Journal of Zhejiang University SCIENCE A
ISSN 1009-3095
http://www.zju.edu.cn/jzus
E-mail: jzus@zju.edu.cn

# Nonlinear decoupling controller design based on least squares support vector regression*

WEN Xiang-jun (文香军)[1], ZHANG Yu-nong (张雨浓)[2], YAN Wei-wu (阎威武)[1], XU Xiao-ming (许晓鸣)[1]

([1]Department of Automatic Control, Shanghai Jiao Tong University, Shanghai 200030, China)

([2]Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow G1 1QE, UK)

E-mail: wenxiangjun@sjtu.edu.cn; ynzhang@ieee.org; yanwwsjtu@sjtu.edu.cn; xmxu@sjtu.edu.cn

Received Apr. 6, 2005; revision accepted Sept. 1, 2005

**Abstract:** Support Vector Machines (SVMs) have been widely used in pattern recognition and have also drawn considerable interest in control areas. Based on a method of least squares SVM (LS-SVM) for multivariate function estimation, a generalized inverse system is developed for the linearization and decoupling control of a general nonlinear continuous system. The approach of inverse modelling via LS-SVM and parameters optimization using the Bayesian evidence framework is discussed in detail. In this paper, complex high-order nonlinear system is decoupled into a number of pseudo-linear Single Input Single Output (SISO) subsystems with linear dynamic components. The poles of pseudo-linear subsystems can be configured to desired positions. The proposed method provides an effective alternative to the controller design of plants whose accurate mathematical model is unknown or state variables are difficult or impossible to measure. Simulation results showed the efficacy of the method.

**Key words:** Support Vector Machine (SVM), Decoupling control, Nonlinear system, Generalized inverse system

**doi:**10.1631/jzus.2006.A0275 **Document code:** A **CLC number:** TP273

## INTRODUCTION

Most practical systems are multivariate nonlinear systems. In general, the MIMO (multiple inputs and multiple outputs) systems are coupled. This coupling affects the effectiveness of a specific loop controller on the corresponding output, and in some case, may become serious and cause many difficulties to the control system design. How to decouple the multivariate systems and design practical controllers is one of the major issues in nonlinear control area.

In recent years, various linearization and decoupling methods have been presented to handle this problem. Some methods are based on differential geometry method, which solves a group of differential equations and linearize nonlinear system with state feedback (Descusse and Moog, 1985; Godbole and Sastry, 1995; Zhang and Wang, 2001; 2002). Some methods consider an inverse model of plant, such as adaptive control of feedback linearizable systems (Hirschorn, 1979) and adaptive inverse control based on adaptive signal processing (Walach and Widrow, 1983). Different from these model-based approaches for nonlinear system control, artificial neural network (ANN) achieves good approximation of nonlinear function, and has the ability of learning from experience without considering the precise plant model, and so, has drawn the intense interest of many researchers in the last decades and led to a large number of successful applications such as direct inverse control (Nguyen and Widrow, 1990; Zhang *et al*., 2002), adaptive control (Chen and Liu, 1994), and generalized inverse neural control (Dai *et al*., 2004; Zhang and Ge, 2003). Unfortunately, most of the ANNs using gradient-based training method like back-propagation, often suffer from the existence of local minima, and it is also not easy to choose a suitable neural network structure like the number of

* Project supported by the National Basic Research Program (973) of China (No. 2002CB312200), and the Hi-Tech Research and Development Program (863) of China (No. 2002AA412010)

hidden neurons.

As a novel breakthrough to ANN, Support Vector Machines (Vapnik, 1998) have proved to be a powerful alternative in many areas. Taking into account that SVM has outperformed many existing methods, there is a lot of interest in this new class of kernel-based technique, especially in the area of control theory. However, the use of SVM in a dynamical system and control context becomes quite complicated (Suykens, 2001), due to the fact that there are stringent requirements for solving online large-scale quadratic programming problems in standard SVM as SVM is basically non-parametric, and the quality and complexity of the SVM solution depend directly on the dimensionality of the input space. As a reformulation of standard SVM, least squares SVM for solving linear KKT systems has been extended to dynamical problems of recurrent neural networks (Suykens and Vandewalle, 2000) and are used in optimal control (Suykens *et al*., 2001). Compared with ANN and standard SVM, LS-SVM control has the following advantages: no number of hidden units has to be determined for the controller, no centers has to be specified for the Gaussian kernel when applying Mercer's condition, and fewer parameters have to be prescribed via the training process, so LS-SVM method is used here to design a generalized inverse controller for nonlinear system decoupling control instead of using ANN and standard SVM technologies.

This paper is organized as follows. We first give a brief review on LS-SVM regression, and then discuss a practical approach based on Bayesian evidence framework for parameters optimization of LS-SVM regression. Followed in Section IV deals with the concepts of generalized inverse system and decoupling controller design based on LS-SVM method. Section V gives details of numerical experiments conducted on a high-order multivariate nonlinear system in order to assess the effectiveness of our method. Finally, Section VI gives conclusions the work done.

## LEAST SQUARES SVM REGRESSION

Given a training dataset $D$ of $l$ samples (input and output pairs $\{x_i, y_i\}_{i=1}^{l}$) independent and identically drawn (i.i.d.) from an unknown probability distribution $\mu(\boldsymbol{X},\boldsymbol{Y})$ on the product space $\boldsymbol{Z}=\boldsymbol{X}\times\boldsymbol{Y}$:

$$D=\{z_1=(x_1,y_1),\ \ldots,\ z_n=(x_l,y_l)\}, \tag{1}$$

where the input data $\boldsymbol{X}$ is assumed to be a compact domain in a Euclidean space $R^{\mathrm{d}}$ and the output data $\boldsymbol{Y}$ is assumed to be a closed subset of $R$. Learning from the training data can be viewed as a multivariate function $f$ approximation that represents the relation between the input $\boldsymbol{X}$ and output $\boldsymbol{Y}$. Nonlinear mapping $\varphi(\cdot)$ is used to map input $\boldsymbol{X}$ into a hypothesis space $R^{n_\mathrm{h}}$ (feature space) in which the learning machine (algorithm) selects a certain function $f$.

