Direct identification of the linear block in Wiener system

Paweł Wachel & Grzegorz Mzyk

Wrocław University of Technology
Poland

24th Nov 2014

The problem

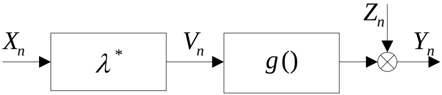

Figure: The Wiener system

$$V_n = \sum_{i=0}^{p} \lambda_i^* X_{n-i}, \quad (1)$$

$$Y_n = g(V_n) + Z_n, \quad (2)$$

Assumptions

A1 The input signal $\{X_n\}$ is an *i.i.d* sequence with zero mean and continuous density function $f(x)$, symmetric around the known point e.g. $x = 0$.

A2 The nonlinearity $g(\cdot)$ has bounded derivative. Furthermore $g(0) = 0$.

A3 The additive noise signal $\{Z_n\}$ is zero–mean *i.i.d.* sequence with finite variance. Both $\{X_n\}$ and $\{Z_n\}$ are independent random sequences.

The idea

$$Y_n = G\left(\mathbf{X}_n; \boldsymbol{\lambda}^*\right) + Z_n, \tag{3}$$

where

$$\begin{aligned} \mathbf{X}_n &= \left(X_n, X_{n-1}, \ldots, X_{n-p}\right)^T \\ \boldsymbol{\lambda}^* &= \left(\lambda_0^*, \lambda_1^*, \ldots, \lambda_p^*\right)^T \end{aligned}$$

and

$$\begin{aligned} G\left(\mathbf{x}; \boldsymbol{\lambda}^*\right) &= g\left(\boldsymbol{\lambda}^{*T}\mathbf{x}\right) \\ \mathbf{x} &\in R^{p+1} \end{aligned}$$

Least squares criterion

$$\mathcal{Q}(\boldsymbol{\lambda}) = E\left\{Y_i - G(\mathbf{X}_i; \boldsymbol{\lambda})\right\}^2 \rightarrow \min_{\boldsymbol{\lambda}}$$

but $G(\cdot)$ is unknown

The gradient of $G(\mathbf{x}; \boldsymbol{\lambda})$ (with respect to $\mathbf{x}$ and for arbitrary $\boldsymbol{\lambda} \in \mathbb{R}^{p+1}$) is proportional to $\boldsymbol{\lambda}$

$$\nabla_{\mathbf{x}} G(\mathbf{x}; \boldsymbol{\lambda}) = c_{\lambda}(\mathbf{x}; \boldsymbol{\lambda})\, \boldsymbol{\lambda}^T, \text{ where } c_{\lambda} = g'\left(\boldsymbol{\lambda}^T \mathbf{x}\right)$$

Taylor's expansion

$$G(\mathbf{X}_i; \boldsymbol{\lambda}) = G(\mathbf{x}; \boldsymbol{\lambda}) + c_{\lambda} \boldsymbol{\lambda}^T (\mathbf{X}_i - \mathbf{x}) + r_i, \tag{4}$$

Kernel least squares criterion (local linear approximation)

$$Q_{\mathbf{x}}(\boldsymbol{\lambda}) = E\left\{\left[Y_i - \left(G(\mathbf{x}; \boldsymbol{\lambda}) + \boldsymbol{\lambda}^T (\mathbf{X}_i - \mathbf{x})\right)\right]^2 K_h\left(\|\mathbf{X}_i - \mathbf{x}\|\right)\right\}, \tag{5}$$

Empirical criterion

For $\mathbf{x} = 0$

$$Q_{\mathbf{0}}(\lambda) = E\left\{\left[Y_i - \boldsymbol{\lambda}^T \mathbf{X}_i\right]^2 K_h\left(\|\mathbf{X}_i\|\right)\right\} \tag{6}$$

$$\arg\min_{\lambda} Q_{\mathbf{0}}(\lambda) = c_{\boldsymbol{\lambda}^*} \boldsymbol{\lambda}^*.$$

$$\hat{Q}_{\mathbf{0}}\left(\boldsymbol{\lambda}\right) = \frac{1}{N} \sum_{i=1}^{N} \left[Y_i - \boldsymbol{\lambda}^T \mathbf{X}_i\right]^2 K_h\left(\|\mathbf{X}_i\|\right),$$

$$\hat{\boldsymbol{\lambda}} = \mathrm{argmin}_{\boldsymbol{\lambda}}\, \hat{Q}\left(\boldsymbol{\lambda}\right)$$

The algorithm

$$\hat{\lambda} = \left[\frac{1}{N} \sum_{i=1}^{N} \mathbf{A}_i \right]^{-1} \left[\frac{1}{N} \sum_{i=1}^{N} \mathbf{B}_i \right], \tag{7}$$

where

$$\mathbf{A}_i = \mathbf{X}_i \mathbf{X}_i^T K \left(\| \mathbf{X}_i \| / h \right), \tag{8}$$

$$\mathbf{B}_i = Y_i \mathbf{X}_i K \left(\| \mathbf{X}_i \| / h \right). \tag{9}$$

The consistency

Theorem

Let assumptions **A1 – A3** *be in force. Then, for* $h = N^{-\alpha}$*, where* $\alpha \in (0, \frac{1}{d})$ *and* $d = p + 1$*, the estimate* $\hat{\lambda}$ *is consistent estimate of the scaled impulse response of the dynamic subsystem, i.e.*

$$\hat{\lambda} \rightarrow c\lambda^* \text{ in probability} \tag{10}$$

as $N \rightarrow \infty$*, where* $c = c_{\lambda^*}$ *is a multiplicative constant.*

