International Symposium on Nonlinear Theory and its Applications
Xi'an, PRC, October 7–11, 2002

Nonparametric nonlinear system modeling

M. Peifer, J. Timmer, H.U. Voss

Freiburg Center for Data Analysis and Modeling
Eckerstrasse 1, 79104 Freiburg, Germany
Tel.: +49-761-203-5886, Fax.: +49-761-203-5967
Email: hv@physik.uni-freiburg.de

Summary

We propose a nonparametric method for nonlinear dynamical system identification from time series. Having only few structural information, a nonparametric approach may be more appropriate than a parametric one for which detailed prior knowledge is needed. Our method is based on nonparametric regression which is used to estimate functions in additive models with a penalized least-squares approach. The optimal penalty term or smoothing parameter is determined via generalized cross validation. The procedure is applied to identify the nonlinear restoring force of helical wire rope isolators.

Keywords System identification, nonparametric modeling, helical wire rope isolators

1. Introduction

We propose a method for nonlinear dynamical system identification from time series for the case that only few structural information about the system is known. Instead of a fit of parameters in given model equations to the data, in this nonparametric approach the functions involved in the equations are estimated themselves. The basic requirement is that the equation can be written as an additive model. Additive models have been successfully applied to various data modelling problems and can be regarded as a generalization of linear regression to nonlinear dependencies [1, 2, 3]. They are given by

$$y = f_1(x_1) + \ldots + f_p(x_p) + \epsilon \; , \tag{1}$$

where the observational error ϵ is usually assumed to be normally distributed and independent of $x_1, \ldots, x_p$. The unknown functions $f_1(x_1), \ldots, f_p(x_p)$ have to be estimated from observations of y given at some design points $(x_1, \ldots, x_p)$. In the following, we will present a method to estimate these functions efficiently and data-adaptive. The basic idea of this procedure is to utilize the additivity in order to built an iterative backfitting algorithm: The additive model is first decomposed into several one-dimensional problems which are solved in each iteration step by smoothing. The reduction to one variable during the iteration process does not suffer from the so-called curse of dimensionality, which is the effect that high dimensional spaces are filled sparsely [4].

The smoothing parameters of each function can be efficiently selected in additive models. The optimal smoothing parameter is balancing the bias to the variance of the estimated curve. It can be estimated by generalized cross validation [5].

2. Estimation of additive models

Consider the simple one-dimensional model

$$y = f(x) + \epsilon \; , \tag{2}$$

where ϵ is a normally distributed error which is independent of x, and the function $f(x)$ is assumed to be unknown. To estimate $f(x)$ from observed data y, one has to smooth the data points to remove the wiggliness which is induced by the error. There are numerous methods of smoothing which are classified by linear and nonlinear smoothers. The smoothers we choose are always linear which is motivated by the fact that confidence bounds can be calculated and there are certain conditions in which backfitting converges, in contrast to nonlinear smoothers, where there are no general convergence conditions. Linear smoothers are mapping the data points linearly to the corresponding estimated function values, or more formally: let $\mathbf{y} = (y_1, \ldots, y_N)^t$ be the data at the so-called design points $\mathbf{x} = (x_1, \ldots, x_N)^t$ ($(\cdot)^t$ being the transposition), then a function estimate $\hat{\mathbf{f}} = \left(\hat{f}_1, \ldots, \hat{f}_N\right)^t$ at $\mathbf{x}$ is $\hat{\mathbf{f}} = S\mathbf{y}$, where S is the $N \times N$ hat-matrix. A smoother is said to be nonlinear if a hat-matrix S does not exist. A computationally comfortable linear method is provided by *smoothing splines*. Here an additional term is added to the usual least-squares functional to penalize the roughness of the estimate:

$$\mathcal{L}(g) = \underbrace{\sum_{i=1}^{N} \frac{(y_i - g(x_i))^2}{\sigma^2}}_{least-squares} + \underbrace{\alpha \int \left[\frac{d^2 g(x)}{dx^2}\right]^2 dx}_{penalty} \quad . \tag{3}$$

The smoothing parameter $\alpha \geq 0$ determines the amount of smoothing, ranging from zero (interpolation) to infinity (linear regression). The minimization of Eq. (3) leads to natural cubic smoothing splines [7, 3]. An overview of some other linear smoothers and their comparison can be found in [1, 2].

Suppose the function in Eq. (2) is known and let $\hat{f}_\alpha$ be an estimate of f, depending on the smoothing parameter α. Then, the mean-squared error

$$\mathrm{MSE}(\alpha) = N^{-1} \sum_{i=1}^{N} \left(\hat{f}_\alpha(x_i) - f(x_i)\right)^2 \tag{4}$$

is an appropriate measure to quantify the goodness of the estimation. Thus, minimizing $\mathrm{MSE}(\alpha)$ gives the optimal smoothing parameter, but the definition of the mean-squared error still contains the unknown function f. The idea to save this concept is to estimate the MSE first, and finally a minimization of the estimated MSE score with respect to α yields an estimate of the optimal smoothing parameter. This can be done by *generalized cross validation*; as a result one obtains the generalized cross validation score [5, 3]

$$\mathrm{GCV}(\alpha) \quad = \quad \frac{N^{-1} \sum_{i=1}^{N} \left(y_i - \hat{f}_\alpha(x_i)\right)^2}{\left(1 - N^{-1} \mathrm{tr}\, S(\alpha)\right)^2} \quad . \tag{5}$$

In this expression $S(\alpha)$ is the hat matrix for smoothing splines and $\mathrm{tr}\, S(\alpha)$ its trace. For a more detailed treatment of an approximative and an exact calculation of $\mathrm{tr}\, S$, see [9, 3].

