Abstract

Least-Squares Support Vector Machines (LS-SVM’s), originating from Stochastic Learning theory, represent a promising approach to identify nonlinear systems via nonparametric estimation of nonlinearities in a computationally and stochastically attractive way. However, application of LS-SVM’s in the identification context is formulated as a linear regression aiming at the minimization of the $\ell_2$ loss in terms of the prediction error. This formulation corresponds to a prejudice of an auto-regressive noise structure, which, especially in the nonlinear context, is often found to be too restrictive in practical applications. In [1], a novel Instrumental Variable (IV) based estimation is integrated into the LS-SVM approach providing, under minor conditions, a consistent identification of nonlinear systems in case of a noise modeling error. It is shown how the cost function of the LS-SVM is modified to achieve an IV-based solution.

In this technical report, a detailed derivation of the results presented in Section 5.2 of [1] is given as a supplement material for interested readers.

1 IV in the dual form

Consider the primal minimization problem (eq. (52) in [1]):

$$\begin{align*}
\min_{\theta \in \mathbb{R}^n} & \quad \frac{1}{2} \theta^T \theta + \frac{\gamma}{2N^2} \|E^T E\|_{\ell_2}^2, \\
\text{s.t.} & \quad e(k) = y(k) - \varphi^T(k) \theta, \quad k = 1, \ldots, N, \\
& \quad \phi_i^T (0) \theta = 0, \quad i = 1, \ldots, n_g.
\end{align*}$$

(1a)

(1b)

(1c)
Introduce the Lagrangian
\[ \mathcal{L}(\theta, e, \alpha, \beta) = \frac{1}{2} \theta^\top \theta + \frac{\gamma}{2N^2} \| \Gamma^\top E \|_2^2 - \sum_{k=1}^{N} \alpha_k (\varphi^\top(k)\theta + e(k) - y(k)) - \sum_{i=1}^{n_g} \beta_i \phi_i^\top(0)\theta_i, \]
with \( \alpha_k \) and \( \beta_i \) being the Lagrangian multiplier. According to [1], the terms \( \varphi^\top(k) \) and \( \theta \) can be decomposed as
\[
\varphi(k) = \begin{bmatrix} 1 & \phi_1^\top(y(k - 1)) & \ldots & \phi_{n_a}^\top(y(k - n_a)) & \phi_{n_a+1}^\top(u(k)) & \ldots & \phi_{n_g}^\top(u(k - n_b)) \end{bmatrix}^\top, \tag{3a}
\]
\[
\theta = \begin{bmatrix} c & \theta_1^\top & \ldots & \theta_{n_g}^\top \end{bmatrix}^\top, \tag{3b}
\]
where \( \phi_i(\cdot) = \begin{bmatrix} \phi_{i,1}(\cdot) & \ldots & \phi_{i,n_H}(\cdot) \end{bmatrix}^\top, \theta_i = \begin{bmatrix} \theta_{i,1} & \ldots & \theta_{i,n_H} \end{bmatrix}^\top \) and \( c \in \mathbb{R}. \)

The global optimum of Problem (1) is obtained when the KKT conditions are fulfilled, i.e.,
\[
\frac{\partial \mathcal{L}}{\partial e} = 0 \rightarrow \quad \alpha_k = \frac{\gamma}{N^2} \Gamma^\top e(k), \tag{4a}
\]
\[
\frac{\partial \mathcal{L}}{\partial \alpha_k} = 0 \rightarrow \quad y(k) = \sum_{i=1}^{n_g} \phi_i^\top(x_i(k))\theta_i + c + e(k), \tag{4b}
\]
\[
\frac{\partial \mathcal{L}}{\partial \beta_i} = 0 \rightarrow \quad 0 = \phi_i^\top(0)\theta_i, \tag{4c}
\]
\[
\frac{\partial \mathcal{L}}{\partial \theta_i} = 0 \rightarrow \quad \theta_i = \sum_{k=1}^{N} \alpha_k \phi_i(x_i(k)) + \beta_i \phi_i(0), \tag{4d}
\]
\[
\frac{\partial \mathcal{L}}{\partial c} = 0 \rightarrow \quad c = \sum_{k=1}^{N} \alpha_k, \tag{4e}
\]
for all \( i = 1, \ldots, n_g \) and \( k = 1, \ldots, N. \)

By substituting (4d) and (4e) into (4b) and (4c), we get
\[
y(k) = \sum_{i=1}^{n_g} \phi_i^\top(x_i(k)) \left( \sum_{k=1}^{N} \alpha_k \phi_i(x_i(k)) + \beta_i \phi_i(0) \right) + \sum_{k=1}^{N} \alpha_k + e(k), \tag{5a}
\]
\[
0 = \phi_i^\top(0) \left( \sum_{k=1}^{N} \alpha_k \phi_i(x_i(k)) + \beta_i \phi_i(0) \right) \left( \sum_{\theta_i} \right), \tag{5b}
\]
for \( k \in \{1, \ldots, N\} \) and \( i \in \{1, \ldots, n_g\} \). Let introduce the following notation (used in [1]):