Due to a generalized representer theorem, feature space endowed with an inner product is a function space with the linear structure of a vector space, in which a linear function estimation is defined:

$$f(x)=\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{\varphi}(x)+\boldsymbol{b}\,. \tag{2}$$

In the case of least squares Support Vector Machines (LS-SVMs) (Suykens *et al*., 2001; 2002), the optimization problem is defined as

$$\min_{\boldsymbol{\omega},\boldsymbol{b},\boldsymbol{e}} J(\boldsymbol{\omega},\boldsymbol{e})=\frac{1}{2}\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{\omega}+\gamma\frac{1}{2}\boldsymbol{e}^{\mathrm{T}}\boldsymbol{e}\,, \tag{3}$$

s.j. $$\boldsymbol{y}_k=\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{\varphi}(x_k)+\boldsymbol{b}+\boldsymbol{e}_k,\ k=1,\ \ldots,\ l\,, \tag{4}$$

where $\boldsymbol{e}\in\boldsymbol{R}^{l\times 1}$ denotes the error vector, regularization parameter $\gamma$ denotes an arbitrary positive real constant.

The conditions for optimality lead to a set of solutions:

$$\begin{bmatrix}\mathbf{0} & \mathbf{1}^{\mathrm{T}}\\ \mathbf{1} & \boldsymbol{K}+\gamma^{-1}\boldsymbol{I}\end{bmatrix}\begin{bmatrix}\boldsymbol{b}\\ \boldsymbol{\alpha}\end{bmatrix}=\begin{bmatrix}\mathbf{0}\\ \boldsymbol{y}\end{bmatrix}, \tag{5}$$

where $\boldsymbol{y}=[y_1,\ y_2,\ \ldots,\ y_l]^{\mathrm{T}}$, $\mathbf{1}_{l\times 1}=[1,\ \ldots,\ 1]^{\mathrm{T}}$, $\boldsymbol{\alpha}=[\alpha_1,\ \ldots,\ \alpha_l]^{\mathrm{T}}$. Application of the Mercer condition (Mercer, 1909) yields

$$\boldsymbol{K}(x_i,x_j)=\boldsymbol{\varphi}(x_i)^{\mathrm{T}}\boldsymbol{\varphi}(x_j),\ i,\ j=1,\ \ldots,\ l\,. \tag{6}$$

The resulting LS-SVM model for function estimation becomes:

$$f(x)=\sum_{k=1}^{l}\boldsymbol{\alpha}_k\boldsymbol{K}(x,x_k)+\boldsymbol{b}\,, \tag{7}$$

where $\boldsymbol{\alpha}_k$, $\boldsymbol{b}$ are the solutions to the linear system Eq.(5).

As one of the most popular kernel functions in machine learning, Gaussian kernel function is selected for controller design in this paper. It takes the following form:

$$\boldsymbol{K}(x_i, x_j) = \exp\left(-\| x_i - x_j \|^2 /(2\delta^2)\right), \tag{8}$$

where $\delta$ denotes the kernel (bandwidth) parameter.

## BAYESIAN PARAMETERS OPTIMIZATION

The Bayesian evidence framework first introduced by Mackay (1992; 1997) has been applied to the design of neural networks with great success. Then Kwok (2000) applied the Bayesian evidence framework to the standard SVM classification algorithm. van Gestel *et al*.(2001; 2002) has extended this integration to the LS-SVM classifier and regression problems. This approach starting from the feature space formulation, while analytic expressions are obtained in the dual space on the different levels of Bayesian inference, which yields the similar expressions of Gaussian Processes (GPs). It is known that this novel approach shows good generalization performances but with very complicated expressions for practical use. In this section, we apply the Bayesian evidence framework to the LS-SVM regression algorithm and proposed a practical approach to select optimal regularization parameter $\gamma$ and optimal kernel parameter $\delta$ of Gaussian kernel. Our method is quite simplified and similar to the Bayesian interpretation of standard SVM.

According to the Bayesian evidence theory, the inference is divided into three distinct levels. Training of the LS-SVM regression can be statistically interpreted in Level 1 inference. The optimal regularization parameter can be inferred in Level 2. The optimal kernel parameter selection can be performed in Level 3.

### Level 1 inference

To be convenient, we divide optimization objective in Eq.(3) by $\gamma$ and then replace $1/\gamma$ by $\lambda$. For a given value of $\lambda$, the first level of inference infers the posterior of $\boldsymbol{\omega}$ by

$$\begin{aligned} p(\boldsymbol{\omega} / D, \lambda, H) &= \frac{p(D / \boldsymbol{\omega}, \theta, H)\, p(\boldsymbol{\omega} / \theta, H)}{P(D / \theta, H)} \\ &\propto p(D / \boldsymbol{\omega}, \lambda, H)\, p(\boldsymbol{\omega} / \lambda, H), \end{aligned} \tag{9}$$

where $D$ is the training dataset and $H$ represents model with parameter vector $\boldsymbol{\omega}$.

Assuming training data are independently identically distributed, and $p(\boldsymbol{\omega}/\lambda,H)$ is the Gaussian probability distribution. We obtain

$$p(D / \boldsymbol{\omega}, \lambda, H) = \prod_{i=1}^{l} p(y_i / x_i, \boldsymbol{\omega}, \lambda, H)\, p(x_i / \boldsymbol{\omega}, \lambda, H), \tag{10}$$

$$p(\boldsymbol{\omega} / \lambda, H) = \left(\frac{\lambda}{2\pi}\right)^{k/2} \exp\left(-\frac{\lambda}{2}\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{\omega}\right), \tag{11}$$

where $p(x_i / \boldsymbol{\omega}, \lambda, H)$ is a constant. Let us assume

$$p(y_i / x_i, \boldsymbol{\omega}, \lambda, H) \propto \exp(-L(y_i, f(x_i))), \tag{12}$$

where $L(y_i, f(x_i))$ denotes the loss function.

The substitution of Eqs.(10), (11), and (12) in Eq.(9) yields

$$p(\boldsymbol{\omega} / D, \lambda, H) \propto \exp\left(-\frac{\lambda}{2}\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{\omega} - \sum_{i=1}^{l} L(y_i, f(x_i))\right). \tag{13}$$

Level 1 inference, training of LS-SVM Eq.(3) can be interpreted as maximizing $p(\boldsymbol{\omega} / D, \lambda, H)$ with respect to $\boldsymbol{\omega}$.

### Level 2 inference

Applying the Bayesian rule in the second level of inference, we obtain the posterior probability of $\lambda$:

$$\begin{aligned} & p(\lambda / D, H) \propto p(D / \lambda, H)\, p(\lambda / H) \propto p(D / \lambda, H) \\ & \propto \int p(D / \boldsymbol{\omega}, \lambda, H)\, P(\boldsymbol{\omega} / \lambda, H)\,\mathrm{d}\boldsymbol{\omega} \\ & \propto \left(\frac{\lambda}{2\pi}\right)^{k/2} \int \exp\left(-\frac{\lambda}{2}\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{\omega} - \sum_{i=1}^{l} L(y_i, f(x_i))\right) \mathrm{d}\boldsymbol{\omega}, \end{aligned} \tag{14}$$

The most possible value of $\lambda$ can be determined by maximizing the posterior probability of $\lambda$ as $p(\lambda/D,H)$.

