Estimation of Reference Indices of Left Ventricular Chamber Function from Echocardiographic Images with Multidimensional Kernel Methods

Ricardo Santiago-Mozos[1], José Luis Rojo-Álvarez[1], J Carlos Antoranz[2], Daniel Rodríguez[2], Mar Desco[2], Alicia Barrio[3], Yolanda Benito[3], Raquel Yotti[3], Javier Bermejo[3]

[1]Dep. of Signal Theory and Communications, Universidad Rey Juan Carlos, Madrid, Spain
[2]Dep. of Mathematical Physics and Fluids, Universidad Nacional de Educación a Distancia, Spain
[3]Dep. of Cardiology, Hospital General Universitario Gregorio Marañón, Madrid, Spain

Abstract

Advanced nonlinear estimation methods can compete with their linear counterparts for the estimation of left ventricular (LV) function indices from color-Doppler M-mode images. We benchmarked three methods: Support Vector Regression, Partial Least Squares and Principal Component Regression using linear and non-linear (Gaussian) kernels. Two reference indices were directly estimated from the images, namely, the peak-systolic elastance (E_{max}) and the time-constant of LV relaxation (τ).

We found linear methods performing slightly better for predicting E_{max}, an easier task, but they were outperformed by non-linear procedures when predicting τ, a harder estimation problem.

1.   Introduction

Some well-known indices of left ventricular (LV) function can be qualitatively related to color-Doppler M-mode (CDMM) images, and non-invasively estimated using robust linear regression methods [1]. In this paper, we challenge the common assumption than linear estimation methods are better than non-linear methods for high dimensional input spaces like CDMM images. We compare three kernel methods for estimation: on the one hand, the conventional Support Vector Regression (SVR) [2]; on the other hand, two methods accounting for implicit or explicit input space reduction, namely Partial Least Squares (PLS) [3], and Principal Components Regression (PCR) [4]. All the kernel based methods were tested using linear and nonlinear (Gaussian) kernels. Two reference indices of LV function were estimated, namely, the peak-systolic elastance (E_{max}) and the time-constant of LV relaxation (τ) [1].

Simultaneous invasive and Doppler tracings were obtained in 9 and 20 mini-pigs for E_{max} and τ, in a high-fidelity experimental setup (274 and 1362 cases of E_{max}

and τ, respectively) of animals undergoing ischaemia as well as load and adrenergic modulation (dobutamine and esmolol infusion). Datasets were simultaneous for hameo-dynamical and CDDM images in each case (beat). CDDM images were time and space aligned and downsampled to 32x32 pixels.

2.   Methods

We denote the mean of the absolute error of the estimation of variable Y for individual m as:

$$\bar{E}_m = 1/N_m \sum_{i=1}^{N_m} |\hat{Y}_m[i] - Y_m[i]| \, ,$$

where N_m is the number of cases recorded for individual m, $Y_m[i]$ and $\hat{Y}_m[i]$ are respectively the measured and estimated variables. We use as error figure $\bar{E}$, defined as the mean of this error for all individuals.

2.1.   Support vector regression

SVR is a well-studied technique that works well with high dimensional spaces like images [2]. Given a training set $\{(\mathbf{x}_i, y_i), \ i = 1, \dots, n\}$ where $\mathbf{x}_i \in \mathbb{R}^d$ and $y_i \in \mathbb{R}$, the SVR finds a function f that estimates y_i as

$$\hat{y}_i = f(\mathbf{x}_i) = \boldsymbol{\phi}^T(\mathbf{x}_i)\mathbf{w} + b = y_i + e_i$$

where $\boldsymbol{\phi} : \mathbb{R}^d \to \mathcal{H}$ is in general a non-linear mapping to the feature space $\mathcal{H}$; $\mathbf{w}$ is the weight vector in this space; b is a bias term and e_i is the residual (error). f is found by minimizing a functional with a regularization term and a loss term as follows:

