Applying Pairing Support Vector Regression Algorithm to GPS GDOP Approximation

Pei-Yi Hao[a] Chao-Yi Wu[b]
Department of Information Management
National Kaohsiung University of Applied Sciences, Kaohsiung, Taiwan
Email: haupy@cc.kuas.edu.tw[a] wububbles@gmail.com[b]

***Abstract*—Global Positioning System (GPS) has extensively been employed in various applications, including the use of GPS to analyze the cognitive diseases and find better treatment. Geometric Dilution of Precision (GDOP) is an indicator showing how well the constellation of GPS satellites is geometrically organized. GPS positioning with a smaller GDOP value usually yields better accuracy. However, the calculation of GDOP is a time- and power-consuming task that requires to solving measurement equations with complicated matrix transformation and inversion. When selecting the one with the lowest GDOP for positioning from many GPS constellations, methods that can fast and accurately calculate GPS GDOP are imperative. Previous works have shown that numerical regression on GPS GDOP can yield satisfactory results and eliminate many calculation steps. This paper employs a new pairing support vector regression algorithm (pair-SVR) to the approximation of GPS GDOP. The pair-SVR determines indirectly the regression function through a pair of nonparallel insensitive upper- and lower-bound functions, each of which is solved by support vector machine (SVM)- type quadratic programming problems (QPP) with smaller-sized. This strategy makes the pair-SVR not only have the faster learning speed than the classical SVR, but also be suitable for many cases, especially when the noise is heteroscedastic. Besides, pair-SVR improves the sparsity than that of twin support vector regression (TSVR) by employing the concept of insensitive zone. This makes the prediction time complexity of pair-SVR is obviously smaller than TSVR. The experimental results show that pair-v-SVR gains better performance for the approximation of GPS GDOP than previous support vector regression machine.**

Keywords- GDOP; GPS; kernel-based method; support vector machine; support vector regression.

I. INTRODUCTION

Cognitive impairment manifests in changed out-of-home mobility. Until recently, the assessment of outdoor mobility relied on the reports of family care-givers and institutional staff and used observational approaches, activity monitoring or behavioural checklists. Shoval et al. [1] apply GPS to analyze the mobility of the people who have Alzheimer's disease and related cognitive diseases. Shoval et al. [2] apply GPS to measure the out-of-home mobility of older adults with differing cognitive functioning. The GPS is a satellite based navigation system that helps users to determine their locations on Earth. A GPS receiver compares the time difference between the signal transmitted by a satellite and the time it was received and calculates the distance between the satellite and GPS receiver. GPS receivers analyze such signals from at least 3 satellites and use triangulation to determine the user's location. GPS, which consists of at least 24 active satellites provides 24-hour, all-weather, worldwide coverage with position, velocity and timing information. Nowadays, GPS has become a popular tool for positioning and navigation. However, the accuracy of GPS positing may unavoidably degrade by two causes [3]: the errors in each observable signal, and the geometry formed by the observables employed for positioning or navigation. Reasons resulting in the former factor include ionospheric delay, tropospheric delay, satellite clock and receiver clock offsets, receiver noise and multi-path problems. The later one is usually referred to as the GDOP, which describes the effect of geometry on the relationship between measurement error and position determination error.

GDOP is an indicator showing how well the constellation of GPS satellites is organized geometrically. Because some receiver device may be restricted to processing a limited number of visible satellites, hence, it needs to select the satellite subset that offers the best or most acceptable solution. Since GDOP provides a simple interpretation of how much positioning precision can be diluted by a unit of measurement error, positioning or navigation can obtain a better quality by choosing the combination of satellites in a satellite constellation with GDOP as small as possible.

Existing methods for calculating GDOP include matrix inversion, closed-form algorithms, maximum volume of tetrahedrons, etc. The most accurate method for determining GDOP is to use matrix inversion to all combinations and select the minimum one. However, this approach is a time- and power- consuming task because it usually requires considerable computational power and a large amount of operations for exhaustively examining all possible combinations of satellites. It would be a computational burden for real time application and mobile device. Closed form methods simplify the computational procedure under specific circumstances, but there still has roundoff errors due to the floating-point operations. Instead of directly calculating the GDOP equations and avoiding the complicated solving of matrix inversion, Simon and El-Sherief [4][5] rephrase the calculation of GDOP as regression/ approximation problems and apply neural networks (NN) to solve such problems. However, solving regression problems using NN usually suffer from the slow training speed and difficulty in determining the network architecture. Besides, the overfitting problem degrades the generalization ability of NN applications when the numbers of features and training samples are large. Wu first employs the support vector regression machine for the approximation of GPS GDOP [6].