Let us define $\boldsymbol{E}_{\boldsymbol{\omega}}=\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{\omega}/2$ , $\boldsymbol{E}_D=\sum_{i=1}^{l} L(y_i, f(x_i))$ to obtain

$$p(\lambda / D,H) \propto \lambda^{k/2} \int \exp[-\lambda \boldsymbol{E}_{\boldsymbol{\omega}}^{\mathrm{MP}} - \boldsymbol{E}_D^{\mathrm{MP}} - \frac{1}{2}(\boldsymbol{\omega}-\boldsymbol{\omega}_{\mathrm{MP}})\boldsymbol{A}(\boldsymbol{\omega}-\boldsymbol{\omega}_{\mathrm{MP}})]\mathrm{d}\boldsymbol{\omega} = \lambda^{k/2}\exp(-\lambda \boldsymbol{E}_{\boldsymbol{\omega}}^{\mathrm{MP}} - \boldsymbol{E}_D^{\mathrm{MP}})(2\pi)^{k/2}\det{}^{-1/2}\boldsymbol{A}, \tag{15}$$

where $\boldsymbol{\omega}_{\mathrm{MP}}$ is the most possible value of parameters $\boldsymbol{\omega}$,

$$\boldsymbol{A} = \frac{\partial^2(\lambda \boldsymbol{E}_{\boldsymbol{\omega}} + \boldsymbol{E}_D)}{\partial \boldsymbol{\omega}^2} = \nabla^2\left(\lambda \boldsymbol{E}_{\boldsymbol{\omega}} + \sum_{i=1}^{l} L(y_i, f(x_i))\right).$$

From Eq.(15), there exists

$$\ln p(\lambda / D,H) \propto \ln p(D/\lambda,H) = -\lambda \boldsymbol{E}_{\boldsymbol{\omega}}^{\mathrm{MP}} - \boldsymbol{E}_D^{\mathrm{MP}} + \frac{k}{2}\ln\lambda - \frac{1}{2}\ln(\det \boldsymbol{A}) + \text{constant}. \tag{16}$$

Maximization of the log-posterior probability of $p(\lambda/D,H)$ with respect to $\lambda$ leads to the most probable value of $\lambda_{\mathrm{MP}}$ obtained by the following equation:

$$2\lambda_{\mathrm{MP}}\boldsymbol{E}_{\boldsymbol{\omega}}^{\mathrm{MP}} = \varsigma, \tag{17}$$

where $\varsigma=k-\lambda\mathrm{trace}\boldsymbol{A}^{-1}$ is called the effective number of parameters.

In the case of LS-SVM regression, used of cost function $L(y_i, f(x_i))=\frac{1}{2}(y_i-\boldsymbol{\omega}\phi(x_i)-b_i)^2$ yields

$$\boldsymbol{A} = \nabla^2\left(\lambda \boldsymbol{E}_{\boldsymbol{\omega}} + \sum_{i=1}^{l} L\left(y_i, f(x_i)\right)\right) = \lambda \boldsymbol{I} + \boldsymbol{B}, \tag{18}$$

where $\boldsymbol{B}=\sum_{i=1}^{l}\boldsymbol{\varphi}(x_i)\boldsymbol{\varphi}(x_i)^{\mathrm{T}}$. Denote the eigenvalues of $\boldsymbol{B}$ by $\rho_m$ yields the effective number of parameters $\varsigma$ of LS-SVM as follows:

$$\varsigma = k - \lambda\mathrm{trace}\boldsymbol{A}^{-1} = \sum_{i=1}^{N}\frac{\rho_i}{\lambda+\rho_i}, \tag{19}$$

where $N$ ($N\leq l$) denotes the number of nonzero eigenvalues of $l\times l$ matrix $\boldsymbol{K}(x_i,x_j)=\boldsymbol{\varphi}(x_i)^{\mathrm{T}}\boldsymbol{\varphi}(x_j)$, $i, j$=1, 2, …, $l$.

**Level 3 inference**

The third level of inference in the evidence framework compares the different models by examining their posterior probabilities $p(H/D)\propto p(D/H)p(H)$ and can be used to find the optimum kernel parameter. Assuming the prior probability $p(H)$ over all possible models is uniform, we have

$$p(H/D) \propto p(D/H) \propto \int p(D/\lambda,H)p(\lambda/H)\mathrm{d}\lambda \propto p(D/\lambda_{\mathrm{MP}},H)/\sqrt{\varsigma}. \tag{20}$$

Therefore

$$\ln p(H/D) = -\lambda_{\mathrm{MP}}\boldsymbol{E}_{\boldsymbol{\omega}}^{\mathrm{MP}} - \boldsymbol{E}_D^{\mathrm{MP}} + \frac{k}{2}\ln\lambda_{\mathrm{MP}} - \frac{1}{2}\ln(\det\boldsymbol{A}) - \frac{1}{2}\ln(k-\lambda_{\mathrm{MP}}\mathrm{trace}\boldsymbol{A}^{-1}) + \text{constant}. \tag{21}$$

The optimum kernel parameter can be obtained by maximizing log-posterior probabilities $\ln p(H/D)$ with respect to the kernel parameter. For practical use, the selection method of the kernel parameter $\delta$ of Gaussian kernel is illustrated in this subsection.

