Proceedings of the 9th WSEAS International Conference on Applied Mathematics, Istanbul, Turkey, May 27-29, 2006 (pp347-350)

A Model Identification Approach Using MINLP Techniques

STEFAN EMET
Department of Mathematics
University of Turku
20014 Turku
FINLAND

Abstract: The present paper presents an approach of identifying the structure of a dynamic system using Mixed Integer Nonlinear Programming (MINLP) techniques. It is shown that the problem can be tackled by minimizing, for example, Akaike's Information Criterion (AIC).
The presented techniques are applied in determining the structure and the parameters of some illustrative Auto-Regressive Moving Average (ARMA) time series. The example problems are solved using the Extended Cutting Plane (ECP) method.

Key–Words: Akaike's Information Criterion, Parameter Estimation, Mixed Integer Non-Linear Programming, Extended Cutting Plane Method

1 Introduction

Given a set of observations y_i, made at time points t_i, $i = 1, \dots, N$, the classical least squares problem of estimating the parameters θ_i, $i = 1, \dots, p$, of a function f, describing the underlying dynamic system, can be formulated as follows:

$$\min_{\theta_1,\dots,\theta_p} \left\{ \sum_{i=1}^{N} (f(t_i,\theta) - y_i)^2 \right\} \tag{1}$$

A tool that can be applied for this purpose is the Akaike's Information Criterion (AIC) [1]:

$$-2\ln L + 2p \tag{2}$$

where L denotes the maximum likelihood function and p is the number of parameters. The logarithm of L, for N normally distributed, independent, random variables ε_k, with a variance σ^2, is given in [3]:

$$\ln L = -\frac{N}{2}\ln(2\pi) + \frac{N}{2}\ln(|\frac{1}{\sigma^2}|) - \frac{1}{2\sigma^2}\sum_{k=1}^{N}\varepsilon_k^2 \tag{3}$$

Substituting (3) in (2) yields

$$N(\ln(2\pi)+\ln(\sigma^2)) + \frac{1}{\sigma^2}\sum_{k=1}^{N}(f(t_k,\theta)-y_k)^2 + 2p \tag{4}$$

If the variance σ^2 is known or predefined [4], the criterion to be minimized can be formulated as follows:

$$\min_{\theta_1,\dots,\theta_p} \left\{ \frac{1}{\sigma^2}\sum_{k=1}^{N}(f(t_k,\theta) - y_k)^2 + 2p \right\} \tag{5}$$

Note, that the objective in (5) consists of the sum of squared residuals and a penalty term given by the number of parameters p. The penalty term can easily be modified in such a way that lower or higher priority is put on the number of parameters, the sample size and/or sample costs etc. In (5) it is assumed that p, which also gives the structure of f, is known. In many cases it is difficult to choose p in advance. One possibility is to solve (5) for different values of p and observe which value gives the best solution [7]. In the following, an alternative approach that enables the solving of both the structure determination and parameter estimation simultaneously, is presented.

2 MINLP Formulation

Introducing a binary variable, δ_i, for the existence of each parameter θ_i, and corresponding lower and upper bounds, $\theta_{i,\min}$ and $\theta_{i,\max}$, the optimization problem (5) can be formulated as follows:

$$\begin{aligned} \min \left\{ \tfrac{1}{\sigma^2}\sum_{i=1}^{N}(f(t_i,\theta) - y_i)^2 + 2\sum_{i=1}^{n}\delta_i \right\} & \\ \theta_i - \theta_{i,\max}\cdot\delta_i \le 0 & \\ -\theta_i + \theta_{i,\min}\cdot\delta_i \le 0 & \\ \delta_i \in \{0,1\},\ i = 1,\dots,n. & \end{aligned} \tag{6}$$

Problem (6) is generally called a Mixed Integer Non-Linear Programming (MINLP) problem. Assume that f is convex and that σ^2 is defined [4], then the objective function in (6) is also convex, which is preferable, because many numerical MINLP-methods provide theoretical guarantees of global optimality for convex problems.

