SELECTION OF REGULARISATION PARAMETERS FOR TOTAL VARIATION DENOISING

V Solo

Dept. of Statistics, Macquarie Univ.
Sydney NSW 2109, Australia
email: vsolo@zen.efs.mq.edu.au

ABSTRACT

We apply a general procedure of the author to choose penalty parameters in total variation denoising.

1. INTRODUCTION

A number of methods have been presented in recent years for the estimation of discontinuous functions in ill-conditioned inverse problems. Thus there is the method of graduated nonconvexity [1], methods based on Markov Random fields [5], Lp regularisation [2] and wavelets [4]: references to other methods can be found in [12] and [7].

More recently the method of total variation denoising has been developed successfully [13], [9] also going under the name anisotropic diffusion [14],[8]. An important feature of this method is that the Tikhonov performance index is nonlinear in the reconstructed function. In any case in all these methods there is the problem of choosing a Tikhonov regularisation or bandwidth parameter, which is the topic of this paper.

There are two well known general approaches capable of delivering bandwidth parameter estimates in non-linear ill-conditioned inverse problems: maximum likelihood and cross validation. Maximum likelihood requires a stochastic model for the function while cross validation treats the function deterministically.

But both methods suffer from a problem of computational complexity. However recently the author [10] has observed that another deterministically based method (unbiased risk estimation) can be extended exactly to nonlinear cases. The technique is computationally much simpler than maximum likelihood or cross validation (but similar to the latter) and beyond solution of the Euler equations associated with the minimisation of the Tikhonov criterion requires only a computation of a trace of a matrix inverse.

In this paper we apply the method to total variation denoising [9], [13]. This seems to be the first development of an automatic method for choosing regularising parameters for that technique.

2. TOTAL VARIATION DENOISING

For simplicity we describe the technique in one dimension. Consider the problem of estimating the possibly discontinuous function f(t) on [0,1] given noisy data

$$y_i = f(i/n) + \epsilon_i \ \ , \ \ i = 1, ..., n$$

where ϵ_i is a white noise of variance σ^2 . The total variation denoising (TVDN) method chooses f to solve the Tikhonov problem

$$\begin{aligned} \hat{f}_\alpha &= arg.min J(f) \\ J(f) &= \frac{1}{2n}\Sigma_1^n (y_i - f_i)^2 + \alpha \int_0^1 |f^{'}| dx \end{aligned}$$

However, the continuous functional in J(f) is not differentiable and this causes difficulties with optimisation. In [9] a steepest descent procedure is used giving a nonlinear diffusion equation. More recently [13] have used a standard mollifier method to perturb the nonsmooth optimisation problem to a smooth one. Thus J(f) is replaced by

$$J(f) = \frac{1}{2n}\Sigma_1^n (y_i - f_i)^2 + \alpha \int_0^1 (|f^{'}|^2 + \gamma^2)^{\frac{1}{2}} dx \quad (2.1)$$

where γ is the small mollifier parameter.

It is worth noting here several important properties that this mollified regularisation function or potential $\rho(x) = (x^2 + \gamma^2)^{\frac{1}{2}} - \gamma$ which have not hitherto been noted. Firstly it obeys the four properties prescribed by [12]- convexity; symmetry; allows discontinuity (ie. $\rho(x) < x^2$ for large x ; for fixed x is monotonic in γ). Of course there is also positivity and uniqueness (ie. $\rho(x) = 0 \Rightarrow x = 0$).

Secondly it obeys all the properties listed by [7] but additionally has the very important feature of being

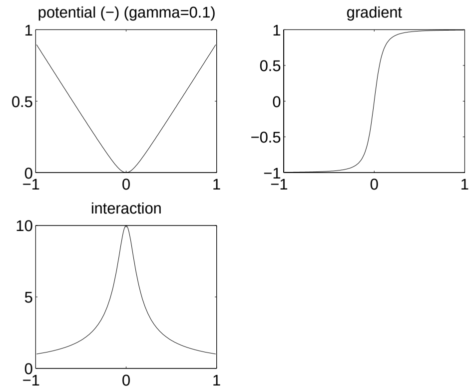

Figure 1: Plot of Potential function characteristics

convex on $(-\infty, \infty)$. The only other known potential obeying this is g_4 of [7]. This global convexity is crucial for the convergence properties mentioned below but its significance seems to have been missed by [7]. Also the anisotropic diffusion algorithm is well posed in the sense of [14] although it does not enjoy the convergence properties described below. In Fig.1 we have plotted an example of the potential function and its characteristics for $\gamma = .1$.

To calculate $\hat{f}$ [13] develop an iterative procedure they call lagged diffusion. We discretise (2.1) and develop a sort of pseudo Gauss Newton algorithm which is much the same as that of [13]. The resulting iteration is

$$(I + n\alpha D^T A^{-1}_{(k-1)} D) f_{(k)} = y \qquad (2.2)$$

where D is an (n-1)xn differencing matrix; $f_{(k)}$ is the kth iterate of $f = (f_1 ... f_n)^T$ and $A_{(k)} = diag(a_2^{(k)} ... a_n^{(k)})$ and $a_r = [(f_r - f_{r-1})^2 n^2 + \gamma^2]^{\frac{1}{2}}$. The matrix $D^T A^{-1} D$ has tridiagonal structure so the solution of (2.2) is rapid. Also it is possible, using arguments similar to those in [3] to show global convergence of (2.2).