To obtain the most possible value of the kernel parameter $\delta$, we set the derivative of $\ln p(H/D)$ with respect to $\delta$ to zero

$$\frac{\partial \ln p(H/D)}{\partial \delta} = 0. \tag{22}$$

Noting that

$$\frac{\partial(\lambda_{\mathrm{MP}}\boldsymbol{E}_{\boldsymbol{\omega}}^{\mathrm{MP}})}{\partial\delta} = -\lambda_{\mathrm{MP}}a_i a_j\frac{\partial K}{\partial \delta} = -\lambda_{\mathrm{MP}}\sum_{i,j=1}^{l} a_i a_j \exp\left(-\frac{(x_i-x_j)^2}{2\delta^2}\right)(x_i-x_j)^2\delta^{-3}, \tag{23}$$

$$\frac{\partial \ln(\det\boldsymbol{A})}{\partial\delta} = \mathrm{trace}\left(\boldsymbol{A}^{-1}\left(\frac{\partial \boldsymbol{A}}{\partial\delta}\right)\right) = \mathrm{trace}\left(\boldsymbol{A}^{-1}\left(\frac{\partial \bar{K}}{\partial\delta}\right)\right), \tag{24}$$

$$\frac{\partial \ln(k-\lambda_{\mathrm{MP}}\mathrm{trace}\boldsymbol{A}^{-1})}{\partial\delta} = \frac{\lambda_{\mathrm{MP}}}{k-\lambda_{\mathrm{MP}}\mathrm{trace}\boldsymbol{A}^{-1}}\mathrm{trace}\left(\boldsymbol{A}^{-2}\left(\frac{\partial\bar{K}}{\partial\delta}\right)\right), \tag{25}$$

By substituting Eqs.(23)~(25) into Eq.(22), we obtain the kernel parameter in the LS-SVM regression:

$$\delta=\left|\left[\frac{\lambda_{\mathrm{MP}}\sum_{i,j=1}^{l}a_i a_j \exp\left(-\frac{(x_i-x_j)^2}{2\delta^2}\right)(x_i-x_j)^2}{\mathrm{trace}\left(\boldsymbol{A}^{-1}\left(\frac{\partial K}{\partial \delta}\right)\right)+\frac{\lambda_{\mathrm{MP}}}{k-\lambda_{\mathrm{MP}}\mathrm{trace}\boldsymbol{A}^{-1}}\mathrm{trace}\left(\boldsymbol{A}^{-2}\left(\frac{\partial K}{\partial \delta}\right)\right)}\right]^{1/3}\right|. \quad (26)$$

It is worth noting that the absolute value of $\delta$ is employed. This is because the kernel width $\delta$ in Gaussian kernel should be positive, and such treatment could enhance the convergence speed of the iterative process. For further perspective of model section method and the Bayesian evidence theory one can refer to the previous work (Yan *et al*., 2004) and the references therein.

## GENERALIZED INVERSE CONTROLLER DESIGN BASED ON LS-SVM METHOD

### Generalized $\alpha$th-order integral inverse system

Given a linear or nonlinear system $\Sigma: u\rightarrow y$, with $p$-dimensional inputs $\boldsymbol{u}=[u_1,u_2,\ldots,u_p]^{\mathrm{T}}$, $p$-dimensional outputs $\boldsymbol{y}=[y_1,y_2,\ldots,y_p]^{\mathrm{T}}$, and an initial state vector $\boldsymbol{x}(t_0)=\boldsymbol{x}_0$. We define a map operator $\phi: \boldsymbol{u}\rightarrow\boldsymbol{y}$, i.e.

$$\boldsymbol{y}(\cdot)=\phi[\boldsymbol{x}_0,\boldsymbol{u}(\cdot)] \text{ or } \boldsymbol{y}=\phi\boldsymbol{u}. \quad (27)$$

The definition of its inverse system is as follows:

**Definition 1** Given a MIMO system $\Sigma$ expressed by Eq.(27). Assume $\Pi$ is a system $\bar{\phi}:\boldsymbol{\varphi}\rightarrow\boldsymbol{u}_d$ with $p$-dimensional inputs and $p$-dimensional outputs, where $\boldsymbol{\varphi}=[\varphi_1,\varphi_2,\ldots,\varphi_p]^{\mathrm{T}}$ and $\boldsymbol{u}_d(t)=[u_{d1},u_{d2},\ldots,u_{dp}]^{\mathrm{T}}$ both are subject to some initial conditions. Let $\varphi_i$ be the $\alpha_i$th-order derivative of $y_{di}$. That is, $\boldsymbol{\varphi}=\boldsymbol{y}_d^{(\alpha)}$, where $\boldsymbol{y}_d=[y_{d1},y_{d2},\ldots,y_{dp}]^{\mathrm{T}}$ is the expected outputs, and $\boldsymbol{\alpha}=[\alpha_1,\alpha_2,\ldots,\alpha_p]$ with integer $\alpha_i\geq 0$. Then $\Pi$ is referred as $\alpha$th-order integral inverse system of $\Sigma$ (or call $\alpha$th-order inversion in short) if $\bar{\phi}$ satisfies

$$\phi\bar{\phi}\boldsymbol{\varphi}=\phi\bar{\phi}(\boldsymbol{y}_d^{(\alpha)})=\phi\boldsymbol{u}_d=\boldsymbol{y}_d. \quad (28)$$

If $\alpha$th-order inversion exists, then the original system $\Sigma$ is said to be invertible.

Connecting an ordinary $\alpha$th-order inversion with its original system, a composite system with a map $M\triangleq\phi\bar{\phi}$ of diagonal matrix property is obtained, it means that there is no cross-interaction between inputs and outputs, the original system $\Sigma$ is decoupled into many SISO independent nonlinear subsystems. For $\alpha$=0, it is known as the identical matrix decoupling system.

According to this notion, the core of decoupling controller design for a general nonlinear dynamic continuous system is to construct an inversion model. If we can design a multivariate LS-SVM approximator to construct a map function $\bar{\phi}(\cdot)$ as an $\alpha$th-order inversion, then connect this inversion with the original system, a pseudo-linear composite system with the property of integral linearization and decoupling will be obtained (Fig.1).

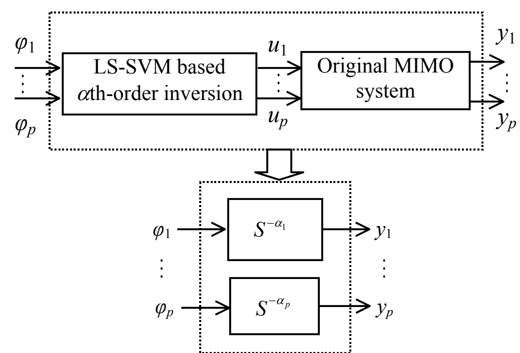


**Fig.1 Decoupling process via $\alpha$th-order inversion**

However, it is not easy to design an additional linear controller as the integrator $S^{-\alpha_i}$ ($i$=1, 2, ..., $p$) is not stable, and it frequently fails when the states of system are difficult to measure. In order to get stable linearization and decoupling of the original system, we consider the poles configuration and decoupling process simultaneously. More generally, let

$$\varphi_i=a_{i0}y_{di}^{(\alpha_i)}+a_{i1}y_{di}^{(\alpha_i-1)}+a_{i2}y_{di}^{(\alpha_i-2)}+\ldots+a_{i\alpha}y_{di}, \quad i=1,\ 2,\ \ldots,\ p, \quad (29)$$

where $a_{i0}$, $a_{i1}$, …, $a_{i\alpha}$ being real and $a_{i0}\neq 0$. If there exists a map $\bar{\phi}_{\mathrm{g}}$ of system $\Pi$ satisfies the relation of Eq.(28), we call it as the generalized ($\alpha$th-order) inversion of the original system $\Sigma$. If we set $a_{i1}=a_{i2}=\ldots=a_{i\alpha}=0$ and $a_{i0}=1$, the ordinary $\alpha$th-order inversion can be derived from generalized inversion as a special case in Eq.(29).