$$L = \frac{1}{2}||\mathbf{w}||^2 + \sum_i \mathcal{L}(e_i)$$

where $\mathcal{L}$ is in our case the robust ϵ-Hubber loss function [1], which increases the flexibility modelling outliers,

namely:

$$\mathcal{L}(e_i) = \begin{cases} 0 & |e_i| \leq \epsilon \\ \frac{1}{2\delta}(|e_i| - \epsilon)^2 & \epsilon \leq |e_i| \leq \epsilon + \delta C \\ C(|e_i| - \epsilon) - \frac{1}{2}\delta C^2 & \epsilon + \delta C \leq |e_i| \end{cases} \quad (1)$$

where ϵ is the insensitive-zone parameter (no loss for errors lower than ϵ), and δC controls the size of the quadratic zone of the loss function. Finally, we minimize the following convex problem

$$L = \frac{1}{2}||\mathbf{w}||^2 + \frac{1}{2\delta}\sum_{i \in I_1}(\xi_i^2 + \xi_i^{*2}) + C\sum_{i \in I_2}(\xi_i + \xi_i^*) - \sum_{i \in I_2}\frac{\delta C^2}{2}$$

with respect to $\mathbf{w}, b, \xi_i, \xi_i^*$ taking into account the following convex constraints:

$$y_i - \phi^T(\mathbf{x}_i)\mathbf{w} - b \leq \epsilon + \xi_i \qquad \forall i$$
$$\phi^T(\mathbf{x}_i)\mathbf{w} + b - y_i \leq \epsilon + \xi_i^* \qquad \forall i$$
$$\xi_i, \xi_i^* \geq 0 \qquad \forall i$$

where ξ_i, ξ_i^* are positive slack variables to penalize the positive and negative errors and I_1 and I_2 are respectively the sets of samples that are in the quadratic and linear loss zone.

Three parameters (C, ϵ, δ) need to be tuned for linear SVR (also the kernel width for the Gaussian case).

2.2. Principal component regression

PCR [4] performs a Principal Component Analysis (PCA) on the zero-mean $n \times d$ predictor matrix $\mathbf{X}$ and then applies Ordinary Least Squares (OLS) to the resulting g principal components (scores) $\mathbf{T} = [\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_g]$ and the zero-mean $n \times k$ dependent variables $\mathbf{Y}$. $\mathbf{X}$ can be factorized as:

$$\mathbf{X} = \mathbf{T}\mathbf{P}^T$$

where $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_g]$ is the loadings matrix ($\mathbf{P}\mathbf{P}^T = \mathbf{I}$) and the equality is for $g = r$ (r is the rank of $\mathbf{X}$). Any number $g \leq r$ of components can be selected. However, these components should be selected carefully: [4] proposes using the variance of the components or the dependence (correlation with $\mathbf{Y}$) depending on the purpose of the analysis but the author warns that low variance components can be useful for predicting $\mathbf{Y}$.

Predictions for $\mathbf{X}$ and $\mathbf{X}_t$ are done using the following:

$$\mathbf{Y} = \mathbf{T}\mathbf{B} + \mathbf{F} \qquad (2)$$
$$\hat{\mathbf{B}} = (\mathbf{T}^T\mathbf{T})^{-1}\mathbf{T}^T\mathbf{Y} \qquad (3)$$
$$\hat{\mathbf{Y}} = \mathbf{T}\hat{\mathbf{B}} = \mathbf{X}\mathbf{P}\hat{\mathbf{B}} \qquad (4)$$
$$\hat{\mathbf{Y}}_t = \mathbf{X}_t\mathbf{P}\hat{\mathbf{B}} \qquad (5)$$

where (2) is the regression model based on the principal components; (3) is the OLS estimator for $\hat{\mathbf{B}}$; and, (4) and (5) are the predictions.