Generalization

Remark

Although the algorithm was derived and analyzed for the input symmetrically distributed around the point $\mathbf{x} = 0$*, it can be generalized for much wider class of densities. Furthermore, for each pair* $(\mathbf{X}_i, Y_i)$ *and* $(\mathbf{X}_j, Y_j)$*,* $i, j = 1, 2, ..., N$*, we can expand* $G(\mathbf{x}; \boldsymbol{\lambda})$ *around the point* $\mathbf{x} = \mathbf{X}_j$ *(provided that* $G(\mathbf{x}; \boldsymbol{\lambda})$ *is differentiable), which leads to*

$$Y_i = Y_j + c_{ij}\boldsymbol{\lambda}^* (\mathbf{X}_i - \mathbf{X}_j) + r_{ij} + Z_i - Z_j,$$

Introducing $Y_{ij} = Y_i - Y_j$*,* $\mathbf{X}_{ij} = \mathbf{X}_i - \mathbf{X}_j$*, and* $Z_{ij} = Z_i - Z_j$*, we obtain that*

$$Y_{ij} = c_{ij}\boldsymbol{\lambda}^* \mathbf{X}_{ij} + r_{ij} + Z_{ij},$$

where $c_{ij} \to \mathrm{const}$*, as* $\|\mathbf{X}_{ij}\| \to 0$*, and the components of* $\mathbf{X}_{ij}$ *are obviously symmetrically distributed.*

Practical aspects (1)

Since the probability of the selection event $P\left(\|\mathbf{X}_{ij}\| < h\right) \sim h^d$ decreases very fast as $h \to 0$ (particularly for large dimension d of the vector $\mathbf{X}_{ij}$) we propose to repeat estimation around all points $\mathbf{x} = \mathbf{X}_j$, $j = 1, 2, \ldots, N$, *i.e.*, to compute

$$\hat{\lambda}_j = \left[\frac{1}{N}\sum_{i=1}^{N}\mathbf{A}_{ij}\right]^{-1}\left[\frac{1}{N}\sum_{i=1}^{N}\mathbf{B}_{ij}\right], \tag{11}$$

where

$$\mathbf{A}_{ij} = \mathbf{X}_{ij}\mathbf{X}_{ij}^T K\left(\|\mathbf{X}_{ij}\| / h\right), \tag{12}$$

$$\mathbf{B}_{ij} = Y_{ij}\mathbf{X}_{ij} K\left(\|\mathbf{X}_{ij}\| / h\right). \tag{13}$$

Practical aspects (2)

This provides the series of estimates $\left\{\hat{\lambda}_j\right\}_{j=1}^{N}$, that can be next averaged as follows

$$\widetilde{\lambda} = \frac{\sum_{j=1}^{N} w_j \hat{\lambda}_j}{\sum_{j=1}^{N} w_j}, \tag{14}$$

where w_j's are properly selected weighting factors.
We propose

$$w_j = \operatorname{sgn} \widehat{\lambda}_j[1] \cdot I(j), \tag{15}$$

where

$$I(j) = \begin{cases} 1, \text{ as } \left\| \widehat{\lambda}_j \right\| \geqslant r \\ 0, \text{ as } \left\| \widehat{\lambda}_j \right\| < r \end{cases}$$

Simulation example

$g(V) = \sin(V), \qquad \lambda^* = [2, 1]^T, \qquad X_n \sim U[-10, 10],$
$Z_n \sim U[-0.1, 0.1]$
The identification experiment was repeated for $N = 1000$, $h = 1$ and various values of the threshold parameter r

$$r(\alpha) = \left\| \boldsymbol{\lambda}_{[\alpha]} \right\|,$$

where $\left\| \boldsymbol{\lambda}_{[\alpha]} \right\|$ denotes a sample α-th order quantile of the random variable $\left\| \widehat{\boldsymbol{\lambda}}_j \right\|$. The following estimation error was computed in $R = 10$ repeats

$$e = \frac{1}{R} \sum_{k=1}^{R} \left\| c_k \widetilde{\boldsymbol{\lambda}}^{(k)} - \boldsymbol{\lambda}^* \right\|,$$

with normalization factor $c_k = \boldsymbol{\lambda}^*[1] / \widetilde{\boldsymbol{\lambda}}^{(k)}[1]$

Simulation example (cont.)

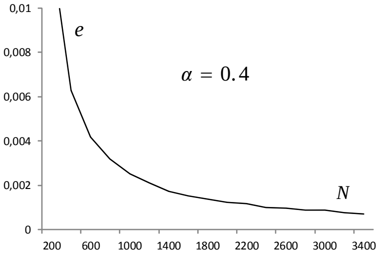

Figure: Estimation error e vs. number of data N and selection parameter α

Pros and Cons

+ weak assumptions, nonparametric nonlinearity, differentiable in the selected points
+ simple algorithm, no nonlinear optimization (only linear least squares + kernel selection)
+ direct recovery of the impulse response (and intraction signal), decomposition of the Wiener system

– slow convergence, ($p + 1$ dimensional space)
– problems with generatization for L-N-L system
– FIR