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

\[ V_n = \sum_{i=0}^{p} \lambda_i^* X_{n-i}, \]  

(1)

\[ Y_n = g(V_n) + Z_n, \]  

(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

\[ Y_n = G(X_n; \lambda^*) + Z_n, \quad (3) \]

where

\[ X_n = (X_n, X_{n-1}, \ldots, X_{n-p})^T \]

\[ \lambda^* = (\lambda_0^*, \lambda_1^*, \ldots, \lambda_p^*)^T \]

and

\[ G(x; \lambda^*) = g\left(\lambda^T x\right) \]

\[ x \in \mathbb{R}^{p+1} \]
Least squares criterion

\[
Q(\lambda) = E \{ Y_i - G(X_i; \lambda) \}^2 \rightarrow \min_{\lambda}
\]

but \( G(\cdot) \) is unknown

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

\[
\nabla_x G(x; \lambda) = c_\lambda(x; \lambda) \lambda^T, \text{ where } c_\lambda = g'(\lambda^T x)
\]

Taylor’s expansion

\[
G(X_i; \lambda) = G(x; \lambda) + c_\lambda \lambda^T (X_i - x) + r_i,
\]

Kernel least squares criterion (local linear approximation)

\[
Q_x(\lambda) = E \left\{ \left[ Y_i - \left( G(x; \lambda) + \lambda^T (X_i - x) \right) \right]^2 K_h(\|X_i - x\|) \right\},
\]
Empirical criterion

For $x = 0$

$$Q_0(\lambda) = E \left\{ \left[ Y_i - \lambda^T X_i \right]^2 K_h(\|X_i\|) \right\}$$

arg min$_{\lambda} Q_0(\lambda) = c_{\lambda^*} \lambda^*.$

$$\hat{Q}_0(\lambda) = \frac{1}{N} \sum_{i=1}^{N} \left[ Y_i - \lambda^T X_i \right]^2 K_h(\|X_i\|),$$

$$\hat{\lambda} = \arg\min_{\lambda} \hat{Q}(\lambda)$$
The algorithm

\[ \hat{\lambda} = \left[ \frac{1}{N} \sum_{i=1}^{N} A_i \right]^{-1} \left[ \frac{1}{N} \sum_{i=1}^{N} B_i \right], \quad (7) \]

where

\[ A_i = X_i X_i^T K \left( \| X_i \| / h \right), \quad (8) \]
\[ B_i = Y_i X_i K \left( \| X_i \| / h \right). \quad (9) \]
Let assumptions $\textbf{A1} - \textbf{A3}$ 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.

$$\hat{\lambda} \to c\lambda^* \text{ in probability} \quad (10)$$

as $N \to \infty$, where $c = c_{\lambda^*}$ is a multiplicative constant.
Although the algorithm was derived and analyzed for the input symmetrically distributed around the point $x = 0$, it can be generalized for much wider class of densities. Furthermore, for each pair $(X_i, Y_i)$ and $(X_j, Y_j)$, $i, j = 1, 2, ..., N$, we can expand $G(x; \lambda)$ around the point $x = X_j$ (provided that $G(x; \lambda)$ is differentiable), which leads to

$$Y_i = Y_j + c_{ij} \lambda^* (X_i - X_j) + r_{ij} + Z_i - Z_j,$$

Introducing $Y_{ij} = Y_i - Y_j$, $X_{ij} = X_i - X_j$, and $Z_{ij} = Z_i - Z_j$, we obtain that

$$Y_{ij} = c_{ij} \lambda^* X_{ij} + r_{ij} + Z_{ij},$$

where $c_{ij} \to \text{const}$, as $\|X_{ij}\| \to 0$, and the components of $X_{ij}$ are obviously symmetrically distributed.
Since the probability of the selection event \( P (\|\mathbf{X}_{ij}\| < h) \sim h^d \) 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

\[
\hat{\lambda}_j = \left[ \frac{1}{N} \sum_{i=1}^{N} A_{ij} \right]^{-1} \left[ \frac{1}{N} \sum_{i=1}^{N} B_{ij} \right], \tag{11}
\]

where

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

\[
B_{ij} = \mathbf{Y}_{ij} \mathbf{X}_{ij} K \left( \|\mathbf{X}_{ij}\| / h \right). \tag{13}
\]
Practical aspects (2)

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

\[
\tilde{\lambda} = \frac{\sum_{j=1}^{N} w_j \hat{\lambda}_j}{\sum_{j=1}^{N} w_j},
\]

(14)

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

\[
w_j = \text{sgn} \left( \hat{\lambda}_j [1] \cdot l(j) \right),
\]

(15)

where

\[
l(j) = \begin{cases} 
1, & \text{as } \| \hat{\lambda}_j \| \geq r \\
0, & \text{as } \| \hat{\lambda}_j \| < r 
\end{cases}
\]
Simulation example

\[ g(V) = \sin(V), \quad \lambda^* = [2, 1]^T, \quad X_n \sim U[-10, 10], \]
\[ Z_n \sim U[-0.1, 0.1] \]

The identification experiment was repeated for \( N = 1000, \ h = 1 \)
and various values of the threshold parameter \( r \)

\[ r(\alpha) = \left\| \lambda_{[\alpha]} \right\|, \]

where \( \left\| \lambda_{[\alpha]} \right\| \) denotes a sample \( \alpha \)-th order quantile of the random variable \( \left\| \tilde{\lambda}_j \right\| \). The following estimation error was computed in \( R = 10 \) repeats

\[ e = \frac{1}{R} \sum_{k=1}^{R} \left\| c_k \tilde{\lambda}^{(k)} - \lambda^* \right\|, \]

with normalization factor \( c_k = \lambda^*[1] / \tilde{\lambda}^{(k)}[1] \)
Simulation example (cont.)

Figure: Estimation error $e$ vs. number of data $N$ and selection parameter $\alpha$
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 interaction signal), decomposition of the Wiener system

− slow convergence, \((p + 1)\) dimensional space
− problems with generalization for L-N-L system
− FIR