AUTOMATIC SMOOTHING BY OPTIMAL SPLINES
Ricardo Biloti,1 Lúcio T. Santos,2 Martin Tygel3
Submitted August 1, 2003/Accepted March 31, 2004

ABSTRACT

We propose a method that is capable to filter out noise as well as suppress outliers of sampled real functions under fairly general conditions. From an a priori selection of the number of knots that define the adjusting spline, but not their location in that curve, the method automatically determines the adjusting cubic spline in a least-squares optimal sense. The method is fast and easily allows for selection of various possible number of knots, adding a desirable flexibility to the procedure. As an illustration, we apply the method to some typical situations found in geophysical problems.

Keywords: Smoothing, cubic splines

RESUMO

Propomos um método capaz de filtrar ruído bem como suprimir outliers de uma função real amostrada, impondo praticamente nenhuma condição. A partir de uma seleção a priori da quantidade de nós que definem a spline, porém não suas localizações, o método automaticamente determina a spline cúbica que melhor ajusta os dados, no sentido de quadrados mínimos. Dado que o método é rápido, torna-se viável a realização de vários ajustes com quantidades distintas de nós, conferindo assim a flexibilidade desejada ao processo. Como exemplo, aplicamos o método a alguns problemas em Geofísica.

Palavras-chave: suavização, spline cúbica

1 Universidade Federal do Paraná - Departamento de Matemática Centro Politécnico, Jd. das Américas - Cx.P. 19081 - CEP 81531-980, Curitiba, PR - Fone/fax: (41) 361 3400 – e-mail: biloti@mat.ufpr.br
2 Departamento de Matemática Aplicada - IMECC - UNICAMP - Caixa Postal 6065 - 13081-970 Campinas (SP) - Fax. (19) 3289-5766 - Tel. (19) 3788-5975 – e-mail: lucio@ime.unicamp.br
3 UNICAMP – IMECC - Praça Sérgio Buarque de Holanda, 651 – Cidade Universitária - Barão Geraldo - Caixa Postal: 6065 - Campinas - São Paulo - 13083-859 – e-mail: tygel@unicamp.br
AUTOMATIC SMOOTHING BY OPTIMAL SPLINES

INTRODUCTION

In experimental sciences we are often required to represent a set of measured data in the form of a smooth curve from which desirable parameters or attributes are to be extracted. A common problem is the presence of noise in the data. In the literature we find several methods that try to filter and/or smooth the data. Many of them provide us, after application, a different sampled dataset that, following some criteria, can be seen as smoother than the original one. We can also think of an interpolation approach, where the noisy data is replaced by corresponding points that belong to an interpolating function.

The natural question is how should we choose the “interpolating points” from the data so as to construct the desired smoothing function. Normally, these points are extracted from the data, in a regular fashion or manually selected.

We propose a method that optimally selects points to define a cubic spline that best represents the data in the least-squares sense. An interesting feature of the method is that the knot points are no longer required to belong to the original data set. As we will see below, the method is suitable, not only for smoothing, but also for discarding outliers. The method is designed to handle datasets composed by samples of rather complicated real functions. It is to be stressed that, by construction, the obtained function is naturally smooth up to second-order derivative.

The proposed method is applied to two important problems in geophysics. The first problem is to recover horizons as part of a macro-velocity model inversion from multi-coverage seismic data and to smooth seismic traveltime attributes (BILOTI; SANTOS, 2002). The second application refers to smoothing well-log data for anomaly detection and inversion purposes.

FORMULATION

Consider a noisy data 
\[ \Omega = \{(x_j, y_j) \in \mathbb{R}^2 | j = 1, ..., M\} \]
be the number of interpolating points and \( \Gamma = \{(X_k, Y_k) \in \mathbb{R}^2 | X_{i-1} < X_i, i = 1, ..., N\} \) be the set of these points that defines the sought-for cubic spline. To obtain the best set \( \Gamma \), in the least-squares sense, we must solve the \( 2N \)-variable problem

\[
\min_{\Gamma} \sum_{j=1}^{M} y_j - s(x_j)^2, \quad \text{s.t.} \quad \begin{cases} 
 s(x_j) \geq \min_{k} x_{j} \\
 s(x_{N}) \leq \max_{j} x_{j}
\end{cases} \quad (1)
\]

To solve this problem, we have employed the optimization solver called GENCAN proposed by Birgin and Martinez (2002). GENCAN is an active-set method for smooth box-constrained minimization. The algorithm combines an unconstrained method, including a line search which aims to add many constraints to the working set at a single iteration, with a recently introduced technique (spectral projected gradient) for dropping constraints from the working set. As usual, the optimization process needs an initial approximation. For this purpose, we chose the initial set as composed by \( M \) regularly sampled pairs on the originally given set \( \Omega \), that is, \((x_j, y_j) = (x_{ji}, y_{ji})\), with \( j = \lfloor 1 + (i-1) \cdot (M-1)/(N-1) \rfloor \) \( \lfloor \cdot \rfloor \) denotes the greater integer less than or equal to \( x \).

Note that we have not made any consideration on how to choose the number, \( N \), of interpolating points. The method is designed to automatically find, in the least-squares sense, the best cubic spline for the specified number of knots \( N \). Of course, for a small number of points \( N \), the obtained spline will not be able to represent more than the general trend of the curve. On the other extreme, for large values of \( N \), the spline will tend to fit even the outliers. Since the method is fast, it is reasonable to estimate the cubic spline for several choices of \( N \). This flexibility can be very useful to the user or interpreter, in the sense that a number of inexpensive trials can be implemented before a final decision on which level of smoothness is the best choice for the problem.