The minimization of Akaike's criterion (6) might, however, in some applications result in redundant parameters. In order to keep the number of parameters of moderate size it is possible to use, for example, the Bayesian Information Criterion (BIC) [1]:

$$\frac{1}{\sigma^2}\sum_{i=1}^{N}(f(t_i,\theta)-y_i)^2+(1+\ln N)p \qquad (7)$$

In (7) the number of observations, that is, the sample size, N, is damping the number of parameters. The corresponding MINLP formulation can be written as follows:

$$\begin{aligned}\min\Big\{\tfrac{1}{\sigma^2}\sum_{i=1}^{N}(f(t_i,\theta)-y_i)^2+(1+\ln N)\sum_{i=1}^{n}\delta_i\Big\}&\\ \theta_i-\theta_{i,\max}\cdot\delta_i\le 0&\\ -\theta_i+\theta_{i,\min}\cdot\delta_i\le 0&\\ \delta_i\in\{0,1\}, i=1,\ldots,n.&\end{aligned} \qquad (8)$$

Another approach which also takes the sample size into consideration is given by the Final Prediction Error (FPE):

$$\frac{N+p}{N-p}\cdot\sum_{i=1}^{N}(f(t_i,\theta)-y_i)^2 \qquad (9)$$

Though f is convex, the function in (9) is nonconvex with respect to p, which might induce difficulties in the solving. Alternatively, a global optimization method can be applied if (9) is to be minimized.

3 Estimation of some time series

An Autoregressive Moving Average (ARMA) series of the order (p,q) can be written as follows:

$$X_t=\theta_0+\sum_{i=1}^{p}\theta_i X_{t-i}+\varepsilon_t+\sum_{i=1}^{q}\beta_i\varepsilon_{t-i} \qquad (10)$$

where $t=1,2,\ldots,N$, $X_0=0$, $\varepsilon_0=0$. The variables ε_t are normally distributed random ones with the mean 0 and the variance 1. Hence, there are $p+1+q+1$ parameters to be estimated ($\theta_0,\ldots,\theta_p,\beta_1,\ldots,\beta_q$ and σ^2). The residuals of (10) can be calculated recursively as follows:

$$\begin{aligned}\epsilon_1&=X_1-\theta_0\\ \epsilon_2&=X_2-\theta_0-\theta_1X_1-\beta_1\epsilon_1\\ \epsilon_3&=X_3-\theta_0-\theta_1X_2-\theta_2X_1-\beta_1\epsilon_2-\beta_2\epsilon_1\\ &\ \vdots\\ \epsilon_{p+1}&=X_{p+1}-\theta_0-\theta_1X_p-\ldots-\theta_pX_1\\ &\quad-\beta_1\epsilon_p-\beta_2\epsilon_{p-1}-\ldots\\ &\ \vdots\\ \epsilon_N&=X_N-\theta_0-\sum_{i=1}^{p}\theta_iX_{N-i}-\sum_{i=1}^{q}\beta_i\epsilon_{N-i}\end{aligned} \qquad (11)$$

Given N data points, $X_1,\ X_2,\ldots,\ X_N$, and assuming that $p>q$ the sum of squares, $\sum_{i=p+1}^{N}\epsilon_i^2$ can be used in problem formulations (6) and (8). Note, that if $q=0$, (10) is a so-called Autoregressive time series of the order p, denoted by AR(p). For AR(p) series the residuals in (11) are linear functions, with respect to the parameters, and, hence the sum of squared residuals is a convex function. Hence, there are theoretical guarantees of finding the global optimal solution for AR(p) series. On the other hand, if $q>0$, it can be noted from (10) that the residuals can be calculated recursively as follows:

$$\begin{aligned}\epsilon_1&=X_1-\theta_0\\ \epsilon_2&=X_2-\theta_0-\theta_1X_1-\beta_1(X_1-\theta_0)\\ &=X_2-\theta_0-(\theta_1+\beta_1)X_1+\underline{\theta_0\beta_1}\\ \epsilon_3&=X_3-\theta_0-\theta_1X_2-\theta_2X_1-\beta_1\epsilon_2-\beta_2\epsilon_1\\ &=X_3-\theta_0-\theta_1X_2-\theta_2X_1\\ &\quad-\beta_1(X_2-\theta_0-(\theta_1+\beta_1)X_1+\underline{\theta_0\beta_1})\\ &\quad-\beta_2(X_1-\theta_0)\\ &=X_3-\theta_0-\theta_1X_2-\theta_2X_1\\ &\quad-\beta_1X_2+\underline{\theta_0\beta_1}+\underline{\theta_1\beta_1}+\underline{\beta_1^2}-\underline{\theta_0\beta_1^2}\\ &\quad-\beta_2X_2+\underline{\theta_0\beta_2}\\ \epsilon_4&=\ldots\end{aligned} \qquad (12)$$

