Stanford Paper

Segmentation of ARX-models using sum-of-norms regularization

Henrik Ohlsson[a,*], Lennart Ljung[a], Stephen Boyd[b]

IEOR 135 meeting, April 3rd

Main idea

Traditional Linear regression: $y(t) = \varphi^T(t)\theta$ $\hat{\theta}(N) = \arg\min_{\theta} \sum_{t=1}^{N} \|y(t) - \varphi^T(t)\theta\|^2$

New Approach: $y(t) = \varphi^T(t)\theta(t)$ $\min_{\theta(t),\ t=1,\ldots,N} \sum_{t=1}^{N} \|y(t) - \varphi^T(t)\theta(t)\|^2$

What is happening here?

We are allowing parameters to be **time variant.**

Potential problem: **overfitting**

New Method

$$\min_{\theta(t)} \sum_{t=1}^{N} \|y(t) - \varphi^T(t)\theta(t)\|^2 + \lambda \sum_{t=2}^{N} \|\theta(t) - \theta(t-1)\|_{\text{reg}}, \qquad (6)$$

Parameter change is penalized by a second term to resolve overfitting

Large λ: worse fit, fewer segments

Small λ: better fit, more segments

λmax yields a time-constant solution

$$\lambda^{\max} = \max_{t=1,\ldots,N-1} \left\| \sum_{\tau=1}^{t} 2(y(\tau) - \varphi^T(\tau)\theta^{\text{const}})\varphi^T(\tau) \right\|_{\text{reg}*}$$

Usually, $(0.01)^*\lambda\text{max} < \lambda < \lambda\text{max}$

More Accuracy

$$\min_{\theta(t)} \sum_{t=1}^{N} \|y(t) - \varphi^T(t)\theta(t)\|^2 + \lambda \sum_{t=2}^{N} w(t)\|\theta(t) - \theta(t-1)\|_{\text{reg}}$$

Allow lambda impact to fluctuate over time for each variable

Iterative procedure for solving w(t):

(1) *Find the parameter estimate.*
Compute the optimal $\theta^{(i)}(t)$ with weighted regularization using weights $w^{(i)}$.

(2) *Update the weights.*
Set $w^{(i+1)}(t) = 1/(\epsilon + \|\theta^{(i)}(t) - \theta^{(i)}(t-1)\|_{\text{reg}})$.