The nonlinear kernel PCR (KPCR) as described in [5] performs kernel PCA [6] and then OLS using the principal components in the feature space. The same concerns appear in KPCR as in PCR, low variance components could be useful for prediction.

One parameter (g) needs to be tuned for PCR (also the kernel width for Gaussian KPCR).

2.3. Partial least squares

PLS for regression (see [3, 7] for a description and an overview) models the zero-mean $n \times d$ predictor matrix $\mathbf{X}$ and the zero-mean $n \times k$ dependent variables $\mathbf{Y}$ as:

$$\mathbf{X} = \mathbf{T}\mathbf{P}^T + \mathbf{E} \qquad (6)$$
$$\mathbf{Y} = \mathbf{U}\mathbf{Q}^T + \mathbf{F} = \mathbf{T}\mathbf{D}\mathbf{Q}^T + \mathbf{F}^* = \mathbf{T}\mathbf{C}^T + \mathbf{F}^*, \quad (7)$$

where $\mathbf{T} = [\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_g]$ and $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_g]$ contain the extracted latent vectors (scores) as columns and g is the number of latent components extracted; $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_g]$ and $\mathbf{Q} = [\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_g]$ are the loading matrices; and, $\mathbf{E}$ and $\mathbf{F}$ represent residual matrices. The second equality in (7) comes from the assumption of $\mathbf{T}$ columns being good predictors of $\mathbf{Y}$ and $\mathbf{U} = \mathbf{T}\mathbf{D} + \mathbf{H}$, where $\mathbf{D}$ is a diagonal matrix. In order to find the score vectors $\mathbf{t}_i, \mathbf{u}_i$ a procedure based on NIPALS algorithm [8] finds $\mathbf{w}_i, \mathbf{c}_i$ vectors such that

$$\text{cov}(\mathbf{t}_i, \mathbf{u}_i)^2 = \text{cov}(\mathbf{X}\mathbf{w}_i, \mathbf{Y}\mathbf{c}_i)^2 = \max_{\substack{||\mathbf{r}||=1 \\ ||\mathbf{s}||=1}} \text{cov}(\mathbf{X}\mathbf{r}, \mathbf{Y}\mathbf{s})^2$$

The procedure starts by a random selection of $\mathbf{u}_i$ vector in the $\mathcal{Y}$-space ($\mathbf{u}_i = \mathbf{Y}$ if $k = 1$). Then, the following steps are repeated until convergence.

1.- $\mathbf{w}_i = \mathbf{X}^T\mathbf{u}_i/||\mathbf{X}^T\mathbf{u}_i||$ 3.- $\mathbf{c}_i = \mathbf{Y}^T\mathbf{t}_i/||\mathbf{Y}^T\mathbf{t}_i||$

2.- $\mathbf{t}_i = \mathbf{X}\mathbf{w}_i$ 4.- $\mathbf{u}_i = \mathbf{Y}\mathbf{c}_i$

$\mathbf{w}_i, \mathbf{t}_i, \mathbf{c}_i$ and $\mathbf{u}_i$ can also be found by solving eigenvalue problems [9]. The loadings can be computed as

$$\mathbf{p}_i = \mathbf{X}^T\mathbf{t}_i/(\mathbf{t}_i^T\mathbf{t}_i) \qquad \mathbf{q}_i = \mathbf{Y}^T\mathbf{u}_i/(\mathbf{u}_i^T\mathbf{u}_i)$$

Then, $\mathbf{X}$ and $\mathbf{Y}$ are deflated as follows:

$$\mathbf{X} = \mathbf{X} - \mathbf{t}_i\mathbf{p}_i^T \qquad (8)$$
$$\mathbf{Y} = \mathbf{Y} - \mathbf{t}_i\mathbf{t}_i^T\mathbf{Y}/(\mathbf{t}_i^T\mathbf{t}_i) = \mathbf{Y} - \mathbf{t}_i\mathbf{c}_i^T \qquad (9)$$

and the algorithm is repeated until the required number of components is reached (see [7] for stopping criteria).