The signomial terms, that are underlined in (12), are in general non-convex which means that the resulting optimization problem will be of a non-convex form. There are, however, strategies for convexifying signomials [9] that can be applied.

The variance of the residual was estimated iteratively (based on previous estimations [6]) using the following formula presented in [7]:

$$\hat{\sigma}_k^2=\frac{1}{N-2p-q-1}\sum_{i=p+1}^{N}\epsilon_i^2 \qquad (13)$$

Proceedings of the 9th WSEAS International Conference on Applied Mathematics, Istanbul, Turkey, May 27-29, 2006 (pp347-350)

The procedure can be started with $\sigma_k^2 = 1$. Computational studies [4] indicate that the above algorithm works and relatively few iterations are needed to achieve convergence. Hence, the optimization problem (6) can be summarized as follows:

$$\begin{aligned} \min \quad & \left\{ \frac{1}{\sigma_k^2} \sum_{i=1}^{N} \epsilon_i^2 + 2 \sum_{i=1}^{n} \delta_i \right\} \\ \text{s.t.} \quad & \theta_k - \theta_{k,\max} \cdot \delta_k \le 0 \\ & -\theta_k + \theta_{k,\min} \cdot \delta_k \le 0 \\ & \beta_l - \beta_{l,\max} \cdot \delta_l \le 0 \\ & -\beta_l + \beta_{l,\min} \cdot \delta_l \le 0 \\ & \delta_i \in \{0,1\},\ i = 1, \dots, n. \end{aligned} \tag{14}$$

where the residuals, ϵ_i, are given in (12). The problem (8) can be formulated similarly using the original objective function in (8) and the constraints in (14).

4 Numerical examples

The MINLP formulations in (6) and (8) were applied on the following set of series:

$$X_t = \epsilon_t + \frac{1}{3} X_{t-1} \tag{15}$$
$$X_t = \epsilon_t + 0.25 X_{t-1} - 0.75 X_{t-2} \tag{16}$$
$$X_t = \epsilon_t + 0.7 X_{t-1} - 0.2 X_{t-2} + 0.5 X_{t-3} \tag{17}$$
$$X_t = \epsilon_t + 0.9 \epsilon_{t-1} \tag{18}$$
$$X_t = \epsilon_t - 0.7 \epsilon_{t-1} + 0.5 \epsilon_{t-2} \tag{19}$$
$$X_t = \epsilon_t + 0.75 \epsilon_{t-1} + 0.75 X_{t-1} \tag{20}$$
$$X_t = \epsilon_t - 0.8 \epsilon_{t-1} + 0.5 X_{t-2} \tag{21}$$

The corresponding MINLP problems were solved with the Extended Cutting Plane method (ECP), [8], which is an extension of Kelley's method [5]. The ECP method is a general purpose MINLP method with applicability to a large variety of problems [2], [6]. The results of determining the order and of estimating the parameters of data generated from series (15)-(21), using the AIC-criterion (6) and the BIC-criterion (8), are presented in Table 1. The profiles of the data and the corresponding estimations of series (16) are illustrated in Figure 1. The results of (15)-(17) using the AIC-criterion are quite good since these problems are convex, and thus the solutions are global optimal solutions. It can be noted, that the series (18)-(21) implied non-convexities, which means that improved solutions could be obtained using convexification strategies [9].

