AUTOMATIC STOPPING CRITERION FOR ANISOTROPIC DIFFUSION

Victor Solo

School of Electrical Engineering
University of New South Wales
Sydney, NSW 2052, AUSTRALIA
email: vsolo@syscon.ee.unsw.edu.au

ABSTRACT

We develop, apparently for the first time, an automatic criterion to choose when to stop the iteration in anisotropic diffusion signal reconstruction.

1. INTRODUCTION

Anisotropic diffusion (ANDI) has become a valuable tool for multiscale nonlinear image analysis for example in edge detection and segmentation. Since it was proposed by [6] there has been a considerable development, especially in theoretical understanding e.g. [15], the special issue in the IEEE Transactions on Image Processing including, [14],[2], [1]. In particular a considerable understanding of the effect of different diffusion coefficients on convergence has been developed e.g. [15],[2]. Also there are close connexions with total variation denoising [8] and smoothed versions of it [13],[3], [12].

A fundamental feature of the ANDI procedure is the necessity to decide when to terminate the iteration. Typically as the diffusion is stepped forward in time (or iteration) the reconstruction improves, then stabilizes and then as time wears on it degrades (gets over-smoothed). The iteration counter or time parameter is in fact a regularizing or tuning parameter for the nonlinear ANDI reconstruction procedure which solves the ill-conditioned inverse problem of estimating a possibly discontinuous signal. As with other regularization methods it is natural to seek an automatic stopping rule. In fact no such methods seem to have been developed for ANDI and our aim here is to develop such an automatic selection rule.

2. ANDI

We treat the one-dimensional problem for simplicity. With continuous data $y(x)$ the ANDI algorithm generates the reconstruction as the solution $\hat{f}(t,x)$ to the inhomogeneous heat equation

$$\frac{\partial \hat{f}}{\partial t} = \frac{\partial}{\partial x}(c(|\frac{\partial \hat{f}}{\partial x}|)\frac{\partial \hat{f}}{\partial x}) \quad (2.1)$$

$$\hat{f}(0,x) = y(x) \quad (2.2)$$

where $c(\cdot)$ is the conduction parameter or diffusion coefficient. A number of authors [2],[15] have pointed out that the ANDI algorithm can be interpreted as a steepest descent algorithm for minimising a nonquadratic criterion

$$J(f) = \int_0^1 \rho(|f(x)|)dx$$

with the start value (2.2). Here $\rho(\cdot)$ is a potential function and then $c(|\xi|) = \frac{\psi(|\xi|)}{|\xi|}$ where $\psi(\xi) = \rho^{'}(\xi)$ is the influence function (a term borrowed from robust statistics)[2]. The ANDI algorithm can thus be written

$$\frac{\partial \hat{f}}{\partial t} = \frac{\partial}{\partial x}(\psi(|\frac{\partial \hat{f}}{\partial x}|)sign(\frac{\partial \hat{f}}{\partial x}))$$

If $\psi(\xi)$ is an odd function then this becomes

$$\frac{\partial \hat{f}}{\partial t} = \frac{\partial}{\partial x}(\psi(\frac{\partial \hat{f}}{\partial x}))$$

Now as made clear by [15][2] the potential function has a crucial impact on the algortihm behaviour. An important set of results of [15] are that:

Theorem.
If $\psi(\infty) = 0$ then the algorithm is ill-posed in that there are an infinite number of global minima.
If $\psi(\infty) \neq 0$ then the algorithm is well-posed in that there is only one stationary point, namely the constant function (or image).
If $\rho(\xi)$ is convex and $\psi(\infty) \neq 0$ then $J(f)$ has a unique global minimum namely the constant image.

It follows that the conduction functions used by [6] namely those corresponding to

$$\psi(\xi) = \xi e^{-\xi^2/\gamma^2}$$

This work was partly supported by the Australian Research Council

$$\psi(\xi) = \frac{\xi}{1+\xi^2/\gamma^2}$$

both give ill-posed ANDI.

A number of well-posed conduction functions have been developed e.g. [15]. We introduce a new potential applied by [13],[3] to the related smoothed total variation denoising reconstruction procedure,

$$\rho(\xi) = \sqrt{\xi^2+\gamma^2} - \gamma$$

The associated influence and conduction functions are easily seen to be

$$\begin{aligned}\psi(\xi) &= \frac{\xi}{\sqrt{\xi^2+\gamma^2}}\\ c(\xi) &= \frac{1}{\sqrt{\xi^2+\gamma^2}}.\end{aligned}$$

We note that this potential function is convex and well-posed and so leads to a well-posed algorithm.