**Generalized inversion based on LS-SVM regression**

To design a generalized inversion of a MIMO system, some theoretic analysis is briefly discussed.

**Lemma 1** (He *et al.*, 2002) Given a MIMO (multiple inputs and multiple outputs) system

$$\Sigma : F(\boldsymbol{Y}^{(N)}, \boldsymbol{X}_Y, \boldsymbol{X}_U, \boldsymbol{U}^{(M)})=0. \quad (30)$$

With $\boldsymbol{X}_Y(t_0)=\boldsymbol{X}_{Y_0}$ and $\boldsymbol{X}_U(t_0)=\boldsymbol{X}_{U_0}$, if $\partial F/\partial \boldsymbol{U}^{(M)}$ is nonsingular and continuous in certain open set $D$, then the system is invertible and there exists a generalized inverse system $\Pi_g$.

Where

$$N=(n_1, n_2, \ldots, n_p), \quad M=(m_1, m_2, \ldots, m_p),$$
$$\boldsymbol{Y}^{(N)}=[y_1^{(n_1)}, y_2^{(n_2)}, \ldots, y_p^{(n_p)}]^{\mathrm{T}},$$
$$\boldsymbol{U}^{(M)}=[u_1^{(m_1)}, u_2^{(m_2)}, \ldots, u_p^{(m_p)}]^{\mathrm{T}},$$
$$\boldsymbol{X}_Y=[\boldsymbol{X}_{Y_1}, \boldsymbol{X}_{Y_2}, \ldots, \boldsymbol{X}_{Y_p}]^{\mathrm{T}},\ \boldsymbol{X}_{Y_i}=[y_i, \dot{y}_i, \ldots, y_i^{(n_i-1)}],$$
$$\boldsymbol{X}_U=[\boldsymbol{X}_{U_1}, \boldsymbol{X}_{U_2}, \ldots, \boldsymbol{X}_{U_p}]^{\mathrm{T}},\ \boldsymbol{X}_{U_i}=[u_i, \dot{u}_i, \ldots, u_i^{(m_i-1)}],$$
$$i=1,\ 2,\ \ldots,\ p,$$

and $F(\cdot)$ are locally analytic functions.

For easy to understand the process of the poles configuration and decoupling process of generalized inversion of MIMO system, let us assume that there exists certain SISO nonlinear subsystem which satisfies

$$f(y_i^{(n)}, y_i^{(n-1)}, \ldots, y_i, u_i^{(m)}, u_i^{(m-1)}, \ldots, u_i)=0, \quad (31)$$

where $f(\cdot)$ is a locally analytic function, and

$$u_i^{(k)}(t_0)=u_{i0}^{(k)},\ k=0,\ 1,\ \ldots,\ m-1, \quad (32)$$

$$y_i^{(l)}(t_0)=y_{i0}^{(l)},\ l=0,\ 1,\ \ldots,\ n-1. \quad (33)$$

According to Lemma 1, the generalized inversion exists if $(\partial f/\partial u)^{(m)}\neq 0$ and is continuous on $D$. Obviously, there must exist a unique solution of $u_i^{(m)}$ according to the implicit function theorem (Nijmeijer and Schaft, 1990), that is

$$u_i^{(m_i)}=g(y_i^{(n_i)}, y_i^{(n_i-1)}, \ldots, y_i, u_i^{(m_i-1)}, \ldots, u_i),\ i\in p. \quad (34)$$

Let

$$\begin{bmatrix} z_{i0} \\ z_{i1} \\ \vdots \\ z_{in_i} \end{bmatrix} = \boldsymbol{A}_i \begin{bmatrix} y_i \\ \dot{y}_i \\ \vdots \\ y_i^{(n_i-1)} \end{bmatrix},\ i\in p, \quad (35)$$

where matrix $\boldsymbol{A}_i$ is the following $(n_i+1)\times(n_i+1)$ non-singular matrix with $a_{ikk}\neq 0$, $k=0, 1, \ldots, n_i$. That is

$$\boldsymbol{A}_i = \begin{bmatrix} a_{i00} & 0 & \cdots & 0 \\ a_{i10} & a_{i21} & \cdots & 0 \\ \cdots & \cdots & \ddots & \cdots \\ a_{in_i0} & a_{in_i1} & \cdots & a_{in_in_i} \end{bmatrix}. \quad (36)$$

From Eq.(35), we obtain

$$\begin{bmatrix} y_i \\ \dot{y}_i \\ \cdots \\ y_i^{(n_i-1)} \end{bmatrix} = \boldsymbol{A}_i^{-1} \begin{bmatrix} z_{i0} \\ z_{i1} \\ \cdots \\ z_{in_i} \end{bmatrix}. \quad (37)$$

Substituting Eq.(37) into Eq.(34) yields

$$u_i^{(m_i)}=\bar{g}(z_{in_i}, z_{i(n_i-1)}, \ldots, z_{i0}, u_i^{(m_i-1)}, \ldots, u_i). \quad (38)$$

Let $\varphi_i=z_{in_i}$ be the input, and $u_i$ be the output. Combining this inverse system $\Pi_i:\varphi_i\rightarrow u_i$ with the original SISO nonlinear subsystem $\Sigma_i:u_i\rightarrow y_i$ yields a pseudo-linear SISO subsystem whose transfer function is

$$g_{ii}(s)=\frac{y_i(s)}{\varphi_i(s)}=\frac{1}{a_{in_in_i}s^{n_i}+\ldots+a_{in_i1}s+a_{in_i0}}. \quad (39)$$

So far, a generalized inversion of the SISO system Eq.(31) can be obtained. Evidently, by regulating the elements of matrix $\boldsymbol{A}_i$ Eq.(36), we can configure the poles of pseudo-linear SISO system Eq.(39) to desired positions.

