SYSTEM IDENTIFICATION IN THE PRESENCE OF TRENDS AND OUTLIERS

Amir Shirdel, Jari M. Böling and Hannu T. Toivonen
Åbo Akademi University,
FIN–20500 Åbo, Finland
E-mail:ashirdel@abo.fi

Introduction

During the operation process many quantities and parameters could be blended with some unwanted data which might be fluctuated in times. In other words, the measured variables in system identification are often disturbed by some kinds of structural disturbances (such as trends, levels shifts or outliers) and noises. It is important to eliminate outliers and trends in the data, as these might otherwise deteriorate the identification accuracy [**?**].According to [**?**] ,Trend filtering applications are so vast and very useful in several areas such as macroeconomics [**?**], [**?**], geophysics [**?**],[**?**],[**?**], financial time series analysis [**?**], social sciences [**?**], revenue management [**?**], and biological and medical sciences [**?**][**?**].

Sparse Representation of Signal

The problem of finding the sparse representation of a signal is that ,when the equations are linear, one would like to determine an object $x_0 \in R^n$ from data $y = \Phi x_0$, where Φ is an $m \times n$ matrix with fewer rows than columns; i.e., $m > n$. The problem is that a system with fewer equations than unknowns usually has infinitely many solutions and thus, it is apparently impossible to identify which of these candidate solutions is indeed the correct one without some additional information [**?**].

In [**?**], by using the property of l_1 regularization in regression, estimated the model parameter of system.

L1 Filtering

L1 trend filtering proposed the minimizer of the weighted sum objective function which can be written in matrix form as

$$(1/2)\sum_{k=1}^{N}\Big(y(k)-x(k)\Big)^2+\lambda\|D_i x\|_1 \qquad (1)$$

λ is a non-negative parameter for controlling the trade-off between smoothness of x and size of the residual. The weighted sum objective (**??**) is strictly convex and coercive in x and so has a unique minimizer [**?**] .

Problem formulation

We consider a linear system

$$y_0(k) = a_1 y_0(k-1) + \cdots + a_n y_0(k-n) + b_1 u(k-1) + \cdots + b_m u(k-m) + e(k) \tag{2}$$

where is $e(k)$ random noise disturbance. It is assumed the measured output is given by

$$y(k) = y_0(k) + d(k) \tag{3}$$

where $d(k)$ is an unknown variable, which we assume can be described as the outlier signal, level shifts and piecewise constant trends. A sequence of outliers can be modeled by Sequence of outliers:

$$d_0(k) = \begin{cases} d_i, & k = k_i, i = 1, \ldots, M_0 \\ 0, & \text{otherwise} \end{cases}$$

Similarly, level shifts are described by piecewise a constant variable

$$d_1(k) = d_i, \;\; k_i \le k < k_{i+1}, i = 1, \ldots, M_1$$

and Sequence of trends are modeled by a piecewise linear signal

$$d_2(k) = d_2(k-1) + \beta_i, k_i \le k < k_{i+1}, i = 1, \ldots, M_2$$

The problem is to identify the system model (**??**) and the variables $d(k)$ from a sequence $\{y(k), u(k), k = 1, \ldots, N\}$ of measured outputs $y(k)$ and known inputs $u(k)$. Note that it is assumed that neither the values nor the time instant k_i of the discontinuities are known. However, it can be assumed that the number of discontinuities M_0, M_1 or M_2 is small in relation to the total data points.

