An innovative procedure for smoothing parameter selection

Gianluca Frasso, Paul H.C. Eilers.

Abstract Smoothing with penalized splines calls for an automatic method to select the size of the penalty parameter λ. We propose a not well known smoothing parameter selection procedure: the L-curve method. AIC and (generalized) cross validation represent the most common choices in this kind of problems even if they indicate light smoothing when the data represent a smooth trend plus correlated noise. In those cases the L-curve is a computationally efficient alternative and robust alternative.

Key words: Whittaker and P-spline smoothers; L-curve; Cross validation.

1 Introduction

Penalized regression has a prominent place in modern smoothing. It combines a rich set of basis functions with a roughness penalty, to tune smoothness of the estimated curve. P-splines combine a B-spline matrix with a penalty on (higher order) differences of their coefficients. A smoother that follows this framework is the Whittaker's smoother, also known as Hodrick-Prescott filter [18]. It is obtained considering equally spaced knots and the identity matrix as regression basis. It is an attractive smoother, because effectively the basis functions disappear and with just one smoothing parameter one can move all the way from a straight line fit to essentially reproducing the data themselves.

Every smoothing procedure call for an automatic smoothing selection procedure. The most common ones are leave-one-out cross-validation (LOO-CV) and AIC (Akaike's Information Criterion) or BIC (Bayesian Information Criterion). They all

Department of Mathematics and Statistics, University of Naples Federico II,
e-mail: gianluca.frasso@unina.it
Department of Biostatistics, Erasmus Medical Centre, Erasmus University in Rotterdam, The Netherlands, e-mail: p.eilers@erasmusmc.nl

share two drawbacks: 1) they require the computation of the effective model dimension, and 2) they are sensitive to serial correlation in the noise around the trend.

We present and alternative approach, based on the L-curve method for ridge regression [10]. This curve is a plot of the logarithm of the magnitude of the penalty term against the log of the sums of squares of the residuals, parameterized by the regularization parameter λ. There is no need to compute the effective model dimension, so using the L-curve makes smoothing of long data series practical. Furthermore this criterion is robust to by correlated noise.

This paper is organized as follows: in Section 2 we show how the L-curve can be built and how to select the optimal λ parameter using it while in Section 2.1 we present a well known smoothing example considering the wood dataset proposed by Pandit and Wu [14].

2 Building the L-curve

Let consider a P-spline smoother obtained solving the minimization problem:

$$\widehat{z} = \underset{z}{\operatorname{argmin}} \|y - Bz\|^2 + \lambda \|Dz\|^2 \tag{1}$$

where B represents the B-spline matrix and the parameter λ tunes the smoothness of the final result. The following quantities can be defined:

$$\{\omega(\lambda); \theta(\lambda)\} = \{\|y - Bz\|^2; \|Dz\|^2\}$$

The L-curve is the parameterized curve:

$$L = \{\psi(\lambda); \phi(\lambda)\} = \{\log(\omega); \log(\theta)\} \tag{2}$$

The curve in (2) has a L-shape. Selecting λ as the parameter corresponding to the corner of the L (the point of maximum curvature) gives the optimal degree of smoothing. The pointwise curvature of (2) can be computed by:

$$k(\lambda) = \frac{\psi' \phi'' - \psi'' \phi'}{[(\psi')^2 + (\phi')^2]^{3/2}} \tag{3}$$

The optimal smoothing parameter is then selected maximizing $k(\lambda)$. However a simpler procedure can be used. Indeed for well-behaved L-curves (i.e. in those cases in which it is possible to distinguish a clear corner point) it is possible to approximate the selection procedure minimizing the Euclidean distances between adjacent points on the curve:

$$\min\{\sqrt{(\Delta \psi)^2 + (\Delta \phi)^2}\} \tag{4}$$

Figure 1 shows the performance of the L-curve in smoothing simulated data and compares the results obtained using the two selection strategies mentioned above.