3. ESTIMATED RISK

We introduce a reconstruction performance measure (the risk or discrepancy)

$$\begin{aligned} R_\alpha &= E||f - \hat{f}_\alpha||^2 \\ &= E\int (f - \hat{f}_\alpha)^2 dx \end{aligned}$$

Ideally we would like to minimise R_α with respect to α . However, R_α is unknown and so not computable and the idea is to find a computable unbiased estimator of R_α and minimise that instead. Remarkably such a statistic can be found. If the ϵ_i are Gaussian then it can be shown [11], [10] that an unbiased estimator of R_α is

$$\begin{aligned} \hat{R}_\alpha &= n\sigma^2 + \Sigma_1^n e_i^2 - 2\sigma^2 \Sigma_1^n \partial e_i / \partial y_i \\ e_i &= y_i - \hat{f}_i = y_i - \hat{f}_{\alpha,i} \end{aligned}$$

The general use of this estimator for ill conditioned inverse problems was suggested in [10] although it has been used in two special cases before.

For the current setting of total variation denoising it can be shown that

$$\begin{aligned} \Sigma_1^n \frac{\partial e_i}{\partial y_i} &= ntrace(I + n\alpha\gamma^2 D^T K^{-1} D)^{-1} \\ K &= diag(a_r^3) \end{aligned}$$

4. RESULTS

In Fig.2 is a plot of SURE for the blocky function used by [13]. We chose $\gamma = .0001, snr = 4$ to compare with their results. Here signal to noise ratio (snr) is a power ratio but is not reexpressed in dB. There is a well defined minimum but SURE is otherwise rather flat near that minimum; this means that several values of α in that vicinity should be tried. In Fig.3 we show the reconstruction corresponding to the minimising $\alpha = .005$. In Fig.4 we show the reconstruction corresponding to the value $\alpha = .01$ used by [13]. It was not commented upon by [13] but the latter estimate shows large biases at the peak and trough associated with the narrow block. On the other hand while the estimate corresponding to the minimser of SURE is a bit less block like it has much smaller bias at those peaks and troughs. The method used by [13] to choose the penalty parameter is known [6] to lead to oversmoothing in Sobolev regularisation and we are seeing that same behaviour here.

5. SUMMARY

In this paper we have presented an automatic method of tuning parameter choice for total variation denoising. The method is computationally much simpler than cross-validation. 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.

6. REFERENCES

[1] A. Blake and A.Zisserman. Visual Reconstruction. MIT Press, Boston, MA, 1987.

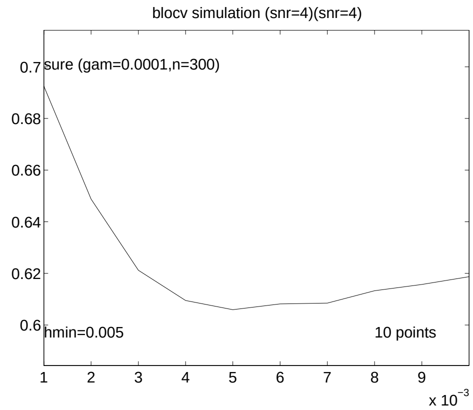

Figure 2: Plot of SURE for blocky function

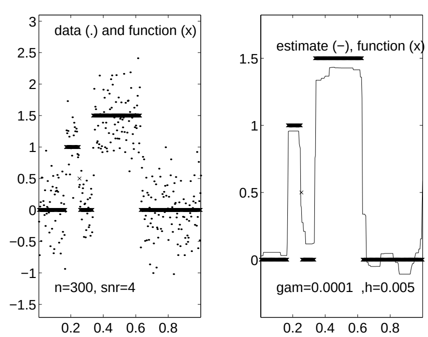

Figure 3: Plot of data and estimate, $\alpha = .005$

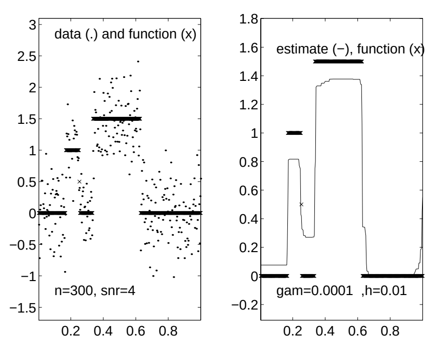

Figure 4: Plot of data and estimate, $\alpha = .01$

[2] C. Bouman and K. Sauer. A generalised Gaussian image model for edge preserving map estimation. IEEE Trans. Im.Proc., 2:296–310, 1993.

[3] D.C. Dobson and C.R.Vogel. Convergence of an iterative method for total variation denoising. SIAM Jl.Num.Anal., to appear, 1996.

[4] D.L. Donoho and I.M. Johnstone. Adapting to unknown smoothness via wavelet shrinkage. Jl. Amer. Stat. Assoc, 90:1200–1224, 1995.

[5] D. Geman and G. Reynolds. Constrained restoration and the recovery of discontinuities. IEEE. Trans. Patt. Anal. Machine Intell., 14:367–383, 1992.

[6] P. Hall and D.M. Titterington. Common strtucture of techniques for choosing smoothing parameters in regression problems. Jl. Royal. Stat. Soc. Ser. B, 49:p.184–198, 1987.

[7] S.Z. Li. On discontinuity-adaptive smoothness priors in computervision. IEEE. Trans. Patt. Anal. Machine Intell., 17:576–586, 1995.

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

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

[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] C. Stein. Estimation of the mean of a multivariate normal distribution. Ann. Stat., 9:1135–1151, 1981.

[12] R.L. Stevenson, B.E. Schmitz, and E.J. Delp. Discontinuity preserving regularization of inverse visual problems. IEEE Trans Sys Man Cybernetics, 24:455–469, 1994.

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

[14] 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.