The estimated $\hat{\mathbf{Y}}$ and $\hat{\mathbf{Y}}_t$ for the training data $\mathbf{X}$ and test data $\mathbf{X}_t$ are respectively given by:

$$\mathbf{B} = \mathbf{W}(\mathbf{P}^T\mathbf{W})^{-1}\mathbf{C}^T \tag{10}$$

$$\hat{\mathbf{Y}} = \mathbf{X}\mathbf{B} = \mathbf{T}\mathbf{C}^T \tag{11}$$

$$\hat{\mathbf{Y}}_t = \mathbf{X}_t\mathbf{B} \tag{12}$$

where $\mathbf{W}$, $\mathbf{P}$ and $\mathbf{C}$ respectively are the matrices with $\mathbf{w}_i$, $\mathbf{c}_i$ and $\mathbf{p}_i$ as columns.

To extend PLS to nonlinear problems, it is common to apply the kernel-trick [10] to the predictor matrix $\mathbf{X}$ and then perform a linear PLS in the feature space [5]. The NIPALS-based kernel version results:

1.- $\quad \mathbf{t}_i = \mathbf{K}\mathbf{u}_i/||\mathbf{K}\mathbf{u}|| \qquad$ 3.- $\quad \mathbf{u}_i = \mathbf{Y}\mathbf{c}_i/||\mathbf{Y}\mathbf{c}_i||$

2.- $\quad \mathbf{c}_i = \mathbf{Y}^T\mathbf{t}_i$

where $\mathbf{K}$ is the kernel matrix of the training data.

One parameter (g) needs to be tuned for PLS (also the kernel width for Gaussian KPLS).

3.   Results and conclusion

We compared the performance of the following methods: SVR with linear kernel (SVR-LIN), non-linear SVR (SVR-RBF), PCR selecting the components to keep using their variance (PCR) or their correlation with $\mathbf{Y}$ (SPCR), Kernel PCR (KPCR), Kernel SPCR (KSPCR), PLS and Kernel PLS (KPLS). We train and evaluate these methods using a leave-one-individual-out setting (a leave-one-out where all the samples of one individual are tested together). Following best practices, test samples were used only once, neither for training nor for parameter selection, which was done using linear grid search. Table 1 shows a comparison of the error for non-linear and linear methods. We observe that non-linear methods outperform linear methods in the prediction of τ and the opposite when predicting E_{max}. SVR achieves better performance both in linear and non-linear cases. Due to the small number of pigs, no significant statistical difference (Mann-Whitney-Wilcoxon paired test) was found for E_{max} neither between SVR-LIN and the following four best methods nor between linear and non-linear counterparts. When predicting τ, statistically significant differences where only found among SVR-RBF vs KPCR and KSPCR vs SPCR.

Figure 1 (a) and (b) show the estimation and Bland-Altman (BA) plots for E_{max} using a linear SVR. We observe that the estimated values follow the measured values for all pigs yet not being able to reach extreme values as we see both in (a) and in the BA plot (almost all values of E_{max} greater than 10 mmHg/ml are above the one-standard-deviation line). The case of τ shown in Figure 1 (c) and (d) is completely different, as the estimation is harder. We observe that the estimation provided by the

Method	E_{max} (mmHg/ml)	τ (ms)
SVR-LIN	**0.92** $\pm$ 0.32	9.8 $\pm$ 4.8
SVR-RBF	0.94 $\pm$ 0.39	**8.9** $\pm$ 3.7
PCR	0.95 $\pm$ 0.30	11.0 $\pm$ 5.6
KPCR	1.02 $\pm$ 0.32	9.8 $\pm$ 3.9
SPCR	0.97 $\pm$ 0.32	11.8 $\pm$ 5.1
KSPCR	0.97 $\pm$ 0.42	9.2 $\pm$ 3.3
PLS	0.93 $\pm$ 0.32	10.7 $\pm$ 5.3
KPLS	0.94 $\pm$ 0.34	9.4 $\pm$ 3.5

Table 1: Mean, $\bar{E}$, of the mean absolute error for all pigs.