For application to noisy data we deal with a discretised version. Consider then the problem of estimating the possibly discontinuous signal $f(t)$ on $[0,1]$ from noisy data

$$y_i = f(\frac{i}{N}) + \epsilon_i, \quad i = 1, \cdots, N$$

where ϵ_i is a white Gaussian noise of variance σ^2 and zero mean.

Because the standard explicit method has stability problems (that are troublesome for the calculations in the next section) we have developed an iterative Crank-Nicholson implicit type scheme [9]. The typical explicit scheme would be (following e.g. [8]) (Here k is a time or iteration counter while i is a spatial coordinate)

$$\begin{aligned}\frac{f_i^{(k+1)} - f_i^{(k)}}{\delta} &= \nabla_- \psi_i^{(k)} \qquad (2.3)\\ &= (\psi_i^{(k)} - \psi_{i-1}^{(k)})\frac{1}{N^{-1}}\\ \psi_i^{(k)} &= \psi(\nabla_+ f_i^{(k)})\\ \nabla_+ f_i^{(k)} &= (f_{i+1}^{(k)} - f_i^{(k)})\frac{1}{N^{-1}}\end{aligned}$$

We can write this as

$$\begin{aligned}\frac{f_i^{(k+1)} - f_i^{(k)}}{\delta} &= \frac{1}{N^{-1}}(c_i^{(k)}\nabla_+ f_i^{(k)} - c_{i-1}^{(k)}\nabla_+ f_{i-1}^{(k)})\\ c_i^{(k)} &= c(\nabla_+ f_i^{(k)})\end{aligned}$$

We now replace this with the implicit scheme

$$\begin{aligned}&\frac{f_i^{(k+1)} - f_i^{(k)}}{\delta}\\ =\ & \frac{1}{2N^{-1}} c_i^{(k)}(\nabla_+ f_i^{(k+1)} + \nabla_+ f_i^{(k)})\\ -\ & \frac{1}{2N^{-1}} c_{i-1}^{(k)}(\nabla_+ f_{i-1}^{(k+1)} + \nabla_+ f_{i-1}^{(k)})\end{aligned}$$

Introducing the average

$$a_i^{(k+1)} = \frac{f_i^{(k+1)} + f_i^{(k)}}{2}$$

enables us to rewrite the update as (with $\rho = \delta N^2$)

$$\begin{aligned}&a_i^{(k+1)} = f_i^{(k)}\\ +\ & \frac{\rho c_i^{(k)}}{2}(a_{i+1}^{(k+1)} - a_i^{(k+1)}) - \frac{\rho c_{i-1}^{(k)}}{2}(a_i^{(k+1)} - a_{i-1}^{(k+1)})\end{aligned}$$

And this suggests the iteration (in n) for $f_i^{(k+1)}$

$$\begin{aligned}&a_i^{(n+1)} = f_i^{(k)}\\ +\ & \frac{\rho c_i^{(k)}}{2}(a_{i+1}^{(n)} - a_i^{(n+1)}) - \frac{\rho c_{i-1}^{(k)}}{2}(a_i^{(n+1)} - a_{i-1}^{(n)})\end{aligned}$$

After some reorganisation this can be written as a weighted average

$$\begin{aligned}&a_i^{(n+1)} = f_i^{(k)} \qquad (2.4)\\ +\ & w_i(a_{i+1}^{(n)} - f_i^{(k)}) + v_i(a_{i-1}^{(n)} - f_i^{(k)})\\ w_i &= \frac{\rho c_i^{(k)}/2}{1+\rho(c_i^{(k)} + c_{i-1}^{(k)})/2}\\ v_i &= \frac{\rho c_{i-1}^{(k)}/2}{1+\rho(c_i^{(k)} + c_{i-1}^{(k)})/2}\end{aligned}$$

We then recover $f_i^{(k+1)} = 2a_i^\infty - f_i^{(k)}$. The algorithm usually converges in a few iterations however; we use reflection boundary conditions. The algorithm collapses to the Jacobi iteration in [9] in the homogeneous case when the conduction function is the identity, however the formulation here as a weighted average seems to be new. In the homogeneous case convergence holds for all values of ρ whereas the explicit algorithm converges only for $\rho \le \frac{1}{2}$.