Results

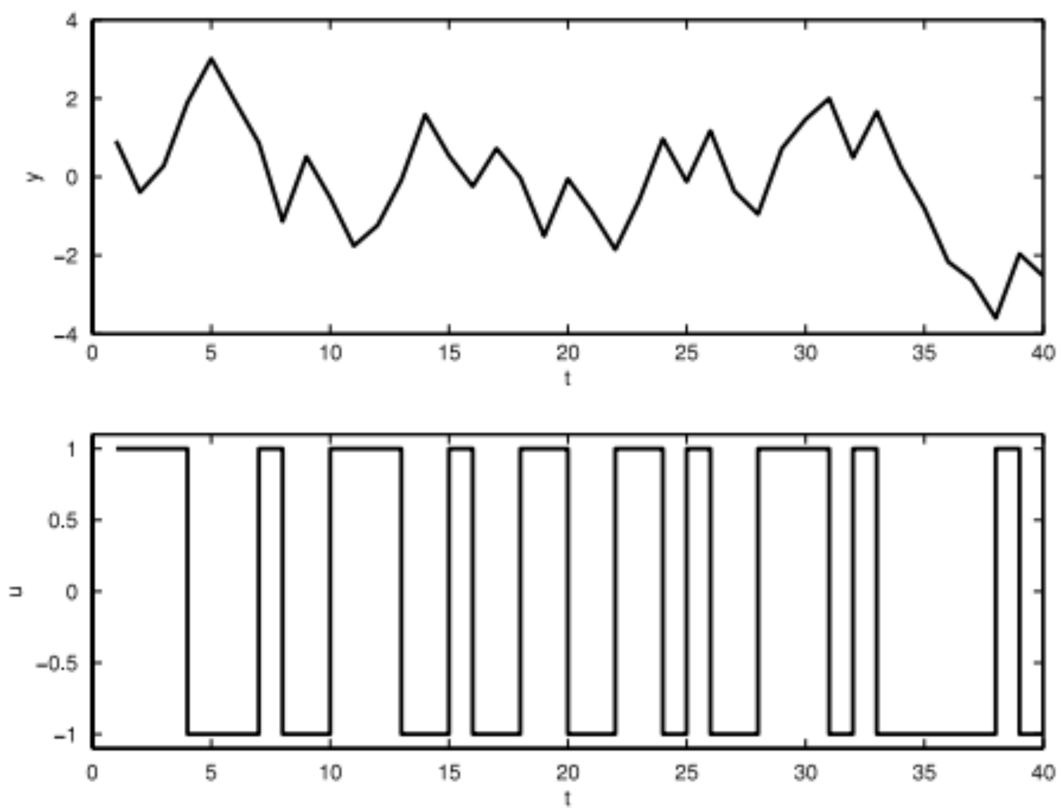

Fig. 1. The data used in Example 1.

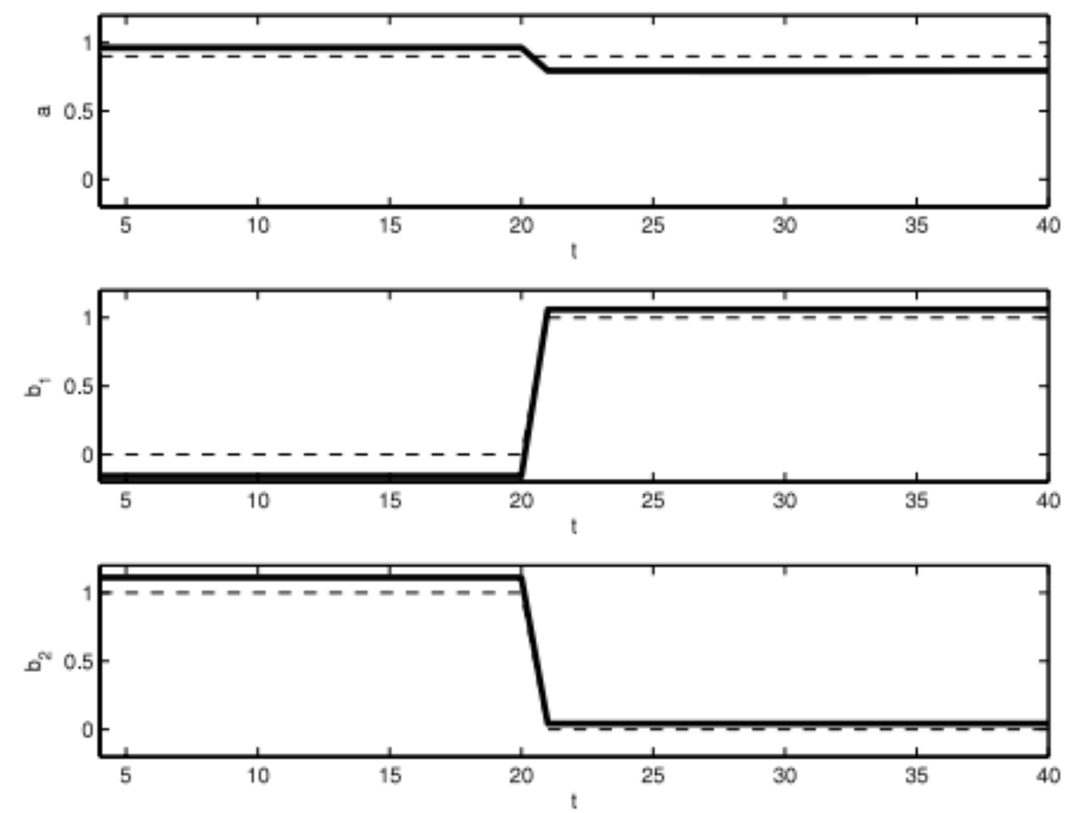

Fig. 2. The parameter estimates in Example 1. Solid lines show the parameter estimates and dashed lines the true parameter values.

