Results on the PASCAL PROMO challenge

Ivan Markovsky

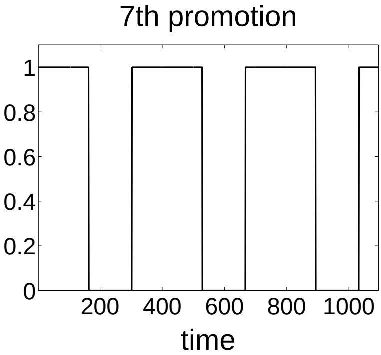
University of Southampton

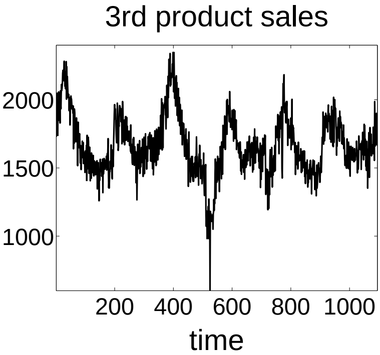
The challenge

Data: consists of two (simulated) time series

$$u_d(1), \ldots, u_d(1095) \quad \in \{0,1\}^{1000} \qquad \text{promotions}$$
$$y_d(1), \ldots, y_d(1095) \quad \in \mathbb{R}^{100} \qquad \text{product sales}$$

Challenge: find ≤ 50 promotions that affect each product sales





Comments

- time series nature of the data $\Longrightarrow$ dynamic phenomenon (the current output may depend on past inputs and outputs)
- it is natural to think of the promotions as inputs (causes) and the sales as outputs (effects)
- multivariable data: $m = 1000$ inputs, $p = 100$ outputs
- $T = 1095$ data points—very few, relative to m and p
- even static linear model $y = Au$ is unidentifiable (A can not be recovered uniquely from $(u_\mathrm{d}, y_\mathrm{d})$) for $T < T_\mathrm{min} := 10^5$
- prior knowledge that a few (≤ 50) inputs affect each output helps ($T_\mathrm{min} = 5000$) but doesn't recover identifiability
- this prior knowledge makes the problem combinatorial

Proposed model

Main assumptions:

1. static input-output relation $y_j(t) = a_j u(t)$
   (this implies that one output can not affect other outputs)

2. there is offset and seasonal component, which is sine, *i.e.*,

$$\text{Base line:} \qquad y_{\mathrm{bl},j}(t) := b_j + c_j \sin(\omega_j t + \phi_j)$$

The model is

$$y_j(t) = y_{\mathrm{bl},j}(t) + Au(t)$$

or, with $Y := \begin{bmatrix} y(1) & \cdots & y(T) \end{bmatrix}$, $U := \begin{bmatrix} u(1) & \cdots & u(T) \end{bmatrix}$, *etc.*,

$$Y = Y_{\mathrm{bl}}(b, c, \omega, \phi) + AU$$

Identification problem

Parameters:

$A \in \mathbb{R}^{p \times m}$	—	input/output (feedthrough) matrix
$b := (b_1, \dots, b_p) \in \mathbb{R}^p$	—	vector of offsets
$c := (c_1, \dots, c_p) \in \mathbb{R}^p$	—	vector of amplitudes
$\omega := (\omega_1, \dots, \omega_p) \in \mathbb{R}^p$	—	vector of frequencies
$\phi := (\phi_1, \dots, \phi_p) \in [-\pi, \pi]^p$	—	vector of phases

Identification problem:

minimize over the parameters $\|Y_\mathrm{d} - Y_\mathrm{bl}(b, c, \omega, \phi) - AU_\mathrm{d}\|$

subject to each row of A has at most 50 nonzero elements.

combinatorial, constrained, nonlinear, least squares problem

Solution approach

Model: $\widehat{y}_j(t) = b_j + c_j \sin(\omega_j t + \phi_j) + Au(t)$

Linear in A, b, c. Nonlinear in ω, ϕ. Combinatorial in A.

Our approach: Split the problem into two stages:

1. Baseline estim.: minimize over b, c, ω, ϕ, assuming $A = 0$.
   Nonlinear LS problem. We use local optimization.

2. I/O function etim.: minimize over A, b, c, with ω, ϕ fixed.
   This is a combinatorial problem. We use the ℓ_1 heuristic.