In the case of MIMO system of Eq.(30), it is known that there exists an equivalent implicit function $G$ on certain open set $D$, which satisfies

$$U^{(M)}=G(Y^{(N)},X_Y,X_U)\,. \quad (40)$$

Similar to the design of the generalized inversion of SISO system, it is easy to deduce

$$U^{(M)}=\sigma(Z_N,X_Z,X_U)\,, \quad (41)$$

where

$X_Z=[X_{Z_1},X_{Z_2},\ldots,X_{Z_p}]^{\mathrm{T}}$,

$X_{Z_i}=[z_{i0},z_{i1},\ldots,z_{i(n_i-1)}]$, $X_U=[X_{U_1},X_{U_2},\ldots,X_{U_p}]^{\mathrm{T}}$,

$X_{U_i}=[u_i,\dot{u}_i,\ldots,u_i^{(m_i-1)}]$, and $Z_N=[z_{1n_1},z_{2n_2},\ldots,z_{pn_p}]^{\mathrm{T}}$, $i\in p$.

Let $\boldsymbol{\Psi}=Z_N=[\varphi_1, \varphi_1, \ldots, \varphi_p]^{\mathrm{T}}$ be the input, we can obtain a generalized inversion of the original MIMO system. By connecting this generalized inverse system with the original system, we obtain a series of pseudo-linear SISO subsystems whose poles are configured according to the matrix $A_i$ in Eq.(36). Fig.2 shows the linearization and decoupling process of MIMO system through a generalized inversion method.

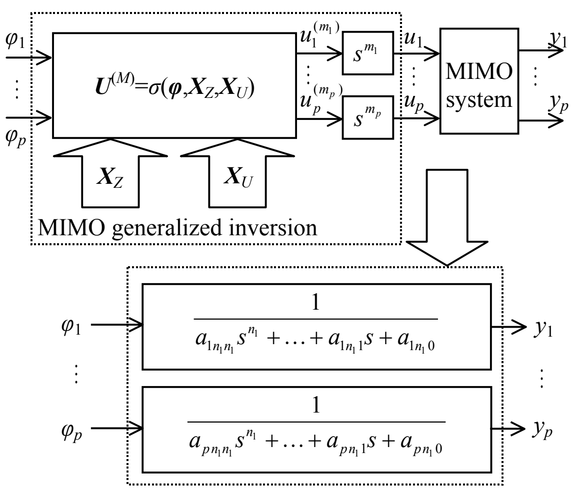


**Fig.2 Decoupling process via generalized inversion**

Taking advantage of the good generalization ability of LS-SVM regression algorithm, we can construct a nonlinear static function approximator $\sigma(\cdot)$ in Eq.(41). By combining this static LS-SVM approximator with some linear dynamic components, such as integral or inertial components, a novel dynamic generalized inverse controller based on LS-SVM regression is proposed for the linearization and decoupling control of a general nonlinear invertible MIMO system Eq.(30). The static LS-SVM approximator is responsible for approximating the static nonlinear mapping described by the analytical expression of the generalized inversion and the linear components are used to represent the dynamics property. Fig.3 shows the dynamic generalized inversion structure based on the static multivariate LS-SVM approximator and dynamic components, where $\varphi_i$ ($i$=1, 2, …, $p$) is the $i$th input of the generalized inverse system.

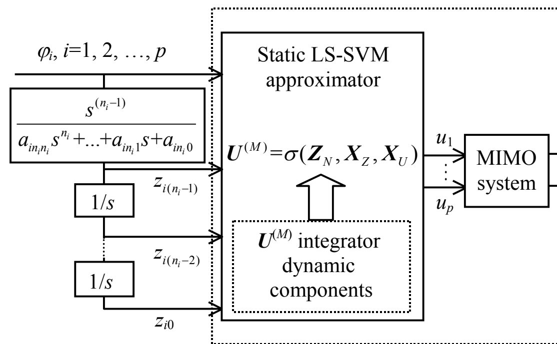


**Fig.3 The structure of generalized inverse system decoupling via LS-SVM**

## Implementation of generalized inversion modelling based on LS-SVM

Since the mathematical model of the plant is frequently not completely known, and as the feedback $y_i^{(\alpha_i-1)}$, …, $\dot{y}_i$ ($i$=1, 2, …, $p$) and the state variables of the plant are not always measurable, the nonlinear property and the complex coupling existing in the plant often make it quite hard for the implementation of generalized inversion in practice. Fortunately, modelling based on LS-SVM is a black-box model based only on input-output measurements of the original nonlinear system. In the modelling procedure, the relationship between the input and output of the plant can be emphasized while the sophisticated inner structure is ignored. The modelling process of a generalized inversion based on LS-SVM method is illustrated in Fig.4.

For simplicity, only the canonical form of $A_i$ in Eq.(36) is considered in this paper. That is

$$A_i=\begin{bmatrix} I_{n_i\times n_i} & \mathbf{0}_{n_i\times 1} \\ a_{in_i0},\ldots,a_{in_i(n_i-1)} & a_{in_in_i}\end{bmatrix},\ (i=1,2,\ldots,p). \quad (42)$$

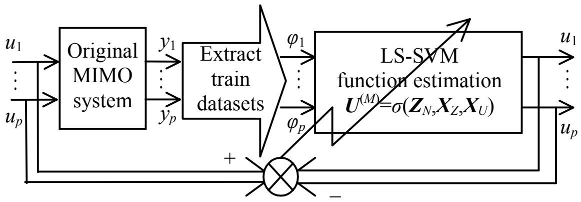


**Fig.4 Generalized inverse modelling via LS-SVM**

The implementation procedure of the LS-SVM based generalized inversion is summarized as follows:

Step 1: Determine the structure of the generalized inversion system.

To get a pseudo-linear composite system with ideal dynamic and static characteristic of a MIMO system, the desired poles for each pseudo-linear subsystem should be prescribed initially. According to the determined poles, one can find the parameters of the matrix $\boldsymbol{A}_i$ ($i$=1, 2, …, $p$), then obtain the linear components that stand for the dynamic part of the generalized inversion. Moreover, the matrix $\boldsymbol{A}_i$ ($i$=1, 2, …, $p$) can determine the training dataset in Eq.(41) with $\{\boldsymbol{Z}_N, \boldsymbol{X}_Z, \boldsymbol{X}_U\}$ as the input and $\{\boldsymbol{U}^{(M)}\}$ the output, respectively.

Step 2: Generate the training datasets.

(1) If the plant has no mathematical model or is only poorly modelled, one should generate 'rich' enough exciting or testing signals, such as square wave or ramp signals, to capture the dynamics of the original system.