Combining (**??**) and (**??**) gives

$$\begin{aligned} y(k) &= a_1\Big(y(k-1) - d(k-1)\Big) + \cdots + a_n\Big(y(k-n) - d(k-n)\Big) + d(k) \\ &\quad + b_1 u(k-1) + \cdots + b_m u(k-m) + e(k) \end{aligned} \tag{4}$$

or

$$y(k) = \theta_a \varphi_y(k) + \theta_b \varphi_u(k) - \theta_a \varphi_d(k) + d(k) + e(k) \tag{5}$$

where

$$\begin{aligned} \theta_a &= [a_1 \; \cdots \; a_n]^{\mathrm{T}} \\ \theta_b &= [b_1 \; \cdots \; b_m]^{\mathrm{T}} \end{aligned} \tag{6}$$

and

$$\begin{aligned} \varphi_y(k) &= [y(k-1) \; \cdots \; y(k-n)]^{\mathrm{T}} \\ \varphi_u(k) &= [u(k-1) \; \cdots \; u(u-m)]^{\mathrm{T}} \\ \varphi_d(k) &= [d(k-1) \; \cdots \; d(k-n)]^{\mathrm{T}} \end{aligned} \tag{7}$$

In identification of the system parameters θ_a and θ_b, and the disturbance sequence $d(k)$, one needs to take into account that the disturbance sequence $\{d(k)\}$ can always be selected so that the model output matches the measured output exactly. However, the disturbance sequence would in general not satisfy the condition that the number of discontinuities be small. In order to satisfy this condition while achieving a small prediction error, sparse optimization will be applied.

A sparse optimization approach

Normally the standard approach in system identification is to remove disturbances by data preprocessing first, then try to estimate the system, but this kind of method is difficult to separate between the effects of known system inputs and unknown disturbances (trends, etc.).
In sparse optimization approach, we are trying to remove disturbances as before but also identifying the model system simultaneously by exploiting sparsity properties of $d(k)$ and minimize an objective function of the form

$$J(\theta_a, \theta_b, d) = \sum_{k=1}^{N} \Big(y(k) - \theta_a \varphi_y(k) - \theta_b \varphi_u(k) + \theta_a \varphi_d(k) - d(k) \Big)^2 + \lambda \|D_i d\|_1 \tag{8}$$

where λ is a positive constant,

$$d = [d(1) \ \cdots \ d(N)]^{\mathrm{T}}$$

and D_i is a weighting matrix selected in accordance with the type of disturbance.

Iterative refinement

As in sparse optimization, for good results reweighting may have to be performed iteratively, and we need to study how these can be interpreted in terms of the primal problem. In [**?**], proposed a weighted formulation of l_1 minimization designed to more democratically penalize nonzero coefficients: larger coefficients are penalized more heavily in the l_1 norm than smaller coefficients, unlike the more democratic penalization of the l_0 norm.
Algorithm 1 shows the sparse optimization with iterative refinement as follows:

Algorithm 1

Step 1. Set W=1 and minimize the cost by using Bilinear Matrix Inequalities

$$J_1(\hat{\theta}_a, \hat{\theta}_b, d) = \sum_{k=1}^{N} \Big(y(k) - \hat{\theta}_a \varphi_y(k) - \hat{\theta}_b \varphi_u(k) + \hat{\theta}_a \varphi_d(k) - \hat{d}(k) \Big)^2 + \lambda \|W D_i \hat{d}\|_1 \tag{9}$$

for the required estimates $\hat{\theta}_a$, $\hat{\theta}_b$ and $\hat{d}(k)$.

Step 2. Use re-weighting to put small weight on the points that have large error and reduce the affection of them on identification parameter of system.

$$W = diag(\frac{1}{\varepsilon + |D_i d(k)|}) \tag{10}$$

Step 3. Minimize the cost with the new weight

$$J_1(\hat{\theta}_a, \hat{\theta}_b, d) = \sum_{k=1}^{N} \Big(y(k) - \hat{\theta}_a \varphi_y(k) - \hat{\theta}_b \varphi_u(k) + \hat{\theta}_a \varphi_d(k) - \hat{d}(k) \Big)^2 + \lambda \|W_{new} D_i \hat{d}\|_1 \tag{11}$$

in each iteration we calculate new weight and apply to our minimization in step 3, if our result is perfect and converge good we exit and Finish the iterative.