This approach simplifies the solution but leads to suboptimality.

Identification of the autonomous term

The problem decouples into p independent problems:

$$\text{minimize} \quad \text{over } b_j, c_j, \omega_j \in \mathbb{R},\ \phi_j \in [-\pi, \pi] \quad \| y_{\mathrm{d},j} - y_{\mathrm{bl},j}(b_j, c_j, \omega_j, \phi_j) \|_2 \tag{1}$$

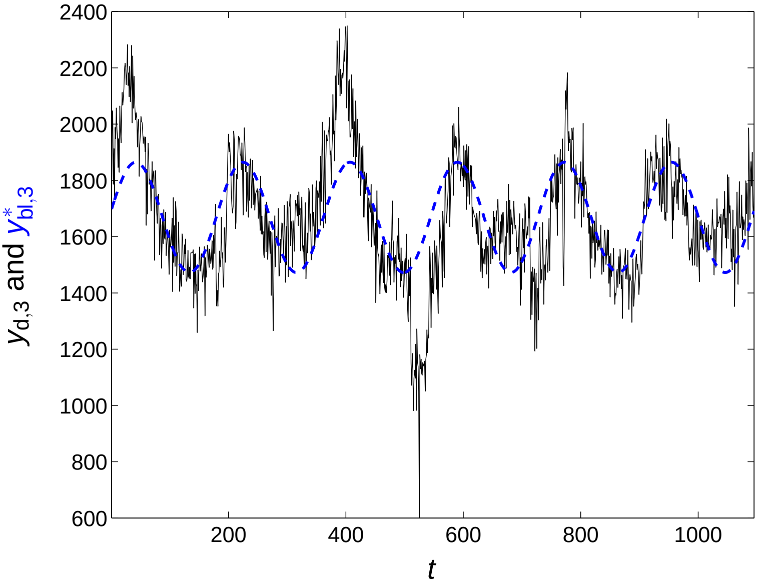
$$(y_{\mathrm{d},j} \text{ — } j\text{th row of } Y_{\mathrm{d}}, \quad y_{\mathrm{bl},j} \text{ — } j\text{th row of } Y_{\mathrm{bl}})$$

A special case of the line spectral estimation problem, for which solution subspace and maximum likelihood (ML) methods exist.

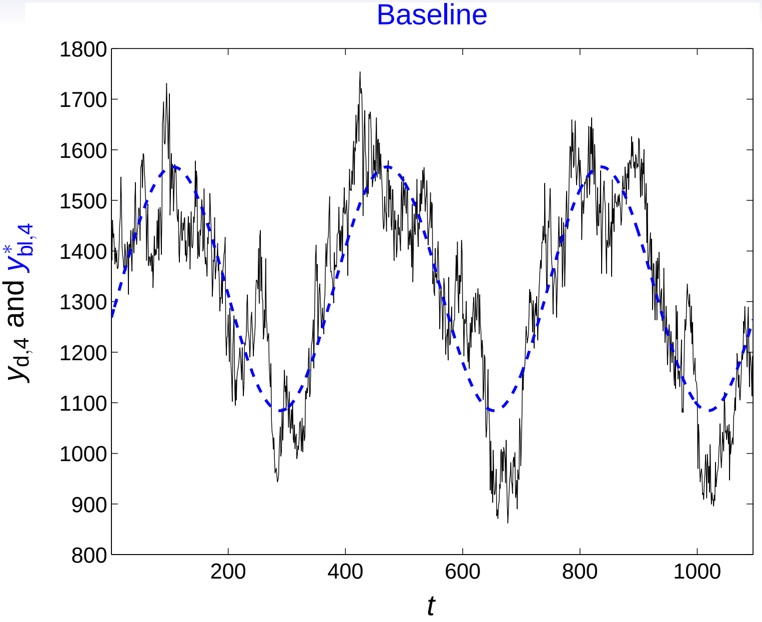
We use the ML approach, *i.e.*, local optimization, assuming $\omega_j = 12\pi/T$ (one year period) or $6\pi/T$ (half year period).