Example 1 (*Changing Time Delay*). This example is from `iddemo11` in the System Identification Toolbox, (Ljung, 2007). Consider the system

$$y(t) + 0.9y(t-1) = u(t-n_k) + e(t).$$

The input u is a ± 1 PRBS (Pseudo-Random Binary Sequence) signal and the additive noise has variance 0.1. At time $t = 20$ the time delay n_k changes from 2 to 1. The data are shown in Fig. 1. An ARX-model

$$y(t) + ay(t-1) = b_1u(t-1) + b_2u(t-2)$$

is used to estimate a, b_1, b_2 with the method described in the previous section. The resulting estimates using $\lambda = 0.1\lambda^{\max}$ are shown in Fig. 2. The solid lines show the estimate and dashed the true parameter values. We clearly see that b_1 jumps from 0 to 1, to "take over" to be the leading term around sample 20. The estimate of the parameter a (correctly) does not change notably.

Results

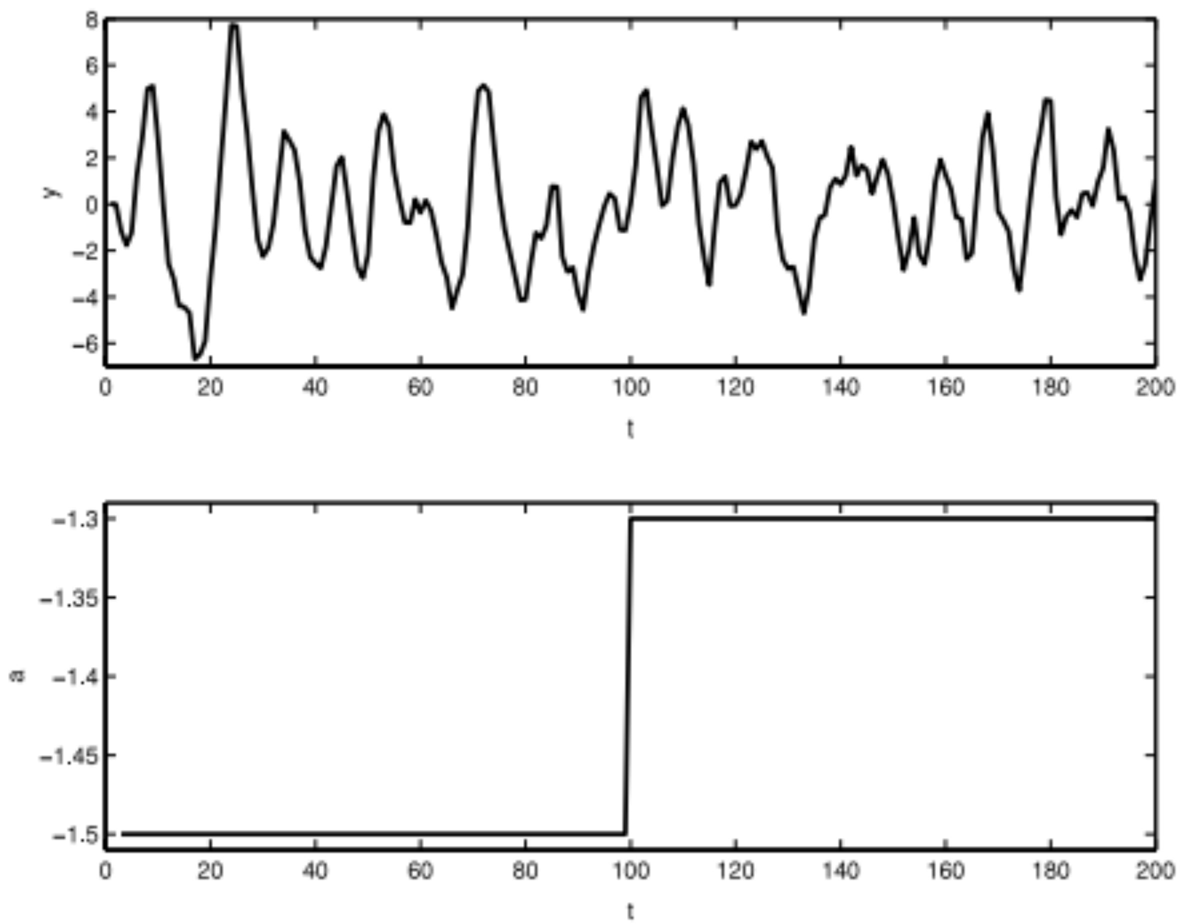

Fig. 3. The time series data (upper plot) and the estimate of a (lower plot) of Example 2.