SVR-RBF is coarser than in the case of E_{max} and a large portion of one pig is predicted using a constant value. In addition, the BA plot shows an almost-linear systematic error in the prediction which could be easily removed to improve the results.

We conclude that non-linear methods for estimation in the case of a high-dimensional input space should not be ruled out. For our data, they were outperformed in the easier estimation problem but they did better in the harder estimation problem. More research is going to be devoted to the conditions that make more suitable the non-linear approach. In addition, easier to tune PCR and PLS methods almost attain the same performance than SVR.

Acknowledgements

References

[1] Rojo-Alvarez J, Bermejo J, Juarez-Caballero V, Yotti R, Cortina C, Garcia-Fernandez M, Antoranz J. Support vector analysis of color-Doppler images: a new approach for estimating indices of left ventricular function. IEEE Transactions on Medical Imaging 2006;25(8):1037–1043.

[2] Smola AJ, Schölkopf B. A tutorial on support vector regression. Statistics and Computing 2004;14(3):199–222.

[3] Rosipal R, Krämer N. Overview and Recent Advances in Partial Least Squares. In Saunders C, Grobelnik M, Gunn S, Shawe-Taylor J (eds.), Subspace, Latent Structure and Feature Selection, volume 3940 of Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg. ISBN 978-3-540-34137-6, 2006; 34–51.