AIC model selection

One of the popular statistical methods for selecting the model of the system in system identification is Akaike Information Criterion. In this part, finding the best lambda is out goal to find better fittting with simpler model in our identification and trends.

Standard AIC is only based on maximum log-likelihood and model order and it typically selects more complex model as the sample size increases.For solving this problem and select better model we need to penalize the maximized log-likelihood with more precisely calibrated factor. In corrected AIC, penalizes complexity more strongly than standard AIC, with less chance of over fitting the model [**?**].

Numerical examples

In this section, we apply the proposed identification de-trending method to an ARX models.

Example 1

The system is defined by

$$\begin{aligned} y_0(k) &= a_1 y_0(k-1) + a_2 y_0(k-2) + b_1 u(k-1) \\ &\quad + b_2 u(k-2) + e(k) \end{aligned} \qquad (12)$$

where the parameter vector

$$\theta = [a_1 \quad a_2 \quad b_1 \quad b_2]^T$$

which is

$$\theta = [1.50 \quad -0.7 \quad 1.00 \quad 0.5]^T, \qquad (13)$$

where $u(k)$ and $e(k)$ is Normally distributed signal with variances 1 and 0.1, and $d(k)$ is a structured disturbance, which was added to the original system to make measured output. we should model the disturbance as

$$d(k) = d_0(k) + d_1(k) + d_2(k) \qquad (14)$$

where $d_0(k), d_1(k), d_2(k)$ are the outlier signal, level shifts and piecewise constant trends.

It is assumed the measured output is given by

$$y(k) = y_0(k) + d(k) \qquad (15)$$

we change the form of our problem to equation **??** and then by minimizing the cost with Iterative refinement we can find the preliminary value of our identification.

In this example we add simultaneously three kinds of disturbances, trends, level shifts and spikes to the system.

The outlier (spikes) trends are

$$d_0(k) = \begin{cases} 3, & k = 330, 760 \\ -3, & k = 135, 255 \\ 0, & \text{otherwise} \end{cases}$$

The Level shifts are

$$d_1(k) = \begin{cases} \text{-3,} & \text{k=1-150} \\ \text{5,} & \text{k=151-250} \\ \text{1,} & \text{k=251-450} \\ \text{5,} & \text{k=500-650} \\ \text{-2,} & \text{k=651-1000} \end{cases}$$

The piece-wise trends are

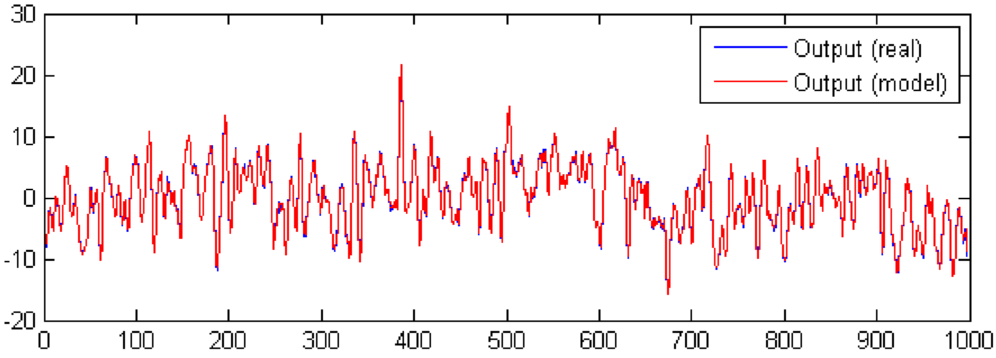
$$d_2(k) = \begin{cases} 2\ldots4, & \text{k=1-80} \\ 4\ldots\text{-1}, & \text{k=80-200} \\ \text{-1}\ldots4, & \text{k=200-400} \\ 4\ldots0, & \text{k=400-600} \\ 0\ldots3, & \text{k=600-880} \\ 3\ldots\text{-2}, & \text{k=880-1000} \end{cases}$$

