An Instrumental Least Squares Support Vector Machine for Nonlinear System Identification: enforcing zero-centering constraints

TUE-CS-2013-001

V. Laurain and R.Tóth and D. Piga

April 18, 2013

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 ℓ_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]):

$$\min_{\theta \in \mathbb{R}^{n_\theta}} \quad \frac{1}{2}\theta^\top\theta + \frac{\gamma}{2N^2}\left\|\Gamma^\top E\right\|_{\ell_2}^2, \tag{1a}$$

$$\text{s.t.} \quad e(k) = y(k) - \varphi^\top(k)\theta, \qquad k = 1, \dots, N, \tag{1b}$$

$$\phi_i^\top(0)\theta_i = 0, \qquad i = 1, \dots, n_{\mathrm{g}}. \tag{1c}$$

Introduce the *Lagrangian*

$$\mathcal{L}(\theta, e, \alpha, \beta) = \frac{1}{2}\theta^\top\theta + \frac{\gamma}{2N^2}\left\|\Gamma^\top E\right\|_{\ell_2}^2 - \sum_{k=1}^{N}\alpha_k\big(\varphi^\top(k)\theta + e(k) - y(k)\big) - \sum_{i=1}^{n_\mathrm{g}}\beta_i\phi_i^\top(0)\theta_i, \quad (2)$$

with α_k and β_i being the *Lagrangian multiplier*. According to [1], the terms $\varphi^\top(k)$ and θ can be decomposed as

$$\varphi(k) = \begin{bmatrix} 1 & \phi_1^\top\big(y(k-1)\big) & \dots & \phi_{n_\mathrm{a}}^\top\big(y(k-n_\mathrm{a})\big) & \phi_{n_\mathrm{a}+1}^\top\big(u(k)\big) & \dots & \phi_{n_\mathrm{g}}^\top\big(u(k-n_\mathrm{b})\big) \end{bmatrix}^\top, \quad (3a)$$

$$\theta = \begin{bmatrix} c & \theta_1^\top & \dots & \theta_{n_\mathrm{g}}^\top \end{bmatrix}^\top, \quad (3b)$$

where $\phi_i(\cdot) = [\ \phi_{i,1}(\cdot) \ \ \dots \ \ \phi_{i,n_\mathrm{H}}(\cdot)\ ]^\top$, $\theta_i = [\theta_{i,1} \ \dots \ \theta_{i,n_\mathrm{H}}]^\top$ and $c \in \mathbb{R}$.

The global optimum of Problem (1) is obtained when the KKT conditions are fulfilled, i.e.,

$$\frac{\partial\mathcal{L}}{\partial e} = 0 \rightarrow \qquad \alpha_k = \frac{\gamma}{N^2}\Gamma\Gamma^\top e(k), \quad (4a)$$

$$\frac{\partial\mathcal{L}}{\partial \alpha_k} = 0 \rightarrow \qquad y(k) = \underbrace{\sum_{i=1}^{n_\mathrm{g}}\phi_i^\top(x_i(k))\theta_i + c}_{\varphi^\top(k)\theta} + e(k), \quad (4b)$$

$$\frac{\partial\mathcal{L}}{\partial \beta_i} = 0 \rightarrow \qquad 0 = \phi_i^\top(0)\theta_i, \quad (4c)$$

$$\frac{\partial\mathcal{L}}{\partial \theta_i} = 0 \rightarrow \qquad \theta_i = \sum_{k=1}^{N}\alpha_k\phi_i(x_i(k)) + \beta_i\phi_i(0), \quad (4d)$$

$$\frac{\partial\mathcal{L}}{\partial c} = 0 \rightarrow \qquad c = \sum_{k=1}^{N}\alpha_k, \quad (4e)$$

for all $i = 1, \dots, n_\mathrm{g}$ and $k = 1, \dots, N$.