An extension of model (2) to a p-dimensional model are additive models of the form

$$y = f_1(x_1) + \ldots + f_p(x_p) + \epsilon \; . \tag{6}$$

The functions $f_1, \ldots, f_p$ are unknown and the error ϵ is still assumed to be normally distributed and independent of $x_1, \ldots, x_p$. Let $\mathbf{y} = (y_1, \ldots y_N)^t$ and $\mathbf{x_j} = (x_{j,1}, \ldots, x_{j,N})^t$ $(j = 1, \ldots, p)$ be the data of an additive model. By posing the twice differentiability of the unknown functions, the penalized least-squares functional

$$\mathcal{L}_p(g_1, \ldots, g_p) = \sum_{i=1}^{N} \frac{(y_i - g_1(x_{1,i}) \ldots - g_p(x_{p,i}))^2}{\sigma^2} + \sum_{j=1}^{p} \alpha_j \int \left[\frac{d^2 g_j(x_j)}{dx_j^2}\right]^2 dx_j \tag{7}$$

can now easily be formulated. An estimate for the unknown functions is thus the minimum of $\mathcal{L}_p$, but a direct minimization is quite clumsy and the optimal choice of the smoothing parameters is a difficult task. Fortunately, there is an iterative solution by the calculation of partial residuals for each function. The new function estimates are provided by stepwise smoothing of the partial residuals and the solution of $\min\{\mathcal{L}_p(g_1, \ldots, g_p)\}$ is obtained by cycling through the functions until the least-squares functional does not change.

Convergence of this *backfitting algorithm* is assured if smoothing splines with fixed smoothing parameters are used [1, 2].

(a)

(b)

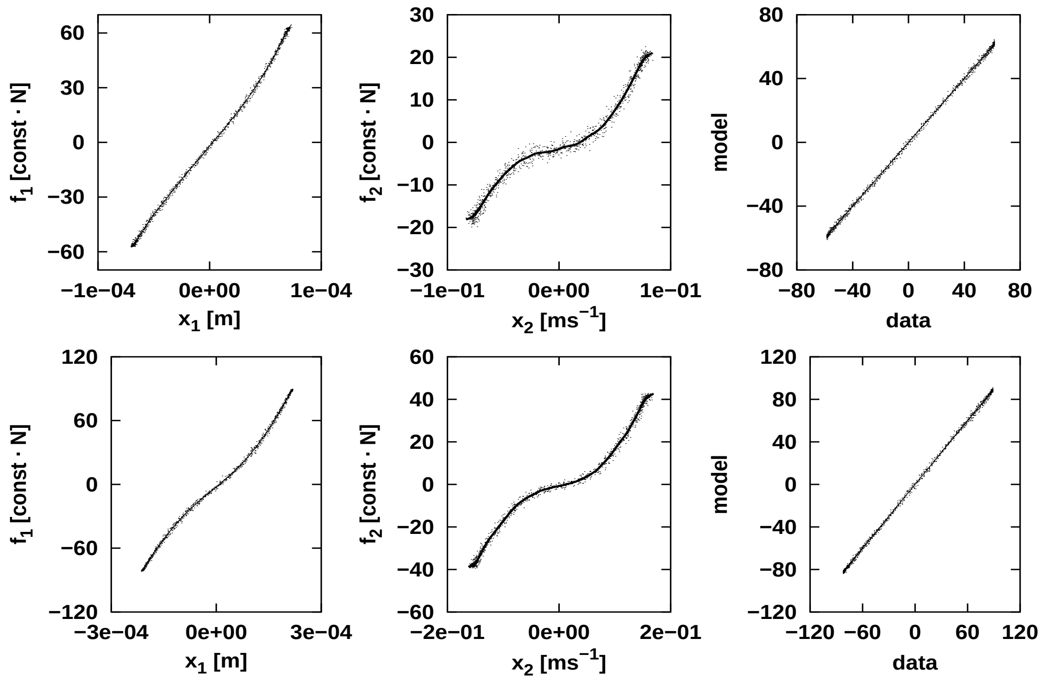

Figure 1: Estimated restoring force f_1 and friction component f_2 at frequency $\nu = 120$ Hz, load mass $m = 2.2$ kg, and excitation amplitude $U = 3$ V (a); $U = 10$ V (b). Dots are indicating the partial residuals of the functions in the first two columns. A comparison between a model based simulation and the data is shown in the third column.