[4] Massy WF. Principal Components Regression in Exploratory Statistical Research. Journal of the American Statistical Association 1965;60:234–256.
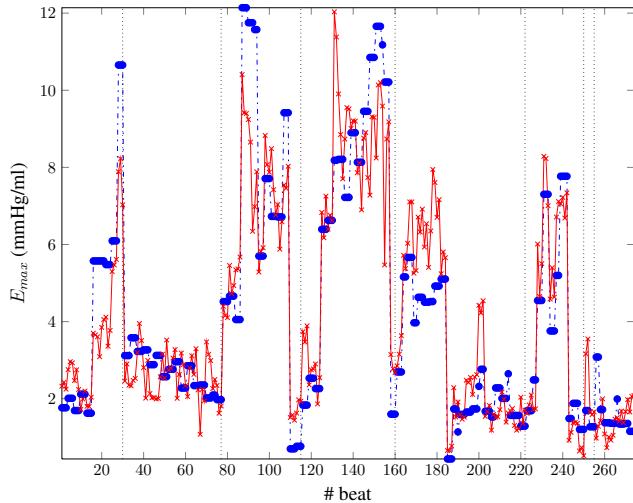
(a) E_{max} Estimation.
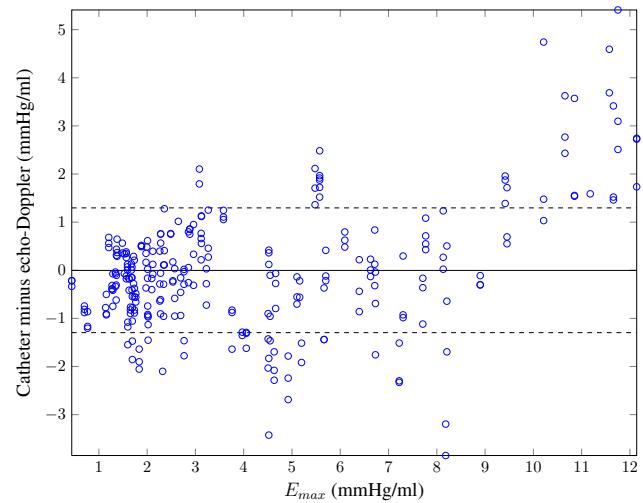
(b) E_{max} Bland-Altman.
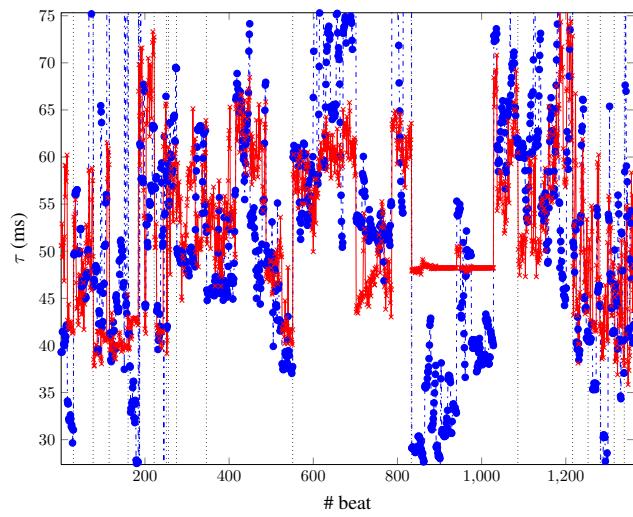
(c) τ Estimation.
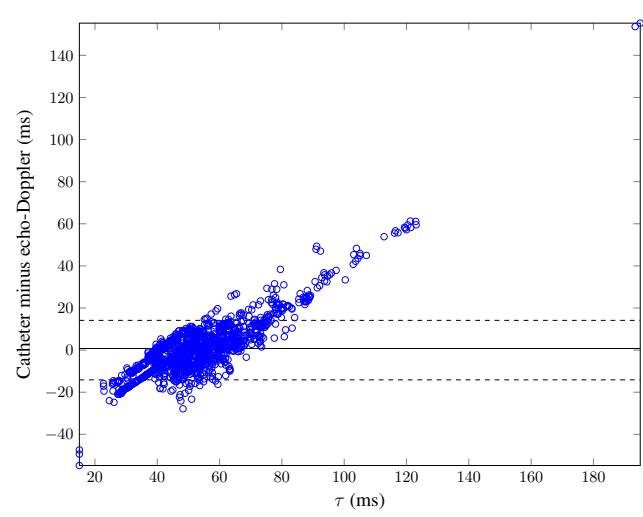
(d) τ Bland-Altman.

Figure 1: Prediction and Bland-Altman plots for E_{max} and τ. The blue line in (a) and (c) represents the catheter measure, the red line the predictions and the vertical lines divide the beats that correspond to each individual. Bland-Altman plots represent the catheter measure versus the catheter minus the estimation and the dotted lines represent one standard deviation of this difference.

[5] Rosipal R, Trejo LJ. Kernel partial least squares regression in reproducing kernel hilbert space. The Journal of Machine Learning Research 2001;2:97–123.

[6] Schölkopf B, Smola AJ, Müller KR. Kernel Principal Component Analysis. In Schölkopf B, Burges CJC, Smola AJ (eds.), Advances in Kernel Methods—Support Vector Learning. Cambridge, MA: MIT Press, 1999; 327–352.

[7] Wold S, Sjöström M, Eriksson L. PLS-regression: a basic tool of chemometrics. Chemometrics and Intelligent Laboratory Systems 2001;58(2):109–130.

[8] Wold H. Nonlinear Estimation by Iterative Least Squares Procedures. In David F (ed.), Researh Papers in Statistics.

Wiley, New York, 1966; 411–444.

[9] Höskuldsson A. PLS regression methods. Journal of Chemometrics 1988;2(3):211–228.

[10] Schölkopf B, Smola A. Learning with Kernels. Cambridge, MA, USA: MIT Press, 2001.

Address for correspondence:

Ricardo Santiago-Mozos
Universidad Rey Juan Carlos, 28943 MADRID, SPAIN
ricardo.santiago.mozos@urjc.es