV result is able to capture a discontinuity in the derivative that the H^1 result smooths away. When we compare the antiderivatives with the original function (Figure 13), we find that they follow the trace much more closely, conforming to smaller-scale fluctuations. Zooming in further, we see the greater curvature penalty on the H^1 curve prevents it from following the sharp corner, thus missing the discontinuous drop in the derivative.

5. Conclusions

We presented a method for regularizing the numerical derivative process, using total-variation regularization. Unlike previously developed methods, the TV method allows for discontinuities in the derivatives, as desired when differentiating data corresponding to nonsmooth functions. We used the lagged diffusivity algorithm, which enjoys proven convergence properties, with one implementation that works rapidly for small problems, and a second more suitable for large problems. The TV regularization allows the derivative to capture more features of the data, while adjusting the regularization parameter controls the scale of fluctuations in the data that are ignored.

Acknowledgments

This paper was supported by the U.S. Department of Energy through the LANL/LDRD Program. The authors wishes to thank John Lighton of Sable Systems International and Edward L. Melanson of the University of Colorado Denver for the data used in the examples.

References

[1] A. N. Tikhonov, "Regularization of incorrectly posed problems," *Soviet Mathematics-Doklady*, vol. 4, pp. 1624–1627, 1963.

[2] M. Green, "Statistics of images, the TV algorithm of Rudin-Osher-Fatemi for image denoising and an improved denoising algorithm," Tech. Rep. 02-55, UCLA Computational and Applied Mathematics, 2002.

[3] T. Le, R. Chartrand, and T. J. Asaki, "A variational approach to reconstructing images corrupted by Poisson noise," *Journal of Mathematical Imaging and Vision*, vol. 27, no. 3, pp. 257–263, 2007.

[4] R. Chartrand and V. Staneva, "Total variation regularisation of images corrupted by non-Gaussian noise using a quasi-Newton method," *IET Image Processing*, vol. 2, no. 6, pp. 295–303, 2008.

[5] J. Cullum, "Numerical differentiation and regularization," *SIAM Journal on Numerical Analysis*, vol. 8, pp. 254–265, 1971.

[6] I. J. Schoenberg, "Spline functions and the problem of graduation," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 52, pp. 947–950, 1964.

[7] C. H. Reinsch, "Smoothing by spline functions. I, II," *Numerische Mathematik*, vol. 10, pp. 177–183, 1967.

[8] M. Hanke and O. Scherzer, "Inverse problems light: numerical differentiation," *The American Mathematical Monthly*, vol. 108, no. 6, pp. 512–521, 2001.

[9] I. Knowles and R. Wallace, "A variational method for numerical differentiation," *Numerische Mathematik*, vol. 70, no. 1, pp. 91–110, 1995.

[10] L. Ambrosio, N. Fusco, and D. Pallara, *Functions of Bounded Variation and Free Discontinuity Problems*, Oxford Mathematical Monographs, The Clarendon Press Oxford University Press, New York, NY, USA, 2000.

[11] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, no. 1–4, pp. 259–268, 1992.
[12] T. J. Asaki, R. Chartrand, K. R. Vixie, and B. Wohlberg, "Abel inversion using total-variation regularization," *Inverse Problems*, vol. 21, no. 6, pp. 1895–1903, 2005.
[13] T. J. Asaki, P. R. Campbell, R. Chartrand, C. E. Powell, K. R. Vixie, and B. E. Wohlberg, "Abel inversion using total variation regularization: applications," *Inverse Problems in Science and Engineering*, vol. 14, no. 8, pp. 873–885, 2006.
[14] C. R. Vogel and M. E. Oman, "Iterative methods for total variation denoising," *SIAM Journal on Scientific Computing*, vol. 17, no. 1, pp. 227–238, 1996.
[15] D. C. Dobson and C. R. Vogel, "Convergence of an iterative method for total variation denoising," *SIAM Journal on Numerical Analysis*, vol. 34, no. 5, pp. 1779–1791, 1997.
[16] C. R. Vogel, *Computational Methods for Inverse Problems*, vol. 23 of *Frontiers in Applied Mathematics*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, Pa, USA, 2002.
[17] E. L. Melanson, J. P. Ingebrigtsen, A. Bergouignan, K. Ohkawara, W. M. Kohrt, and J. R. B. Lighton, "A new approach for flow-through respirometry measurements in humans," *American Journal of Physiology*, vol. 298, no. 6, pp. R1571–R1579, 2010.