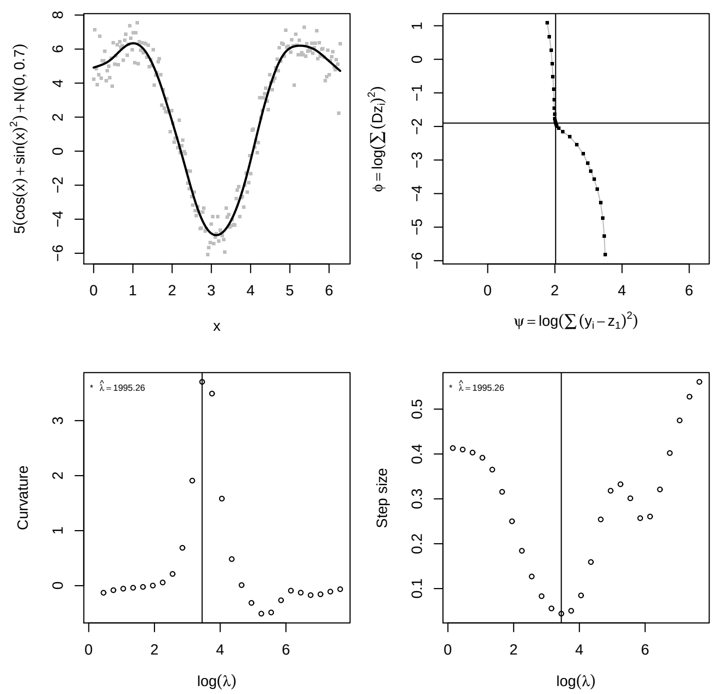

Fig. 1 P-spline smoothing of simulated data using a L-curve approach. The first panel shows the obtained smoothing function (line) and the second the associated L-curve. The lower panels show the curvature function and Euclidean distance between adjacent points of L-curve. The values of these functions are plotted against different values of $\log(\lambda)$.

2.1 The shape of the L-curve

Using simulated data we can evaluate how the characteristics of the data impact on the shape of the L-curve. This issue is summarized in figure 2. The second panel shows the L-curve for a Whittaker smoother applied to data simulated using the following scheme: $y = 10^{c_j}\sin(x_i) + N(0,1)$ with $x_i = 1,...,2\pi$ for $i = 1,...,200$ and $c_j = 2.5,...,-2$ for $j = 1,...,7$. It is clear that the convex region tends to disappear when the white noise component tends to be dominant on the trend component.

Also the characteristics of the error component influence the shape of the L-curve. First of all the the variability of the white noise component plays a role. The third panel of figure 2 shows that higher the variability is, less sharp the L-curve appears. These results where obtained using 200 observation simulated as follows: $y = \sin(x) + N(0,\sigma_j)$ with $x = 1,...,2\pi$ and $\sigma_j \in \{0.01, 0.02, 0.05, 0.1, 0.2, 0.5, 1\}$.

The L-curve is particularly useful if we want to smooth data with autocorrelated noise. However the shape of the curve depends on the strength of the serial correlation of the error component. The fourth panel of figure 2 shows some results obtained using the Whittaker smoother on a set of data simulated as follows: $y = 3\sin(x_i) + AR(1, \rho_j, \sigma = 1)$ with $x_i = 1, ..., 2\pi$ for $i = 1, ..., 200$ and $\rho_j = 0, ..., 0.9$ for $j = 1, ..., 7$ where ρ indicates the autocorrelation coefficient. Poorly autocorrelated noise produces really sharp L-curves while higher degrees of serial correlation reduce the sharpness. In any case the curves show clear convex areas.

In addition to these considerations we found also that the L-curve seems to be less sharp in smoothing spline regression than in ridge regression analysis (first panel of Figure 2). In our opinion it does not invalidate the applicability of the methodology. Indeed we believe that, as long as a convex area is well distinguishable, the procedure can be considered reliable.