APPLICATIONS

We now illustrate the application of the proposed method to some common practical situations. We start by testing the ability of the method to smooth a sequence of the four datasets of increasing difficulty, shown in Figure 1. We next apply the procedure to two problems related to seismic imaging and inversion, shown in Figures 2 and 3, respectively.

General Situations

Figure 1(a) shows that the method efficiently handles and removes white noise. In the next example (Figure 1(b)), besides the noise, some outliers were added. Again, the optimized cubic spline represents the data very well. In Figure 1(c), we see that the obtained curve is able to well describe abrupt variations of the data. Finally, in Figure 1(d), we see that the method is robust enough to provide good results even in the presence of discontinuities. Note that, in particular, there are no Runge effects near the discontinuities.
HORIZON RECONSTRUCTION

We present the results of an algorithm for reconstruction of interfaces and inversion of attributes from 2D-multi-coverage seismic data. As reported in Biloti et al. (2001), the procedure has been successfully applied to invert a layered macro-velocity model from the data. Figure 2(a) depicts the inverted model, where the estimated interfaces (solid lines) were approximated by conventional cubic splines (constructed by selection of knots among the sampled set). In Figure 2(b), we can see the improvement of the inverted model, when the interfaces were constructed upon the application of the proposed technique.

WELL-LOG ANALYSIS

Well logs play the important role of linking rock parameters to seismic data. As an example, impedance functions derived from well logs are generally used for identification and characterization of reservoir anomalies. Well data (e.g., P- and S-velocities and density) are, in general, very noisy, so it may be desirable to consider the parameters as smooth functions of depth (or time). Figure 3 shows the application of the method to smooth a couple real data well logs. On the top of the figure, the new method was used with $N = 20$ knots. On the bottom of the figure, the number of knots was $N = 30$.

CONCLUSIONS

We presented an automatic method for smoothing and outlier suppression of data sets that consist of sampled real function points. After an a priori selection of the number of knots, the procedure automatically finds the location of the interpolating points, in such a way that the resulting smoothing function (a cubic spline) is optimal in the least-squares sense. As the method is fast, it allows the user to apply the procedure to different numbers of knots, so as to choose the degree of smoothness that best fits the data. A particular feature of the method is its ability to adjust to abrupt discontinuities on the data.
The few illustrations presented in the text show a wide applicability of the method. As further applications, the method could be useful for purposes such as tomographic inversion and automatic picking.

Acknowledgements

This work was kindly supported by CNPq & FAPESP (Brazil) and the sponsors of the Wave Inversion Technology (WIT) Consortium, Karlsruhe, Germany.

REFERENCES


Figure 3 – P- and S-velocities and density well logs, in black, and adjusted cubic spline, in red, with (a) $N=20$ and (b) $N=30$.

Figura 3 – Velocidades de onda P e S e densidades em poço, em preto, e splines cúbicas ajustadas, em vermelho, com (a) $N = 20$ e (b) $N = 30$. 
NOTES ABOUT THE AUTHORS

Ricardo Biloti is graduated in Applied Mathematics (UNICAMP/1995). Master in Applied Mathematics (UNICAMP/1998). PhD in Applied Mathematics (UNICAMP/2001). In 2001, he spent two months in visit to the Institute of Geophysics of the University of Karlsruhe, Germany. Since 2002, he is Assistant Professor at the Department of Mathematics at the Federal University of Paraná, working in the groups of Scientific Computing, Optimization. His areas of activity concentrate in Numerical Analysis and applications to Geophysics. His professional interests include multiparametric seismic processing and estimation of models of velocity. He is member of SEG, SBMAC and SIAM.

Lúcio T. Santos is graduated in Applied Mathematics (UNICAMP/1982), Master in Applied Mathematics (UNICAMP/1985), PhD in Electrical Engineering (UNICAMP/1991) and Distinguished Professor (UNICAMP/1999). From 1985 to 1988 he was Assistant Professor at USP/São Carlos and since 1998 he is Professor at the Department of Applied Mathematics at IMECC/UNICAMP, working in the groups of Computational Geophysics and Optimization. From August/1994 to September/1995 he did a Post-Doctoral at Rice University (EUA) and during January/February of 1998, 1999 and 2001 he was Visiting Professor at the Institute of Geophysics of the University of Karlsruhe (Germany). His professional interests include seismic modeling and imaging and optimization of non-linear. His recent research involves new approximations for P-P reflection coefficient, alternative attributes for analysis of AVO, multiparametric expressions for travel times and finite difference methods for the local wave equation. He is member of SBGf and SEG.

Martin Tygel is graduated in Physics (UERJ/1969), Master in Mathematics (PUC-Rio/1973) and PhD in Mathematics (Stanford University/1979). He is Professor at the UNICAMP since 1984, having been a teacher at UFRN (Natal/1979-1980), PPPG-UFBA (Salvador/1981-1983) and Institute of Geophysics of the University of Karlsruhe (Germany) in 1990. He is the recipient of Schlumberger (EAGE/2002) and Zeferino Vaz (UNICAMP/1997 and 2003) prizes. He is a founder of the Laboratory of Computational Geophysics at UNICAMP. The Laboratory is one of the members of the Consórcio Wave Inversion Technology (WIT), with headquarters in Karlsruhe. He is a consultant for Petrobras since 1991 on processing and imaging seismic data. He is Associate Editor of the Geophysics (SEG).