The SVMs have been very successful in many fields. Peng proposed a TSVR for data regression [9]. In TSVR, a pair of smaller sized QPPs is solved rather than the large single QPP in the SVR. This strategy makes the training speed of TSVR is faster than classical SVR machine. However, the major disadvantage of TSVR is its prediction speed is significantly slow due to the loss of sparsity. In TSVR, the number of basis function used for estimating the final regression function is equal to the number of training samples. Therefore, predicting using TSVR is a time-consuming task for large-scale data set. In many real applications, the prediction speed is more important than training speed. Hence, it is necessary to improve the sparsity of TSVR, since a sparse regression model means a low economy for storage requirement and a high efficiency for the real time prediction.

In the spirit of TSVR, this paper proposes a novel pair-SVR, which seeks a pair of nonparallel bound function of regression model by solving two related SVM-type problems, each of which is smaller than conventional SVM. The major benefit of the proposed pair-SVR is the efficiency for both learning and prediction. We improve the sparsity of TSVR by adopting an insensitive zone that is determined by a pair of nonparallel upper bound and lower bound function. Only samples outside the insensitive zone are captured as SVs, and only those SVs construct the final regression model. In general, the number of SV is very few. This makes the prediction time cost of pair-SVR is obviously smaller than TSVR. Besides, the strategy of solving two QPPs with smaller-sized instead of a single large QPP makes the training time complexity of pair-SVR approximately 4 times smaller than that of a classical SVR.

The rest of this paper is organized as follows. Section II gives a brief overview of GDOP and TSVR. Section III describes a modification of TSVR, called pair-SVR, and applies pair-SVR for GDOP approximation. Experiments are presented in Section IV, and some concluding remarks are given in Section V.

II. BACKGROUND

A. Geometric Dilution of Precision

In GPS applications, the GDOP is often used to select a subset of satellites from all visible ones. In order to determine the position of a receiver, pseudoranges from n (≥4) satellites must be used at the same time. By linearizing the pseudorange equation with Taylor's series expansion at the approximate (or nominal) receiver position, the relationship between pseudorange difference ($\Delta\rho_i$) and positioning difference (Δx_i) can be summarized as follows [2]:

$$\begin{bmatrix} \Delta\rho_1 \\ \Delta\rho_2 \\ \vdots \\ \Delta\rho_n \end{bmatrix} = \begin{bmatrix} e_{11} & e_{12} & e_{13} & 1 \\ e_{21} & e_{22} & e_{23} & 1 \\ & \vdots & & 1 \\ e_{n1} & e_{n2} & e_{n3} & 1 \end{bmatrix} \begin{bmatrix} \Delta x_u \\ \Delta y_u \\ \Delta z_u \\ c\Delta t_b \end{bmatrix} + \begin{bmatrix} v_{\rho 1} \\ v_{\rho 2} \\ \vdots \\ v_{\rho n} \end{bmatrix} \quad (1)$$

Equation (1) can have a general form represented as

$$\mathbf{z} = \mathbf{H}\mathbf{x} + \mathbf{v} \quad (2)$$

where the geometry matrix $\mathbf{H}$ is n × 4, n ≥ 4, since it is necessary to use at least 4 satellites to determine a position in a 3-dimension space. Such an overdetermined system like (2) has no exact solution. However, the n columns of $\mathbf{H}$ are linearly independent since they are signals received from individual satellites independently. The linear least squares solution can be obtained by solving the normal equations

$$\mathbf{H}'\mathbf{z} = \mathbf{H}'\mathbf{H}\mathbf{x} + \mathbf{H}'\mathbf{v} \quad (3)$$

$\mathbf{H}$ has full rank and $\mathbf{M} = \mathbf{H}'\mathbf{H}$ is invertible, then we can have

$$\hat{\mathbf{x}} = (\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}'\mathbf{z} \quad (4)$$