(2) Sample inputs, outputs and states of the original system (if possible).

(3) Calculate the derivatives of each output to obtain the required training datasets.

Step 3: Train the inversion model based on LS-SVM within Bayesian framework.

(1) Given certain initial value for regularization parameter $\gamma$ and kernel parameter $\delta$, the generalized inversion model based on LS-SVM can be obtained.

(2) The second level of inference determines the most possible value of the regularization parameter $\gamma$.

(3) The third level of inference determines the most possible value of the kernel parameter $\delta$.

(4) Retrain the LS-SVM inversion model using the most possible values of the regularization parameter $\gamma$ and kernel parameters $\delta$ until enough accuracy is reached.

Step 4: connect the optimal generalized inversion model with the original system.

So far, a practical generalized inversion based on LS-SVM for decoupling control has been proposed.

## SIMULATION RESULTS

Given a nonlinear MIMO (two inputs two outputs) system $\Sigma$ (He *et al.*, 2002):

$$\begin{aligned} &\ddot{y}_1 + 0.8\dot{y}_1 + 2y_1 - \sin(\ddot{y}_2) = u_1, \\ &\dddot{y}_2 + 1.8\ddot{y}_2 + 2.5\dot{y}_2 + 3y_2 + 0.8\dot{y}_2 = u_2. \end{aligned} \tag{43}$$

Let the expected transfer function be

$$G(s) = \mathrm{diag}\{g_{11}(s), g_{22}(s)\}, \tag{44}$$

where $g_{11}(s)=(s^2+1.1414s+1)^{-1}$ with two poles being $-0.707\pm0.707$j and $g_{22}(s)=(0.1s^3+1.1414s^2+1.514s+1)^{-1}$ with three poles being $-0.707\pm0.707$j and $-10$. We assume that we have no prior knowledge of the real mathematical model except the parameters of the highest order of inputs and outputs.

According to Lemma 1, the system Eq.(43) is obviously invertible. Then we can obtain the structure of generalized inversion from Eq.(36), i.e.

$$\boldsymbol{A}_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1.1414 & 1 \end{bmatrix},$$

and

$$\boldsymbol{A}_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 1.514 & 1.1414 & 0.1 \end{bmatrix}.$$

We can construct a generalized inverse system based on LS-SVM as illustrated in Fig.3. For training the LS-SVM inversion model, we first generate two random signals as inputs to the original system Eq.(43). In our experiment, we selected multi-amplitude varying-step square wave as testing signals. By sampling the inputs and outputs at high speed, we obtain data $\{y_1, y_2, u_1, u_2\}$. After computing the derivatives $\{\dot{y}_1, \ddot{y}_1, \dot{y}_2, \ddot{y}_2, \dddot{y}_2\}$ off-line, and let

$$\begin{bmatrix} z_{10} \\ z_{11} \\ z_{12} \end{bmatrix} = \boldsymbol{A}_1 \begin{bmatrix} y_1 \\ \dot{y}_1 \\ \ddot{y}_1 \end{bmatrix} \text{ and } \begin{bmatrix} z_{20} \\ z_{21} \\ z_{22} \\ z_{23} \end{bmatrix} = \boldsymbol{A}_2 \begin{bmatrix} y_2 \\ \dot{y}_2 \\ \ddot{y}_2 \\ \dddot{y}_2 \end{bmatrix},$$

we can obtain the training dataset of the generalized inversion. Finally, we train the LS-SVM inversion model using Bayesian parameter optimization and connecting the optimal LS-SVM inversion model with the original system. In our experiment, the training dataset consisted of 300 samples, Gaussian kernel is selected and the regularization parameter and kernel parameter are optimized efficiently with $\{\gamma_1=4127, \delta_1=13.55\}$ and $\{\gamma_2=3795, \delta_2=15.33\}$.

The simulation results are illustrated in Fig.5 and Fig.6. Fig.5 shows the response of the original system with square wave input, and Fig.6 shows the response of the compensated system via the novel approach. The dash line denotes compensated system output, and solid line denotes desired output. It is worth noting that they can hardly be distinguished from each other in Fig.6.

The simulation results showed that, the outputs of decoupled MIMO systems could follow closely the expected outputs. This means that the LS-SVM-based $\alpha$th-order inverse system controller has good decoupling effects and good control performance. Furthermore, it is straightforward to design a closed loop stabilization controller by using an additional closed-loop linear controller like a PID controller.

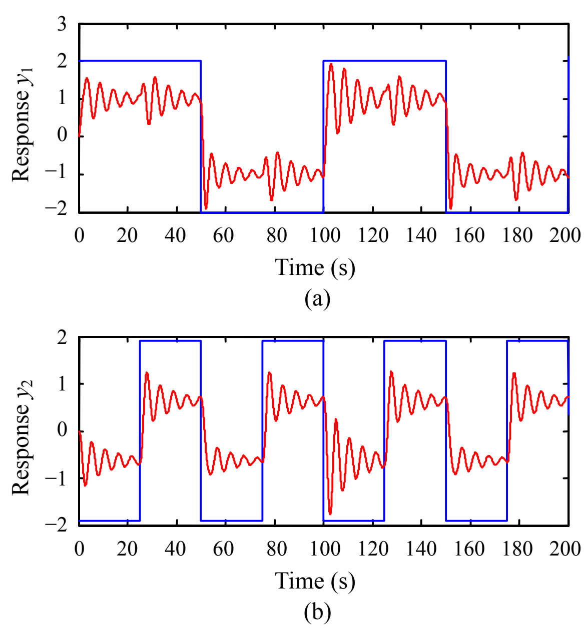


**Fig.5 The response of the original system with square wave input. (a) The original response $y_1$; (b) The original response $y_2$**

## CONCLUSION

In this paper, we discussed a LS-SVM-based generalized inverse of decoupling control for multivariate nonlinear systems. By using a multivariate LS-SVM inverse controller and a number of linear components, the nonlinear system can be decoupled into a number of pseudo-linear SISO subsystems and regulated to desired response. A case study showed that our method is effective. This technique does not require accurate mathematical models of the plants, and compared with standard SVM, is a more practical