3. Application

The suggested method is applied on a mechanical system to identify the nonlinear restoring force of helical wire rope isolators. The experiment was proposed by VTT Technical Research Centre of Finland within the framework of COST (European Cooperation in the field of Scientific and Technical Research). It consists of a bottom plate, a top plate and the wire rope isolators. These wire rope isolators are mounted between the top and the bottom plate. The bottom plate is excited by an electro-dynamic shaker and the top plate is the load for the isolators.

Measured are the acceleration a_1 of the bottom plate, the acceleration a_2 of the top plate, the displacement x_1 between the upper and the lower plate and some forces which are not used in our analysis. The behavior of the isolators are recorded under several experimental conditions. The load mass m_2 is 2.2 kg, and there are two kinds of excitation, harmonic and white noise. The amplitude and frequency was varied for the harmonic excitation and different variances were chosen for the random excitation.

Let y be the inertial force of the upper plate and $x_2 = \dot{x}_1$ be the relative velocity between the two plates. The observed oscillations exhibit a hysteresis which is smeared out in the case of white noise excitation. It is therefore likely to find a nonlinear restoring force and a non-vanishing friction component. With these variables,

$$y + f(x_1, x_2) = 0 \tag{8}$$

is the equation of motion of the upper plate. We further assume that the unknown function f is additive. Hence, $f(x_1, x_2) = f_1(x_1) + f_2(x_2)$, where f_1 is the stiffness force and f_2 is a damping term. In the case of white noise excitation, Kerschen et al. [10] proposed a model for this system. This model successfully explains the data and we therefore consider only the harmonic case in our analysis.

Since there is no measurement for the velocity x_2, we estimate it by numerical differentiation of the displacement x_1 which is very sensitive to noise. For sufficiently large amplitudes of the excitation force, the observational noise is negligible small, such that a high accuracy of x_2 was achieved.

Figure 1 shows the results of the analysis under different experimental conditions. In both cases, the restoring force and the damping are nonlinear. The major difference between these results and increased load mass is a change of the nonlinearity with respect to symmetry. Whereas for small load mass both the restoring force and the friction component are almost point-symmetric, this property is lost for a larger load mass of $m = 5.8$ kg (not shown).

The fitted model can be used for predictions of the observable y. These predictions are in good accordance with the data (Fig. 1, third column). According to the assumptions, the residuals (the differences between the predictions and the measurements) should be a realization of a normally distributed random variable. To verify this, we performed a Kolmogorov-Smirnov test and could not reject the null hypothesis in any case. The condition of an independent observation error was verified by calculating the power spectrum of the residuals, which were found to be flat.

4. Conclusion

We proposed a method for estimating nonlinear additive models and applied it to estimate functions in ordinary differential equations, using a completely data-adaptive procedure based on smoothing and generalized cross-validation. A case not considered here are noisy design points which can induce strongly biased function estimates, the so-called errors in variables problem. If this turns out to be a problem, this method can be used as an explorative tool to formulate a parametric model. Such a two-step strategy has been successfully applied in [11, 12].

Acknowledgment

We would like to thank G. Kerschen for stimulating discussions and comments, and the VTT Technical Research Centre of Finland and COST for making the data available.

References

[1] Buja A., Hastie T., and Tibshirani R. Linear smoothers and additive models. *Ann. Statist.*, 17:453–555, 1989.

[2] Hastie T.J. and Tibshirani R.J. *Generalized Additive Models.* Chapman and Hall, London, 1990.

[3] Green P.J. and Silverman B.W. *Nonparametric Regression and Generalized Linear Models.* Chapman and Hall, London, 1994.

[4] Thompson J.R. and Tapia R. A. *Nonparametric Function Estimation, Modeling, and Simulation.* SIAM, Philadelphia, 1990.

[5] Craven P. and Wahba G. Smoothing noisy data with spline functions. *Num. Math.*, 31:377–403, 1979.

[6] Härdle W. *Applied Nonparametric Regression.* Cambridge University Press, Cambrige, 1990.

[7] Reinsch C.H. Smoothing by spline functions. *Num. Math.*, 10:177–183, 1967.

[8] Silverman B.W. Spline smoothing: The equivalent variable kernel method. *Ann. Statist.*, 12:898–916, 1984.

[9] Silverman B.W. A fast and efficient cross-validation method for smoothing parameter choice in spline regression. *J. Amer. Statist. Assoc.*, 79:584–589, 1984.

[10] Kerschen G., Lenaerts V., and Golinval J.C. Identification of wire rope isolators using the restoring force surface method. In *International Conference on Structural System Identification, Kassel*, 2001.

[11] Voss H.U., Horbelt W., and Timmer J. Continuous nonlinear delayed-feedback dynamics from noisy observations. In: *Proceedings of the 6th Experimental Chaos Conference, Potsdam*, 2001.

[12] Timmer J., Rust H., Horbelt W., and Voss H.U. Parameteric, nonparametric and parametric modelling of a chaotic circuit time series. *Phys. Lett. A*, 274:123–134, 2000.