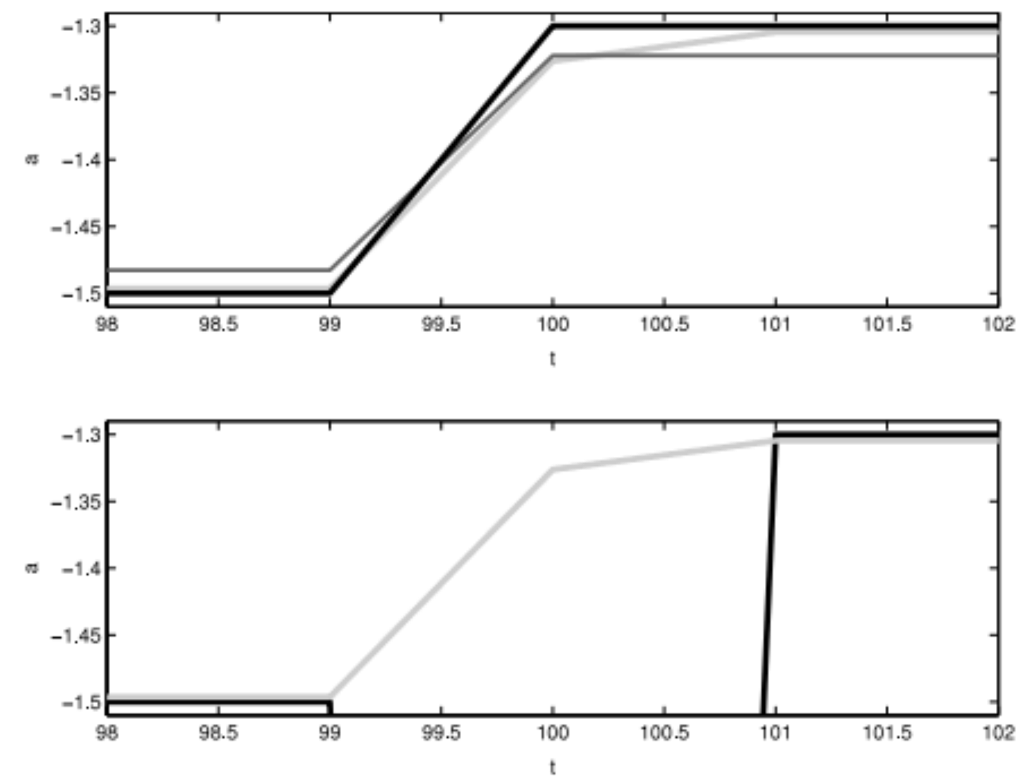

Fig. 4. Estimates of a in Example 2 with (top plot) and without (bottom plot) iterative refinement. Thick black line, estimate after least-squares has been applied to segments without changes in a and light-gray thick line, estimate given by (6). In the top plot, the gray thin lines show estimates of a after one and two iterative refinements (the two lines are not distinguishable). Without iterative refinement (bottom plot) a is estimated to -5.1 at $t = 100$.

Example 2 (*Changing Time Series Dynamics*). Consider the time series

$$y(t) + ay(t-1) + 0.7y(t-2) = e(t)$$

with $e(t) \sim \mathcal{N}(0, 1)$. At time $t = 100$ the value of a changes from -1.5 to -1.3. The output data and the estimate of a are shown in Fig. 3. $\lambda = 0.01\lambda^{\max}$ was used.

To motivate the iterative refinement procedure suggested in Section 2.3, let us see what happens if it is removed. Fig. 4 shows the estimate of a (around $t = 100$) with and without the refinement iteration. As shown by the figure, (6) incorrectly estimates the change at $t = 100$ and gives an estimate having a change both at $t = 100$ and $t = 101$. Using iterative refinement, however, this does not occur. Without iterative refinement, a is estimated to -5.1 at $t = 100$.