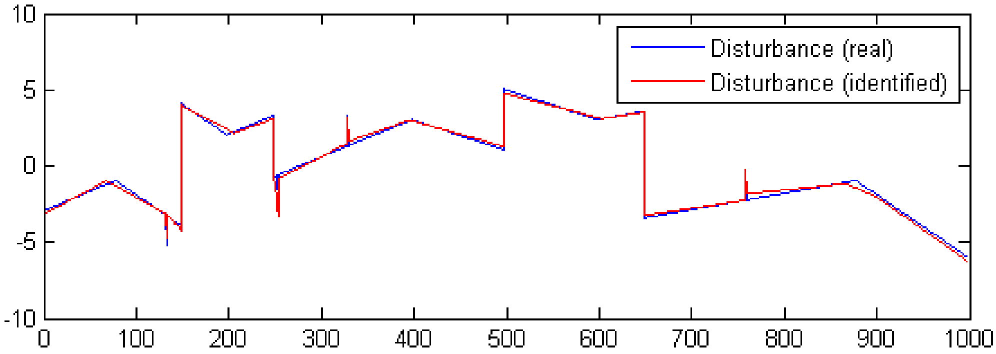
Table 1: Estimated system parameter in *Example* 1

$*$	a_1	a_2	b_1	b_2
θ_{real}	1.50	-0.7	1.0	0.5
$\hat{\theta}$	1.4971	-0.6960	1.0017	0.4987
θ_{LS}	1.5056	-0.6445	0.9966	0.4777

Table 2: Average of RMS value of the prediction error in *Example* 1

RMS E	0.3226
RMS PE	0.3269
RMS PELS	0.7476





Example 2: Distillation column

In this example we used the distillation column data [**?**] to identify system parameter and disturbances. The inputs are V (reboiling flow) and L(reflux) and output is overhead product of column. The inputs are PRBS signals that simultaneously apply to the system. By using AIC method,w e found the delay of system 6 and the parameter $\lambda = 0.4$.

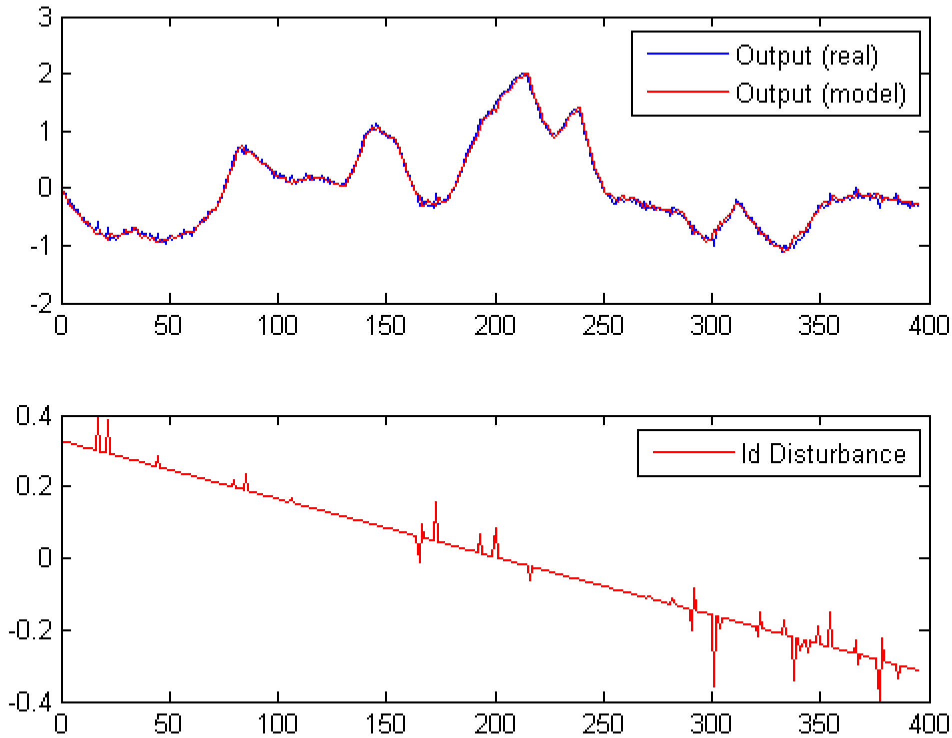
The following table shows the comparison between the error of predicted model with ordinary least square and our method predicted error.

