

# 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  $\ell_2$  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]):

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

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

$$\phi_i^\top(0)\theta_i = 0, \quad i = 1, \dots, n_g. \quad (1c)$$

Introduce the *Lagrangian*

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

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

$$\varphi(k) = [ 1 \quad \phi_1^\top(y(k-1)) \quad \dots \quad \phi_{n_a}^\top(y(k-n_a)) \quad \phi_{n_a+1}^\top(u(k)) \quad \dots \quad \phi_{n_g}^\top(u(k-n_b)) ]^\top, \quad (3a)$$

$$\theta = [ c \quad \theta_1^\top \quad \dots \quad \theta_{n_g}^\top ]^\top, \quad (3b)$$

where  $\phi_i(\bullet) = [ \phi_{i,1}(\bullet) \quad \dots \quad \phi_{i,n_H}(\bullet) ]^\top$ ,  $\theta_i = [\theta_{i,1} \quad \dots \quad \theta_{i,n_H}]^\top$  and  $c \in \mathbb{R}$ .

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

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

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

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

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

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

for all  $i = 1, \dots, n_g$  and  $k = 1, \dots, N$ .

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

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

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

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

$$E = [e(1) \ \dots \ e(N)]^\top, \quad (6a)$$

$$Y = [y(1) \ \dots \ y(N)]^\top, \quad (6b)$$

$$\alpha = [\alpha_1 \ \dots \ \alpha_N]^\top, \quad (6c)$$

$$\beta = [\beta_1 \ \dots \ \beta_{n_g}]^\top, \quad (6d)$$

$$1_N = [1 \ \dots \ 1]^\top \in \mathbb{R}^N, \quad (6e)$$

$$0_{n_g} = [0 \ \dots \ 0]^\top \in \mathbb{R}^{n_g}, \quad (6f)$$

$$\Phi_i = [\phi_i(x_i(1)) \ \dots \ \phi_i(x_i(N))]^\top, \quad (6g)$$

$$D_\Phi = [\Phi_1 \phi_1(0) \ \dots \ \Phi_{n_g} \phi_{n_g}(0)]^\top, \quad (6h)$$

$$D_0 = \text{diag} \left( \phi_1^\top(0) \phi_1(0), \dots, \phi_{n_g}^\top(0) \phi_{n_g}(0) \right). \quad (6i)$$

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

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

$$0_{n_g} = D_\Phi^\top \alpha + D_0 \beta. \quad (7b)$$

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

$$\begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} \frac{1}{N^2} H G + \frac{1}{\gamma} I_N & \frac{1}{N^2} H D_\Phi \\ \frac{1}{N} D_\Phi^\top & \frac{1}{N} D_0 \end{bmatrix}^{-1} \begin{bmatrix} \frac{1}{N^2} H Y \\ 0_{n_g} \end{bmatrix}, \quad (8)$$

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

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

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

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

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

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.,

$$\hat{\theta}_D = \begin{bmatrix} c \\ \theta_1 \\ \vdots \\ \theta_{n_g} \end{bmatrix} = \begin{bmatrix} 1_N^\top \alpha \\ \Phi_1^\top \alpha + \beta_1 \phi_1(0) \\ \vdots \\ \Phi_{n_g}^\top \alpha + \beta_{n_g} \phi_{n_g}(0) \end{bmatrix}. \quad (12)$$

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

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

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

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

## 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.