GDOP becomes a linear least-squares solution of a linearized pseudorange equation by taking the difference between the estimated and the true positions. Therefore

$$\tilde{\mathbf{x}} = (\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}'\mathbf{v} \quad (5)$$

The quality of this solution is evaluated by cov(·), which denotes the covariance of a measurement.

$$\mathrm{cov}(\hat{\mathbf{x}}) = (\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}'\,\mathrm{cov}(\mathbf{v})\left((\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}'\right)' \quad (6)$$

If all components of $\mathbf{v}$ are pairwise uncorrelated and have variance σ^2, cov($\mathbf{v}$) can be normalized to an identity matrix cov($\mathbf{v}$)=σ^2I, and a simplified expression of (6) can be obtained as

$$\mathrm{cov}(\hat{\mathbf{x}}) = \sigma^2(\mathbf{H}'\mathbf{H})^{-1} \quad (7)$$

This quantifies the magnification of pseudorange errors onto the user position errors. Let $\mathbf{M} = \mathbf{H}'\mathbf{H}$ be the measurement matrix. The GDOP factor is defined as

$$GDOP = \sqrt{trace(\mathbf{H}'\mathbf{H}^{-1})} = \sqrt{\frac{trace\left[adj(\mathbf{H}'\mathbf{H})\right]}{\det(\mathbf{H}'\mathbf{H})}} \quad (8)$$

B. Twin Support Vector Regression

The TSVR finds a pair of nonparallel functions around the data points [9]. In general, it considers the following pair of functions for the nonlinear case:

$$f_1(\mathbf{x}) = \mathbf{w}_1^T\mathbf{K}(\mathbf{A},\mathbf{x}) + b_1 \text{ and } f_2(\mathbf{x}) = \mathbf{w}_2^T\mathbf{K}(\mathbf{A},\mathbf{x}) + b_2$$

each one determines the ε -insensitive down- or up-bound function, respectively. The functions $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$ are obtained by solving the following pair of QPPs:

$$\underset{\mathbf{w}_1, b_1}{\text{minimize}} \quad \frac{1}{2}\left\|\mathbf{Y} - \mathbf{e}\varepsilon_1 - (\mathbf{K}\mathbf{w}_1 + \mathbf{e}b_1)\right\|^2 + \frac{C_1}{N}\mathbf{e}^T\boldsymbol{\xi} \quad (9)$$
$$\text{subject to} \quad \mathbf{Y} - (\mathbf{K}\mathbf{w}_1 + \mathbf{e}b_1) \geq \mathbf{e}\varepsilon_1 - \boldsymbol{\xi}, \quad \boldsymbol{\xi} \geq 0$$

$$\underset{\mathbf{w}_1, b_1}{\text{minimize}} \quad \frac{1}{2}\left\|\mathbf{Y} - \mathbf{e}\varepsilon_2 - (\mathbf{K}\mathbf{w}_2 + \mathbf{e}b_2)\right\|^2 + \frac{C_2}{N}\mathbf{e}^T\boldsymbol{\xi} \quad (10)$$
$$\text{subject to} \quad \mathbf{Y} - (\mathbf{K}\mathbf{w}_2 + \mathbf{e}b_2) \geq \mathbf{e}\varepsilon_2 - \boldsymbol{\xi}, \quad \boldsymbol{\xi} \geq 0$$

where ε_1, $\varepsilon_2 \geq 0$ are the insensitive parameters. C_1, $C_2 \geq 0$ are the regularization parameters. $\mathbf{e}$ are vector of ones of N dimensions. Y is the target vector $\mathbf{Y}=(y_1,..,y_N)^T$. $\boldsymbol{\xi}$ is the slack vector $\boldsymbol{\xi}=(\xi_1,.., \xi_N)^T$. $\mathbf{K}(\mathbf{A},\mathbf{x})$ is the column vector $(k(\mathbf{A}_1,\mathbf{x}),...,k(\mathbf{A}_N,\mathbf{x}))^T$ where $\mathbf{A}_i$ are the ith training sample (row vector). $\mathbf{K}$ is the N by N kernel matrix such that $\mathbf{K}_{ij}=k(\mathbf{A}_i,\mathbf{A}_j)$. By considering the Karush-Kuhn-Tucker (KKT) conditions for the Lagrangian functions of (9) and 10), we obtain the dual QPPs, which are