3. TUNING PARAMETER SELECTION

Here we regard diffusion time $h = k\delta$ as a regularizing or tuning parameter to be chosen. We use the symbol h here rather than t to emphasize that h is really a regularizing characteristic feature size. If h is too small, the reconstructed signal is very noisy; if h is too large it is smooth and discontinuities are lost. There are numerous methods that have been developed for choosing regularizing parmeters in ill-conditioned inverse problems. These are reviewed in [10] where the approach to be applied here was developed. We note that methods such as AIC [5] or MDL [7] are not obviously applicable because they require the tuning parameter to be a model dimension. A method such as cross-validation [5] would be computationally prohibitive because it requires that ANDI be repeated over and over as data points are left out one at a time.

We use a simple quadratic measure of reconstruction quality. At time h (or iteration k) this risk function is defined as

$$R_h = E||f - f^{(k)}||^2 = E\int_0^1 (f - f^{(k)})^2 dx$$

The discrete version is then

$$R_h = \frac{1}{N}\Sigma_1^N E(f(\frac{i}{N}) - f^{(k)}(\frac{i}{N}))^2$$

Ideally we would like to choose the stopping time h to minimize R_h. However R_h cannot be computed if only because $f(x)$ is unknown.

The idea is to find an empirically computable, statistically unbiassed estimator of R_h and minimize that instead. Using only the Gaussian assumption and a simple integration by parts argument it can be shown [10] that an empirically computable unbiassed estimator of R_h is

$$\hat{R}_h = \frac{1}{N}\Sigma_1^N e_i^2 - \sigma^2 + \frac{2\sigma^2}{N}\omega_h$$

$$\omega_h = trace(\frac{\partial f^{(k)T}}{\partial y})$$

$$e_i = y_i - f^{(k)}(\frac{i}{N})$$

The idea then is to plot $\hat{R}_h$ for a minimum in h. This is called SURE (Stein's unbiassed risk estimator for the originator, in a different context, of the integration by parts argument: see [10]).

In the current setting we generate $\frac{\partial f^{(k)T}}{\partial y}$ iteration by iteration. Differentiate through (2.3) and set $g_i = \frac{\partial f^{(k)T}}{\partial y}$ to find for $i = 1, \cdots, N$

$$\frac{g_i^{(k+1)} - g_i^{(k)}}{\delta} \quad (3.1)$$

$$= N^2((g_{i+1}^{(k)} - g_i^{(k)})\dot{\psi}_i^{(k)} - (g_i^{(k)} - g_{i-1}^{(k)})\dot{\psi}_{i-1}^{(k)}))$$

with boundary conditions $\dot{\psi}_0^{(k)} = 0 = \dot{\psi}_{N+1}^{(k)}$ where

$$\dot{\psi}(\xi) = \frac{d}{d\xi}\psi(\xi) = \frac{\gamma^2}{(\xi^2 + \gamma^2)^{3/2}}$$

This algorithm exhibits very slow convergence as well as stability problems and so we use a natural extension of the algorithm described in section 2 instead. The computation then consists of advancing (2.4) and the iterative extension of (3.1) in parallel and computing $\hat{R}_h$ at each step. We then plot $\hat{R}_h$ to find the minimizing h value. We emphasize again that the approach based on SURE does not rely on any particular algorithm, potential function or the one-dimensional nature of the illustration used here.

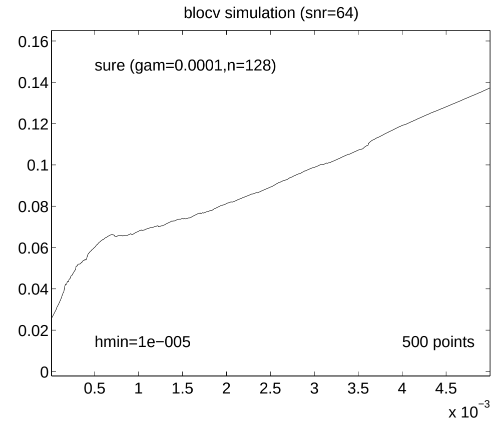

Fig. 1. Plot of SURE for blocky function showing shallow local minimum.

4. RESULTS

In Fig.1 is a plot of SURE for the blocky function used by [13],[11] with N=128 points. We chose $\gamma = .0001, snr = 64$ to compare with their results. Here signal to noise ratio (snr) is a power ratio . Since ANDI starts from the data the error sum of squares is zero there (but the trace term is large). As the fitting proceeds the error sum of squares rises bu the trace drops. Thus the minimising value of SURE is the first local minimum. This kind of behaviour is known for tuning parameter selection for linear estimators [4]. In Fig.1 this first local minimum is rather shallow but evident; this means that several values of h in that vicinity should be tried. In Fig.2 we show the reconstruction corresponding to the local minimising $h = .00065$. The reconstruction is still somewhat noisy but the narrow block is well reconstructed although some bias shows. Bias is also evident in the reconstruction of the top level of the larger block. For comparison in Fig.3 we show the reconstruction towards the higher end of the flat area of the local minimum. There is not a lot to choose between the two reconstructions. The larger value of h provides a slightly smoother reconstruction with a little less bias.