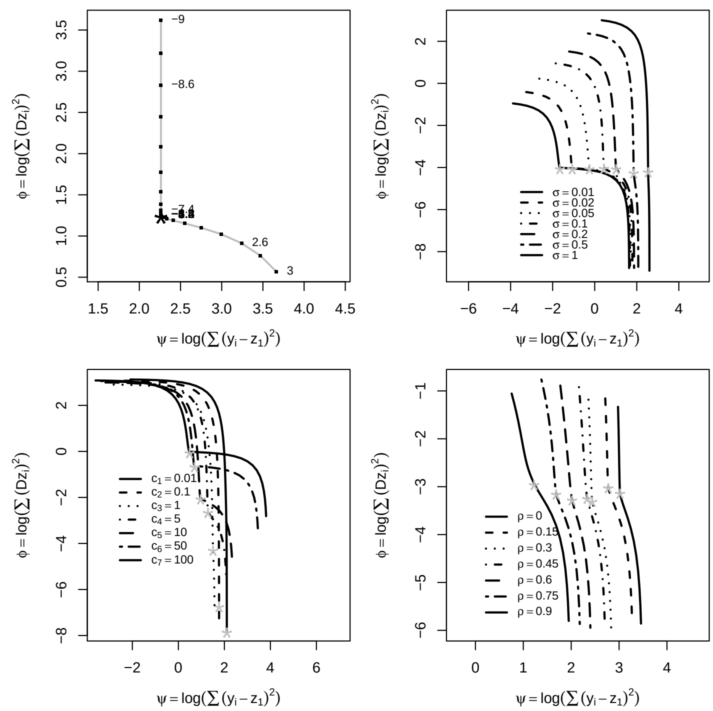

Fig. 2 The first panel shows the L-curve obtained for a simulated ridge regression example. The second panel shows seven L-curves obtained for a Whittaker smoother estimated on data with different variability for the white noise component. The third panel shows seven L-curves obtained for a Whittaker smoother estimated on data with different weights for the signal component. The fourth panel shows seven L-curves for a Whittaker smoother estimated on data with an increasing autocorrelation coefficient for the noise component.

3 A real data example

An interesting application that we would like to show concerns the smoothing of the wood data. This dataset was proposed by Pandit and Wu [14]. It describes 320 measurements of a block of wood that was subject to grinding. In figure 3 the profile height at different distances is drawn. The profile variation follows a curve determined by the radius of the grinding stone. We can use the Whittaker smoother to analyze these data and compare the performances of the L-curve and the cross validation for the smoothing parameter selection. The fitted curves and the related selection criteria are shown in figure 3. Also in this case the smoothing procedure built using the L-curve efficiently reproduces the trend in the data while the filter based on cross validation does not.

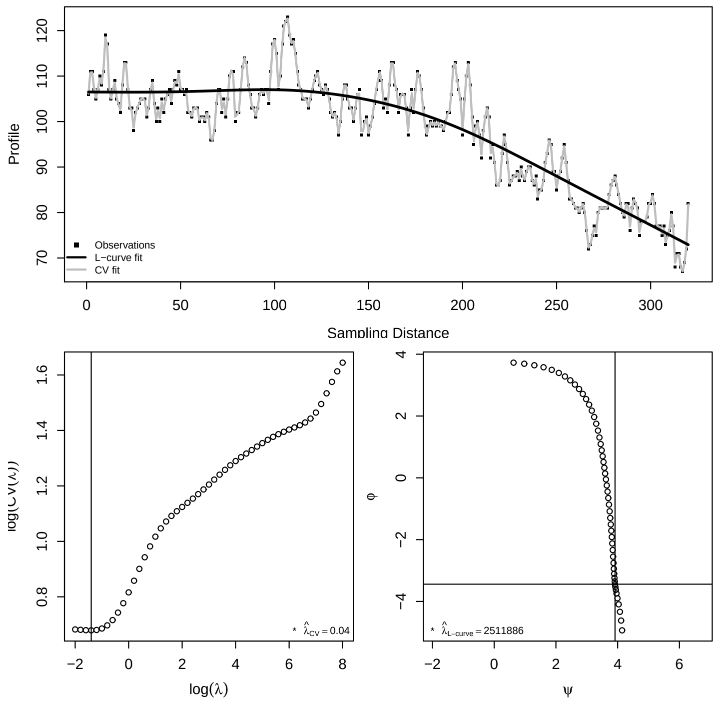

Fig. 3 Profile of a block of wood subject to grinding. The upper panel shows the result obtained selecting the λ parameter of a Whittaker smoother using the cross validation and the L-curve methods. The lower panels show the cross validation and L-curve profiles and indicate the selected smoothing parameters. For both selection methods we considered $\log(\lambda) \in [-4,6]$.