Furthermore, we eliminate the “linear” parameters b_j, c_j by projection $\quad \rightsquigarrow \quad$ VARPRO method

Baseline



2400
2200
2000
1800
1600
1400
1200
1000
800
600
200
400
600
800
1000
t
$y_{d,3}$ and $y^*_{bl,3}$


Baseline
1800
1700
1600
1500
1400
1300
1200
1100
1000
900
800
200
400
600
800
1000
t
$y_{d,4}$ and $y^*_{bl,4}$

Identification of the term involving the inputs

Problem:

$$\begin{array}{ll} \text{minimize} & \text{over } b_j, c_j, a_j \quad \|y_{\mathrm{d},j} - y_{\mathrm{bl},j}(b_j, c_j, \phi_j^*, \omega_j^*) - a_j^\top U_\mathrm{d}\|_2 \\ \text{subject to} & a_j \text{ has at most 50 nonzero elements} \end{array} \tag{2}$$

Proposed heuristic:

$$\begin{array}{ll} \text{minimize} & \text{over } b_j, c_j, a_j \quad \|y_{\mathrm{d},j} - y_{\mathrm{bl},j}(b_j, c_j, \phi_j^*, \omega_j^*) - a_j^\top U_\mathrm{d}\|_2 \\ \text{subject to} & \|a_j\|_1 \leq \gamma_j \end{array} \tag{3}$$

$\gamma_j > 0$ is parameter controlling the sparsity vs accuracy trade-off

Choice of the regularization parameter γ_j

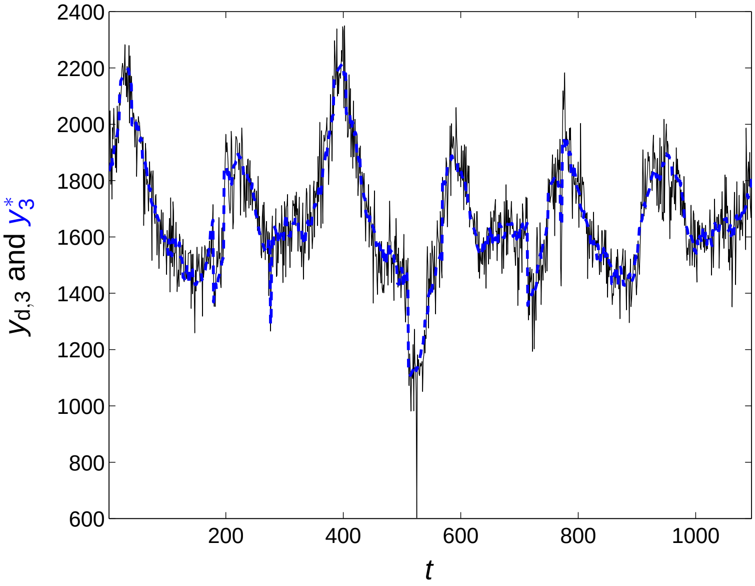
If we fix the nonzero elements to be the first 10 elements, the optimal solution (with this choice of the nonzero elements) is

$$a_j := \left[(y_{\mathrm{d}j} - y_{\mathrm{bl},j}) U_{\mathrm{d}}(1{:}10,:)^+ \quad 0_{1\times(m-10)} \right]$$

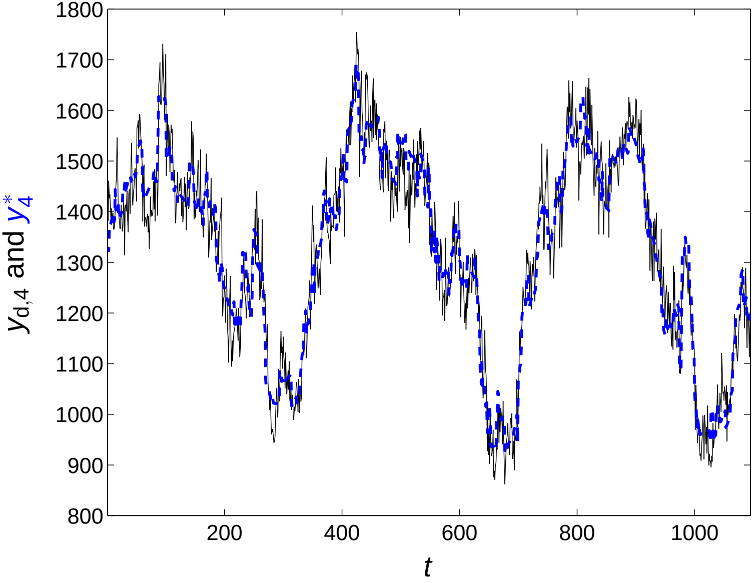
Let a^* be the optimal solution over all choices of the nonzero elements.