\[
E = [e(1) \ldots e(N)]^T, \tag{6a}
\]

\[
Y = [y(1) \ldots y(N)]^T, \tag{6b}
\]

\[
\alpha = [\alpha_1 \ldots \alpha_N]^T, \tag{6c}
\]

\[
\beta = [\beta_1 \ldots \beta_{n_g}]^T, \tag{6d}
\]

\[
1_N = [1 \ldots 1]^T \in \mathbb{R}^N, \tag{6e}
\]

\[
0_{n_g} = [0 \ldots 0]^T \in \mathbb{R}^{n_g}, \tag{6f}
\]

\[
\Phi_i = [\phi_i(x_i(1)) \ldots \phi_i(x_i(N))]^T, \tag{6g}
\]

\[
D_\Phi = [\Phi_1 \Phi_0(0) \ldots \Phi_{n_g} \phi_{n_g}(0)]^T, \tag{6h}
\]

\[
D_0 = \text{diag}(\phi_1(0) \phi_1(0), \ldots, \phi_{n_g}(0) \phi_{n_g}(0)). \tag{6i}
\]

Eqs. (5) can also be written in the matrix form

\[
E = Y - \left(1_N 1_N^T + \sum_{i=1}^{n_g} \Phi_i \Phi_i^T\right)\alpha - D_\Phi \beta, \tag{7a}
\]

\[
0_{n_g} = D_\Phi^T \alpha + D_0 \beta. \tag{7b}
\]

Then substitution of (7a) into (4a) leads to the solution:

\[
\begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \left[ H G + \frac{1}{\gamma} I_N \right]^{-1} \left[ -H Y \right], \tag{8}
\]

where \( H = \Gamma \Gamma^T \) and \( G = 1_N 1_N^T + \sum_{i=1}^{n_g} \Phi_i \Phi_i^T \). Note that the \((i, j)\)-th entry of the matrix \( G^{(i)} \)

is given by

\[
[G^{(i)}]_{j,k} = \langle \phi_i(x_i(j)), \phi_i(x_i(k)) \rangle = K^{(i)}(x_i(j), x_i(k)), \tag{9}
\]

with \( K^{(i)}(x_i(j), x_i(k)) \) being a positive definite kernel function defining the inner product \( \langle \phi_i(x_i(j)), \phi_i(x_i(k)) \rangle \). Similarly, the entries of the matrices \( D_\Phi \) and \( D_0 \) can be defined in terms of a kernel function as

\[
[D_\Phi]_{i,k} = \langle \phi_i(x_i(k)), \phi_i(0) \rangle = K_{\Phi,0}^{(i)}(x_i(k), 0), \tag{10}
\]

\[
[D_0]_{i,i} = \langle \phi_i(0), \phi_i(0) \rangle = K_{0,0}^{(i)}(0, 0). \tag{11}
\]

Once the Lagrangian multipliers \( \alpha \) and \( \beta \) are computed through (8), the estimate \( \hat{\theta} \) of the model parameters \( \theta \) is obtained from (4d) and (4e), i.e.,

\[
\hat{\theta}_D = \begin{bmatrix} c \\ \theta_1 \\ \vdots \\ \theta_{n_g} \end{bmatrix} = \begin{bmatrix} 1_N^T \alpha \\ \Phi_1^T \alpha + \beta_1 \phi_1(0) \\ \vdots \\ \Phi_{n_g}^T \alpha + \beta_{n_g} \phi_{n_g}(0) \end{bmatrix}. \tag{12}
\]
The estimate of the nonlinear functions $\phi_i^\top(\cdot)\theta_i$ can be then obtained from (12) and (4d), i.e.,

$$\phi_i^\top(\cdot)\theta_i = \phi_i^\top(\cdot) \left( \phi_i(0) \beta_i + \sum_{k=1}^{N} \alpha_k \phi_i(x_i(k)) \right) = \phi_i^\top(0)^\top \beta_i + \sum_{k=1}^{N} \alpha_k \phi_i^\top(\cdot) \phi_i(x_i(k)) = K^{(i)}(0, \cdot) \beta_i + \sum_{k=1}^{N} \alpha_k K^{(i)}(x_i(k), \cdot). \quad (13a)$$

$$= \phi_i^\top(\cdot) \phi_i(0) \beta_i + \sum_{k=1}^{N} \alpha_k \phi_i^\top(\cdot) \phi_i(x_i(k)) = K^{(i)}(0, \cdot) \beta_i + \sum_{k=1}^{N} \alpha_k K^{(i)}(x_i(k), \cdot) = K^{(i)}(x_i(0), \cdot) \beta_i + \sum_{k=1}^{N} \alpha_k K^{(i)}(x_i(k), \cdot). \quad (13b)$$

$$= \phi_i^\top(\cdot) \phi_i(0) \beta_i + \sum_{k=1}^{N} \alpha_k \phi_i^\top(\cdot) \phi_i(x_i(k)) = K^{(i)}(0, \cdot) \beta_i + \sum_{k=1}^{N} \alpha_k K^{(i)}(x_i(k), \cdot). \quad (13c)$$

References