5. SUMMARY

In this paper we have presented an automatic method of tuning parameter choice for anisotropic diffusion, apparently for the first time. The method has modest computational requirements, whereas for example, cross-validation is computationally prohibitive. Future work will deal with estimation of the noise variance, extension of the technique to handle correlated noise and some theoretical performance analysis of the method.

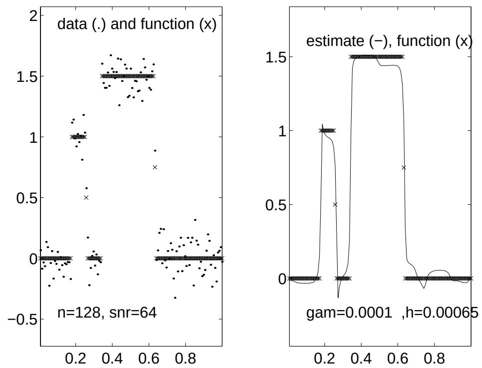

Fig. 2. Plot of data and estimate, $h = .00065$

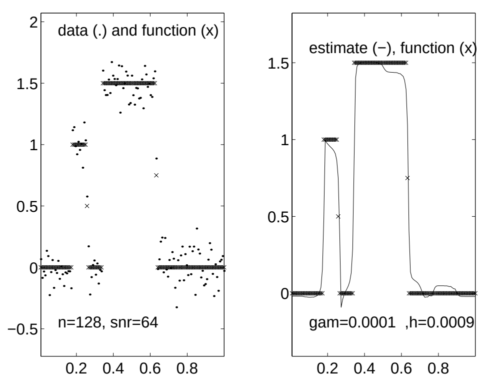

Fig. 3. Plot of data and estimate, $h = .00090$

6. REFERENCES

[1] S Acton. Multigrid anisotropic diffusion. *IEEE Trans. Im.Proc.*, 7:280–291, 1998.

[2] M.J. Black, G. Sapiro, D.H. Marimont, and D. Heeger. Robust anisotropic diffusion. *IEEE Trans. Im.Proc.*, 7:421–432, 1998.

[3] P. Charbonnier, L. Blanc-Feraud, G. Aubert, and M. Barlaud. Deterministic edge-preserving regularization in computed imaging. *IEEE Trans. Im.Proc.*, 5, 1997.

[4] J. Fan and I Gijbels. *Local Polynomial Modelling and its applications*. Chapman and Hall, London, 1996.

[5] H. Linhart and W. Zucchini. *Model selection*. J. Wiley, New York, 1986.

[6] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE. Trans. Patt. Anal. Machine Intell.*, 12:629–639, 1990.

[7] J Rissanen. *Stochastic Complexity in Statistical Enquiry*. World Scientific, Singapore, 1989.

[8] L.I. Rudin, S.Osher, and E.Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.

[9] G.D. Smith. *Numerical Solution of Partial Differential Equations*. Oxford University Press, London, 1975.

[10] V. Solo. A SURE-fired way to choose smoothing parameters in ill-conditioned inverse problems. In *Proc. IEEE ICIP96*. IEEE, IEEE Press, 1996.

[11] V Solo. Selection of regularization parameters for total variation denoising. In *Proc. ICASSP99*. IEEE, 1999.

[12] S. Teboul, L. Blanc-Feraud, G. Aubert, and M. Barlaud. Variational approach for edge-preserving regularization using coupled pde's. *IEEE Trans. Im.Proc.*, 7:387–397, 1998.

[13] C.R. Vogel and M.E.Oman. Iterative methods for total variation denoising. *SIAM Jl. Sci. Stat.Comp.*, 1996.

[14] J. Weickert, B.M. ter Haar Romeny, and M.A. Viergever. Efficient and reliable schemes for nonlinear diffusion filtering. *IEEE Trans. Im.Proc.*, 7:398–410, 1998.

[15] Y.L. You, W.Xu, A.Tannenbaum, and M.Kaveh. Behavioural analysis of anisotropic diffusion in image processing. *IEEE Trans. Im.Proc.*, 5:1539–1553, 1996.