$$\begin{aligned} &\max \quad -\frac{1}{2}\boldsymbol{\alpha}^T\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\boldsymbol{\alpha}+\mathbf{f}^T\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\boldsymbol{\alpha}-\mathbf{f}^T\boldsymbol{\alpha} \\ &\text{s.t.} \quad 0 \leq \boldsymbol{\alpha} \leq \frac{C_1}{N}\mathbf{e} \end{aligned} \tag{11}$$

and

$$\begin{aligned} &\max \quad -\frac{1}{2}\boldsymbol{\beta}^T\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\boldsymbol{\beta}-\mathbf{h}^T\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\boldsymbol{\beta}+\mathbf{h}^T\boldsymbol{\beta} \\ &\text{s.t.} \quad 0 \leq \boldsymbol{\beta} \leq \frac{C_2}{N}\mathbf{e} \end{aligned} \tag{12}$$

where $\mathbf{H}=[\mathbf{K}\ \mathbf{e}]$, $\mathbf{f}=\mathbf{Y}-\varepsilon_1\mathbf{e}$, and $\mathbf{h}=\mathbf{Y}+\varepsilon_2\mathbf{e}$. After optimizing (11) and (12), we obtain the augmented vectors for $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$, which are

$$\begin{bmatrix}\mathbf{w}_1\\ b_1\end{bmatrix}=(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T(\mathbf{f}-\boldsymbol{\alpha}) \quad \begin{bmatrix}\mathbf{w}_2\\ b_2\end{bmatrix}=(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T(\mathbf{h}+\boldsymbol{\beta}) \tag{13}$$

Then, the estimated regressor is constructed by as follows:

$$\begin{aligned} f(\mathbf{x}) &= \frac{1}{2}(f_1(\mathbf{x})+f_2(\mathbf{x})) \\ &= \frac{1}{2}(\mathbf{w}_1^T+\mathbf{w}_2^T)\mathbf{K}(\mathbf{A},\mathbf{x})+\frac{1}{2}(b_1+b_2) \end{aligned} \tag{14}$$

III. GDOP APPROXIMATION USING PAIRING SUPPORT VECTOR REGRESSION ALGORITHM

A. Pairing Support Vector Regression Algorithm

Motivated by TSVM, the goal of the proposed pair-SVR algorithm is to estimate a pair of nonparallel function $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$ by solving two SVM type QPPs with smaller size, each of which determines the upper bound and lower bound of the insensitive zone, such that the insensitive zone includes all training samples with smallest size. According to the concept of kernel-based learning, a non-linear function is obtained via a linear learning machine in a kernel-introduced feature space while the capacity of the learning machine is controlled by a parameter that is independent to the dimensionality of the space. The basic concept is that a nonlinear regression function is estimated via simply mapping the training data vector $\mathbf{x}_i$ by Φ: $R^n \rightarrow$ F into a high-dimensional feature space F. Therefore, the proposed pair-SVR aims at estimating the following two functions:

$$f_1(\mathbf{x})=\langle\mathbf{w}_1\cdot\Phi(\mathbf{x})\rangle+b_1, \text{ where } \mathbf{w}\in F,\ \mathbf{x}\in R^n,\ b\in R,$$

$$f_2(\mathbf{x})=\langle\mathbf{w}_2\cdot\Phi(\mathbf{x})\rangle+b_2, \text{ where } \mathbf{w}\in F,\ \mathbf{x}\in R^n,\ b\in R$$

For the estimation of $f_1(\mathbf{x})=\langle\mathbf{w}_1\cdot\Phi(\mathbf{x})\rangle+b_1$, the upper bound function of the insensitive zone, we force the upper bound function $f_1(\mathbf{x})$ to move downward via minimizing $\|\mathbf{w}_1\|^2$ and b_1 in the objective function, and requires all training data $(\mathbf{x}_i, y_i)$ to be below the upper bound function in the constraint, simultaneously. Hence, the problem of estimating the $\mathbf{w}_1$ and b_1 is equivalent to solve the following QPP:

$$\underset{w_1,b_1,\xi_{1i}}{\text{minimize}}\ \frac{1}{2}\|\mathbf{w}_1\|^2+b_1+C_1\sum_{i=1}^{N}\xi_{1i} \tag{15}$$

subject to

$$\langle\mathbf{w}_1\cdot\Phi(\mathbf{x}_i)\rangle+b_1 \geq y_i-\xi_{1i}$$

and $\xi_{1i}\geq 0$ for i=1,…,N.