Since $\|a_j^*\|_1 = \gamma_j$, a heuristic choice for γ_j is $\gamma_j := \|a_j\|_1$.

Complete model (baseline and 24 inputs)



Complete model (baseline and 25 inputs)



Nonuniqueness of the solution

For uniqueness of A, we need U_{d} to be full row rank.

Special cases that lead to rank deficiency of U:

- Zero inputs can't affect the output. Removing them leads to an equivalent reduced model. For maximum sparsity, assign zero weights in A to those inputs.

- Inputs that are multiples of other inputs lead to essential nonuniqueness that can not be recovered by the sparsity.

Preprocessing step: remove redundant inputs.

Algorithm

1. Input: $U_\mathrm{d} \in \mathbb{R}^{m\times T}$ and $Y_\mathrm{d} \in \mathbb{R}^{p\times T}$.

2. Preprocessing: detect and remove redundant inputs.

3. For $j = 1$ to p

   3.1 Identify the baseline $\rightsquigarrow$ $(\omega_j^*, \phi_j^*, c_j^*, a_j^*)$

   3.2 Identify the I/O relation $\rightsquigarrow$ (b_j^*, c_j^*, a_j^*), sparsity pattern of a_j^*

   3.3 Solve (2) with fixed sparsity pattern, $\phi_j = \phi_j^*$ and $\omega_j = \omega_j^*$
   $\rightsquigarrow$ (b_j^*, c_j^*, a_j^*)

4. Postprocessing: add zero rows in A^* corresponding to the removed inputs

5. Output: $Y_\mathrm{bl}(b^*, c^*, \omega^*, \phi^*)$ and A^*

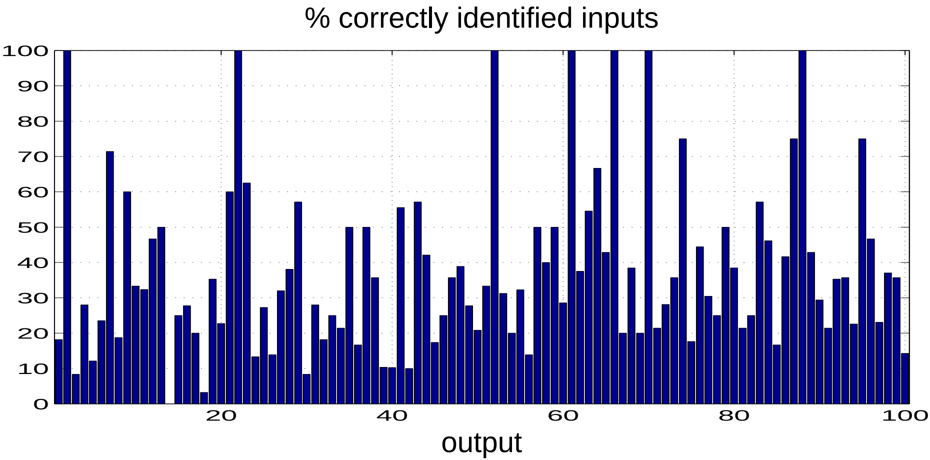
Identification of the baseline:

1. Let f', ϕ_j' be minimum value/point of (1) with $\omega_j = 6\pi/T$.
2. Let f'', ϕ_j'' be minimum value/point of (1) with $\omega_j = 12\pi/T$.
3. If $f' < f''$, $\omega_j^* := 6\pi/T$, $\phi_j^* := \phi_j'$, else $\omega_j^* := 12\pi/T$, $\phi_j^* := \phi_j''$.

Identification of the baseline:

1. Let $\gamma_j := \|(y_{\mathrm{d},j} - y_{\mathrm{bl},j})U_{\mathrm{d}}(1{:}10,:)^+\|_1$.
2. Let a_j' be solution to (3) with $\phi_j = \phi_j^*$, $\omega_j = \omega_j^*$.
3. Determine the sparsity pattern of a_j'.

Results on the PROMO challenge



Total: 2321 true inputs, 1796 identified inputs, of which 507 correct.

Code: `http://www.ecs.soton.ac.uk/~im/challenge.tar`