Table 1: Results of series (15)-(21).

series	estimated model
(15)	$\varepsilon_t + 0.339X_{t-1}$
	$\varepsilon_t + 0.339X_{t-1}$
(16)	$\varepsilon_t + 0.1233X_{t-1} - 0.7562X_{t-2}$
	$\varepsilon_t - 0.7571X_{t-2}$
(17)	$\varepsilon_t + 0.776X_{t-1} - 0.242X_{t-2} + 0.434X_{t-3}$
	$\varepsilon_t + 0.660X_{t-1} + 0.309X_{t-3}$
(18)	$\varepsilon_t + 0.904\varepsilon_{t-1}$
	$\varepsilon_t + 0.916\varepsilon_{t-1}$
(19)	$\varepsilon_t - 0.794\varepsilon_{t-1} + 0.450\varepsilon_{t-2}$
	$\varepsilon_t - 0.822\varepsilon_{t-1} + 0.468\varepsilon_{t-2}$
(20)	$\varepsilon_t + 0.623\varepsilon_{t-1} + 0.745X_{t-1}$
	$\varepsilon_t + 0.620\varepsilon_{t-1} + 0.746X_{t-1}$
(21)	$\varepsilon_t - 0.872\varepsilon_{t-1} - 0.371X_{t-2}$
	$\varepsilon_t - 0.854\varepsilon_{t-1} - 0.347X_{t-2}$

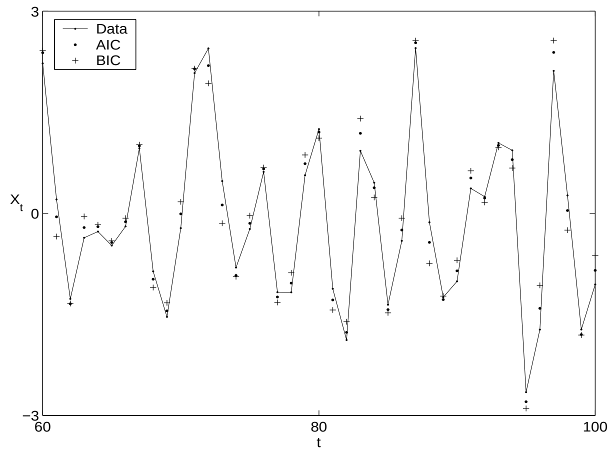

Fig. 1. Generated and estimated data of series (16).

5 Conclusion

In the present paper, techniques for determining the structure and estimating parameters of a system were presented. It was shown that such problems can be modeled and solved as MINLP problems by minimizing the Akaike's information criterion.

Examples of joint model structure determination and parameter estimation of some time series were finally illustrated. The examples were numerically solved using the ECP-method that has been proven efficient on many complex engineering problems. The results were encouraging, the presented methods can in an analogous way also be applied on multivariate systems.

References:

[1] H. Akaike, A New Look at Statistical Model Identification. *IEEE Trans. Automat. Contr.* 19, 1974, pp. 716–722.

[2] I. Castillo, J. Westerlund, S. Emet and T. Westerlund: Optimization of Block Layout Design Problems with Unequal Areas: A Comparison of MILP and MINLP Optimization Methods. *Computers & Chemical Engineering* 30, 2005, pp. 54–69.

[3] A. M. Goodwin and R. L. Payne: *Dynamic System Identification. Experimental Design and Data Analysis*. Academic Press, 1977.

[4] S. Karrila and T. Westerlund, An Elementary Derivation of the Maximum Likelihood Estimator of the Covariance Matrix, and an Illustrative Determinant Inequality, *AUTOMATICA* 27, 1990, pp. 425–426.

[5] J. Kelley, The cutting-plane method for solving convex programs. *Journal of SIAM* 8, 1960, pp. 703–712.

[6] H. Skrifvars, S. Leyffer and T. Westerlund, Comparison of certain MINLP algorithms when applied to a model structure determination and parameter estimation problem. *Computers chem. Engng* 22, 1998, pp. 1829–1835.

[7] M. B. Priestley: *Spectral Analysis an Time Series*. Academic Press Inc., 1981.

[8] T. Westerlund and R. Pörn, Solving Pseudo-Convex Mixed Integer Problems by Cutting Plane Techniques. *Optimization and Engineering* 3, 2002, pp. 253–280.

[9] T. Westerlund, Some Transformation Techniques in Global Optimization. *Global Optimization, from Theory to Implementations*, Liberti L. and Maculan N. (eds.), 2006, pp. 45–74.