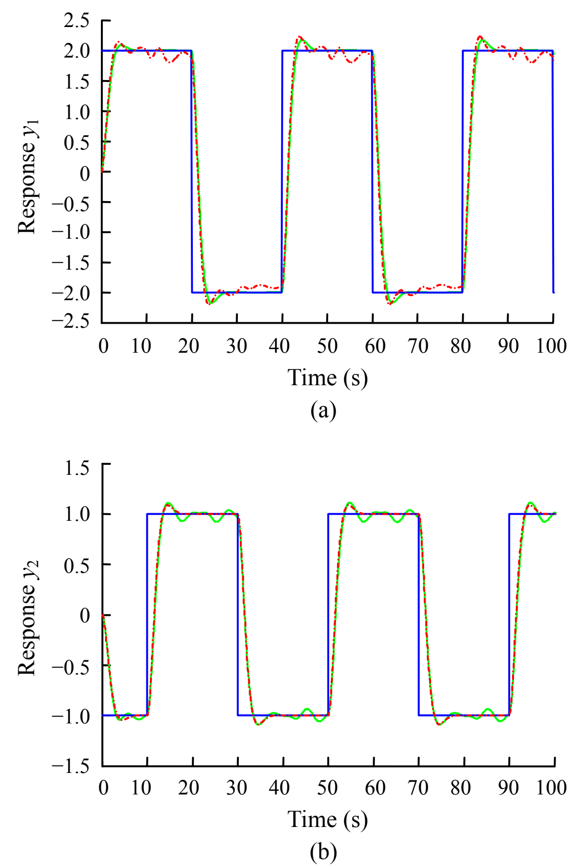


**Fig.6 The response of the decoupled system with square wave input. (a) The compensated system response $y_1$; (b) The compensated system response $y_2$**

alternative in nonlinear control. In general, the LS-SVM methodology might offer a better opportunity in the area of control theory although its application still needs further study.

## References

Chen, F., Liu, C., 1994. Adaptively controlling nonlinear continuous-time systems using multiplayer neural networks. *IEEE Transactions on Automatic Control*, **39**(6):1306-1310. [doi:10.1109/9.293202]

Dai, X., Zhang, K., Zhang, T., Lu, X., 2004. ANN generalized inversion control of turbo generator governor. *IEE Proceedings-Generation, Transmission and Distribution*, **151**(3):327-333. [doi:10.1049/ip-gtd:20040380]

Descusse, J., Moog, C., 1985. Decoupling with dynamic compensation for strong invertible affine nonlinear systems. *International Journal of Control*, **42**(6):1387-1398.

Godbole, D.N., Sastry, S.S., 1995. Approximate decoupling and asymptotic tracking for MIMO systems. *IEEE Transactions on Automatic Control*, **40**(3):441-450. [doi:10.1109/9.376056]

He, D., Dai, X., Wang, Q., 2002. Generalized ANN inverse control method. *Control Theory and Applications*, **19**(1):34-40 (in Chinese).

Hirschorn, R.M., 1979. Invertibility of multivariable nonlinear control systems. *IEEE Transactions on Automatic Control*, **24**(6):855-865. [doi:10.1109/TAC.1979.1102181]

Kwok, J.T., 2000. The evidence framework applied to Support Vector Machines. *IEEE Transaction on Neural Network*, **11**(5):1162-1173. [doi:10.1109/72.870047]

Mackay, D.J.C., 1992. Bayesian interpolation. *Neural Computing*, **4**:415-447.

Mackay, D.J.C., 1997. Probable network and plausible predictions-A review of practical Bayesian methods for supervised neural networks. *Network Computation in Neural Systems*, **6**:1222-1267.

Mercer, J., 1909. Functions of positive and negative type and their connection with the theory of integral equations. *Transactions of the London Philosophical Society A*, **209**:415-446.

Nguyen, D.H., Widrow, B., 1990. Neural networks for self-learning control systems. *IEEE Control Systems Magazine*, **10**(3):18-23. [doi:10.1109/37.55119]

Nijmeijer, H., Schaft, A., 1990. Nonlinear Dynamical Control Systems. Springer-Verlag, New York.

Suykens, J.A.K., 2001. Support vector machines: a nonlinear modelling and control perspective. *European Journal of Control*, **7**(2-3):311-327.

Suykens, J.A.K., Vandewalle, J., 2000. Recurrent least squares support vector machines. *IEEE Transactions on Circuits and Systems, Part I*, **47**(7):1109-1114. [doi:10.1109/81.855471]

Suykens, J.A.K., Vandewalle, J., de Moor, B., 2001. Optimal control by least squares Support Vector Machines. *Neural Networks*, **14**(1):23-35. [doi:10.1016/S0893-6080(00)00077-0]

Suykens, J.A.K., Gestel, T., de Brabanter, J., de Moor, B., Vandewalle, J., 2002. Least Squares Support Vector Machines. World Scientific, Singapore.

van Gestel, T., Suykens, J.A.K., Baestaens, D., Lambrechts, A., Lanckriet, G., Vandaele, B., de Moor, B., 2001. Financial time series prediction using least squares support vector machines within the evidence framework. *IEEE Transaction on Neural Network*, **12**(4):809-821. [doi:10.1109/72.935093]

van Gestel, T., Suykens, J.A.K, Lanckriet, G., Lambrechts, A., de Moor, B., Vandewalle, J., 2002. Bayesian framework for least squares Support Vector Machine classifiers, Gaussian processes and Kernel Fisher discriminant analysis. *Neural Computation*, **15**(5):1115-1148. [doi:10.1162/089976602753633411]

Vapnik, V., 1998. The Nature of Statistical Learning Theory. Springer-Verlag, New York.

Walach, E., Widrow, B., 1983. Adaptive Signal Processing for Adaptive Control. IFAC workshop on Adaptive Systems in Control and Signal Processing, San Francisco, CA.

Yan, W., Shao, H., Wang, X., 2004. Soft sensing modelling based on Support Vector Machine and Bayesian model selection. *Computers and Chemical Engineering*, **28**:1489-1498.

Zhang, Y., Wang, J., 2001. Recurrent neural networks for nonlinear output regulation. *Automatica*, **37**:1161-1173. [doi:10.1016/S0005-1098(01)00092-9]

Zhang, Y., Wang, J., 2002. Global exponential stability of recurrent neural networks for synthesizing linear feedback control systems via pole assignment. *IEEE Transactions on Neural Networks*, **13**(3):633-644. [doi:10.1109/TNN.2002.1000129]

Zhang, Y., Ge, S.S., 2003. A General Recurrent Neural Network Model for Time-varying Matrix Inversion. The IEEE 42nd IEEE Conference on Decision and Control, p.6169-6174.

Zhang, Y., Jiang, D., Wang, J. 2002. A recurrent neural network for solving Sylvester equation with time-varying coefficients. *IEEE Transactions on Neural Networks*, **13**(5):1053-1063. [doi:10.1109/TNN.2002.1031938]