Table 3: Estimated system parameter in *Example* 2

$*$	a_1	a_2	b_{V1}	b_{V2}	b_{L2}	b_{L2}
θ_{LS}	0.6220	0.3370	0.0105	0.0110	-0.0322	-0.0059
$\hat{\theta}$	0.6198	0.3355	0.0105	0.0127	-0.0324	-0.0077

Table 4: Average of RMS value of the prediction error in *Example* 2

RMSPE LS	0.064±0.005
RMS PE	0.053±0.005



References

[1] [System identification theory for the user,second edition, by Lennart Ljung]:

[2] L-1 trend filtering, S.-J. Kim, K. Koh, S. Boyd, and D. Gorinevsky.2009. SIAM Review, problems and techniques section, 51(2):339360, May 2009

[3] R. Hodrick and E. Prescott, Postwar U.S. business cycles: An empirical investigation, J.Money, Credit, and Banking, 29 (1997), pp. 116.

[4] K. Singleton, Econometric issues in the analysis of equilibrium business cycle models, J.Monetary Econom., 21 (1988), pp. 361386.

[5] R. Baillie and S. Chung, Modeling and forecasting from trend stationary long memory models with applications to climatology, Internat. J. Forecast., 18 (2002), pp. 215226.

[6] P. Bloomfeld, Trends in global temperature, Climate Change, 21 (1992), pp. 116.

[7] P. Bloomfeld and D. Nychka, Climate spectra and detecting climate change, Climate Change, 21 (1992), pp. 275287.

[8] R. Tsay, Analysis of Financial Time Series, 2nd ed., Wiley-Interscience, Hoboken, NJ, 2005.

[9] S. Levitt, Understanding why crime fell in the 1990s: Four factors that explain the declineand six that do not, J. Econom. Perspectives, 18 (2004), pp. 163190.

[10] R. Talluri and G. van Ryzin, The Theory and Practice of Revenue Management, Kluwer Academic, Boston, MA, 2004.

[11] W. Link and J. Sauer, Estimating equations estimates of trend, Bird Populations, 2 (1994),pp. 2332.

[12] S. Greenland and M. Longnecker, Methods for trend estimation from summarized doseresponse-data, with applications to meta-analysis, Amer. J. Epidemiology, 135(1992),pp. 13011309.

[13] "Outliers, Level Shifts, and Variance Changes in Time Series." Ruey S. Tsay; Journal of Forecasting, 1988, 7(1), pp. 1.

[14] Tibshirani, R. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society, Series B, Vol 58, No. 1, pp. 267288, 1996.

[15] A.MOGHTADERI, P.BORGNAT and P.FLANDRIN, TREND FILTERING: EMPIRICAL MODE DECOMPOSITIONS VERSUS l1 AND HODRICKPRESCOTT,Advances in Adaptive Data Analysis Vol. 3, No. 1,2 (2011) 4161,World Scientic Publishing Company

[16] Emmanuel J. Cands , Michael B. Wakin and Stephen P. Boyd , Enhancing Sparsity by Reweighted L_1 Minimization, J Fourier Anal Appl (2008) 14: 877905

[17] Model Selection and Model Averaging by CLAESKENS, G. and HJORT, N. L. 2008, page 44.

[18] Jari boling,korterik hagbolom, report 08-01, process control lab, Abo akademi