Figure 3 shows clearly that the CV tends to suggest a too small λ parameter while the L-curve procedure guarantees a parameter large enough to catch the signal behind the data.

4 Discussions

In this work a L-curve procedure for the selection of the smoothing parameter in a P-splines framework is presented. The L-curve selects the optimal smoothing parameter comparing the goodness of fit and the roughness of the final estimates. The optimal smoothing parameter is selected maximizing the local curvature on the 'L', i.e. locating its corner. This method shows excellent performances in practice. We have no compelling explanations of why the L-curve works so well. Of course, the corner of an L-shaped curve is a special point, but it is not clear why it marks a good choice of smoothing parameter. The relative changes of both the penalty and the size of the residuals are small there, and approximately equal, and apparently that matters. The insensitivity to serial correlation in the noise is also hard to explain. The L-curve criterion could lead to no reliable results in some cases. It usually happens smoothing data approaching to a pure white noise or when the signal component underlying the data tends to disappear. Indeed in those cases the L-curve tends to be globally concave.

The L-curve procedure offers many opportunities for further research. We believe that it is possible to generalize this methodology. Our future research will concentrate on a L-curve criterion for multivariate smoothing analyzes and on a L-curve based procedure suitable for spatially adaptive smoothing problems. We will also study the applicability of this procedure to a generalized linear model setting for smoothing of counts and binary data.

References

1. I. D. Currie and M. Durban. Flexible smoothing with P-splines: a unified approach. *Statistical Modelling*, 4:pp. 333–349, 2002.
2. C. de Boor. *A Practical Guide to Splines*. Springer-Verlag, 1978.
3. Paul Dierckx. *Curve and Surface Fitting with Splines*. Oxford University Press, 1995.
4. Paul H. Eilers. A perfect smoother. *Analytical chemistry*, 75(14):3631–3636, July 2003.
5. Paul H. C. Eilers and Brian D. Marx. Flexible smoothing with B-splines and penalties. *Statistical Science*, 11(2):pp. 115–121, 1996.
6. Paul H. C. Eilers and Brian D. Marx. Splines, knots, and penalties. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(6):637–653, 2010.
7. G. Frasso and P. H. Eilers. Smoothing parameter selection using the L-curve. *Technical report, Erasmus Medical Center, Erasmus Universiteit, Rotterdam, The Netherlands*, Submitted, 2012.
8. P. C. Hansen. The L-curve and its use in the numerical treatment of inverse problems. In *Computational Inverse Problems in Electrocardiology, ed. P. Johnston, Advances in Computational Bioengineering*, pages 119–142. WIT Press, 2000.
9. Per C. Hansen. Analysis of Discrete Ill-Posed Problems by Means of the L-curve. *SIAM Review*, 34(4):pp. 561–580, 1992.
10. Per C. Hansen and Dianne P. O'Leary. The use of the L-Curve in the regularization of discrete ill-posed problems. *SIAM J. SCI. COMPUT.*, 14(6):pp. 1487–1503, 1993.
11. T. J. Hastie and R. J. Tibshirani. *Generalized additive models*. London: Chapman & Hall, 1990.

12. Robert J. Hodrick and Edward C. Prescott. Postwar U.S. Business Cycles: An Empirical Investigation. *Journal of Money, Credit and Banking*, 29(1):pp. 1–16, 1997.
13. Goeran Kauermann, Tatyana Krivobokova, and Willi Semmler. Filtering time series with penalized splines. *Studies in Nonlinear Dynamics & Econometrics*, 15(2):2, 2011.
14. S. M. Pandit and S. M. Wu. *Time series and system analysis with applications*. Krieger, 1993.
15. A Pressley. *Elementary Differential Geometry*. Springer-Verlag, 2001.
16. T. Reginska. A regularization parameter in discrete ill-posed problems. *SIAM J. Sci. Comput.*, 17:740–749, May 1996.
17. G. Wahba. *Spline models for observational data*, volume 59 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1990.
18. E. T. Whittaker. On a new method of graduation. *Proceedings of the Edinburgh Mathematical Society*, 41:63–75, 1922.