By substituting (4d) and (4e) into (4b) and (4c), we get

$$y(k) = \sum_{i=1}^{n_\mathrm{g}}\phi_i^\top(x_i(k))\underbrace{\left(\sum_{k=1}^{N}\alpha_k\phi_i(x_i(k)) + \beta_i\phi_i(0)\right)}_{\theta_i} + \underbrace{\sum_{k=1}^{N}\alpha_k}_{c} + e(k), \quad (5a)$$

$$0 = \phi_i^\top(0)\underbrace{\left(\sum_{k=1}^{N}\alpha_k\phi_i(x_i(k)) + \beta_i\phi_i(0)\right)}_{\theta_i}, \quad (5b)$$

for $k \in \{1, \dots, N\}$ and $i \in \{1, \dots, n_g\}$. Let introduce the following notation (used in [1]):

$$E = [e(1) \quad \dots \quad e(N)]^\top, \tag{6a}$$
$$Y = [y(1) \quad \dots \quad y(N)]^\top, \tag{6b}$$
$$\alpha = [\alpha_1 \quad \dots \quad \alpha_N]^\top, \tag{6c}$$
$$\beta = [\beta_1 \quad \dots \quad \beta_{n_g}]^\top, \tag{6d}$$
$$1_N = [1 \quad \dots \quad 1]^\top \in \mathbb{R}^N, \tag{6e}$$
$$0_{n_g} = [0 \quad \dots \quad 0]^\top \in \mathbb{R}^{n_g}, \tag{6f}$$
$$\Phi_i = [\ \phi_i(x_i(1)) \quad \dots \quad \phi_i(x_i(N))\ ]^\top, \tag{6g}$$
$$D_\Phi = [\ \Phi_1\phi_1(0) \quad \dots \quad \Phi_{n_g}\phi_{n_g}(0)\ ]^\top, \tag{6h}$$
$$D_0 = \mathrm{diag}\left(\phi_1^\top(0)\phi_1(0),\ \dots,\ \phi_{n_g}^\top(0)\phi_{n_g}(0)\right). \tag{6i}$$

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

$$E = Y - \left(1_N 1_N^\top + \sum_{i=1}^{n_g} \Phi_i \Phi_i^\top\right)\alpha - D_\Phi \beta, \tag{7a}$$
$$0_{n_g} = D_\Phi^\top \alpha + D_0 \beta. \tag{7b}$$

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

$$\begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} \frac{1}{N^2}HG + \frac{1}{\gamma}I_N & \frac{1}{N^2}HD_\Phi \\ \frac{1}{N}D_\Phi^\top & \frac{1}{N}D_0 \end{bmatrix}^{-1} \begin{bmatrix} \frac{1}{N^2}HY \\ 0_{n_g} \end{bmatrix}, \tag{8}$$

where $H = \Gamma\Gamma^\top$ and $G = 1_N 1_N^\top + \sum_{i=1}^{n_g} \underbrace{\Phi\Phi_i^\top}_{G^{(i)}}$. Note that the (i,j)-th entry of the matrix $G^{(i)}$ is given by

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

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

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

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

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

$$\hat{\theta}_D = \begin{bmatrix} c \\ \theta_1 \\ \vdots \\ \theta_{n_g} \end{bmatrix} = \begin{bmatrix} 1_N^\top \alpha \\ \Phi_i^\top \alpha + \beta_1 \phi_1(0) \\ \vdots \\ \Phi_{n_g}^\top \alpha + \beta_{n_g} \phi_{n_g}(0) \end{bmatrix}. \tag{12}$$

The estimate of the nonlinear functions $\phi_i^\top(\centerdot)\theta_i$ can be then obtained from (12) and (4d), i.e.,

$$\phi_i^\top(\centerdot)\theta_i = \phi_i^\top(\centerdot)\left(\phi_i(0)\beta_i + \sum_{k=1}^{N}\alpha_k\phi_i(x_i(k))\right) = \tag{13a}$$

$$= \underbrace{\phi_i^\top(\centerdot)\phi_i(0)}_{K^{(i)}\left(0,\centerdot\right)}\beta_i + \sum_{k=1}^{N}\alpha_k\underbrace{\phi_i^\top(\centerdot)\phi_i(x_i(k))}_{K^{(i)}\left(x_i(k),\centerdot\right)} = \tag{13b}$$

$$= K^{(i)}\big(0,\centerdot\big)\beta_i + \sum_{k=1}^{N}\alpha_k K^{(i)}\big(x_i(k),\centerdot\big). \tag{13c}$$

References

[1] V. Laurain, R. Tóth, D. Piga, and W. X. Zheng. An instrumental least squares support vector machine for nonlinear system identification. *Submitted to Automatica*, 2013.