We can find the solution of this QPP in dual variables by finding the saddle point of the Lagrangian:

$$\begin{aligned} L &= \frac{1}{2}\|\mathbf{w}_1\|^2+b_1+C_1\sum_{i=1}^{N}\xi_{1i} \\ &-\sum_{i=1}^{N}\alpha_{1i}\left[\langle\mathbf{w}_1\cdot\Phi(\mathbf{x}_i)\rangle+b_1-y_i+\xi_{1i}\right]-\sum_{i=1}^{N}\beta_{1i}\xi_{1i} \end{aligned} \tag{16}$$

where α_{1i} and β_{1i} are the nonnegative Lagrange multipliers. Differentiating L with respect to $\mathbf{w}_1$, b_1 and ξ_{1i} and setting the result to zero, we obtain:

$$\partial L/\partial\mathbf{w}_1=0 \Rightarrow \mathbf{w}_1=\sum_{i=1}^{N}\alpha_{1i}\Phi(\mathbf{x}_i), \tag{17}$$

$$\partial L/\partial b_1=0 \Rightarrow \sum_{i=1}^{N}\alpha_{1i}=1, \tag{18}$$

$$\partial L/\partial\xi_{1i}=0 \Rightarrow \alpha_i=C_1-\beta_{1i} \text{ and } \alpha_{1i}\leq C_1, \tag{19}$$

Substituting Eqs. (17)-(19) into L, we obtain the following dual problem

$$\max \quad \frac{-1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_{1i}\alpha_{1j}\langle\Phi(\mathbf{x}_i)\cdot\Phi(\mathbf{x}_j)\rangle+\sum_{i=1}^{N}\alpha_{1i}y_i \tag{20}$$

$$\text{subject to} \quad \sum_{i=1}^{N}\alpha_{1i}=1, \text{ and } \alpha_{1i}\in[0,C_1].$$

Parameter b_1 can be calculated from the KKT conditions:

$$\alpha_{1i}\left[\langle\mathbf{w}_1\cdot\Phi(\mathbf{x}_i)\rangle+b_1-y_i+\xi_{1i}\right]=0, \tag{21}$$

$$(C_1-\alpha_{1i})\xi_{1i}=0 \tag{22}$$

For some $\alpha_{1i}\in(0,C_1)$, we have $\xi_{1i}=0$ and moreover the second factor in (21) equals to zero. Hence, b_1 can be calculated as follows:

$$b_1=y_i-\langle\mathbf{w}_1\cdot\Phi(\mathbf{x}_i)\rangle \text{ for some } \alpha_{1i}\in(0,C_1).$$

Finally, the upper bound function of the regression model is

$$f_1(\mathbf{x})=\sum_{i=1}^{N}\alpha_{1i}k(\mathbf{x},\mathbf{x}_i)+b_1. \tag{23}$$

For the estimation of $f_2(\mathbf{x}) = \langle \mathbf{w}_2 \cdot \Phi(\mathbf{x}) \rangle + b_2$, the lower bound function of the insensitive zone, intuitively, we should force $f_2(\mathbf{x})$ to move upward via maximizing $\|\mathbf{w}_2\|^2$ and b_2 in the objective function, and requires$ all training data $(\mathbf{x}_i, y_i)$ to be above the lower bound function in the constraint, simultaneously. However, maximizing $\|\mathbf{w}_2\|^2$ violates the principle of sparsity regression model. Hence, we apply the following trick to estimate the lower bound function. First, we multiplies the desired target y_i by -1 and estimates a mirroring function of $f_2(\mathbf{x})$. We force the mirroring function $\bar{f}_2(\mathbf{x}) = \langle \bar{\mathbf{w}}_2 \cdot \Phi(\mathbf{x}) \rangle + \bar{b}_2$ to move downward, and require all instances $(\mathbf{x}_i, -y_i)$ to be below the mirroring function, simultaneously. Finally, the lower bound function is $f_2(\mathbf{x}) = -\bar{f}_2(\mathbf{x})$. The problem for seeking $\bar{f}_2(\mathbf{x}) = \langle \bar{\mathbf{w}}_2 \cdot \Phi(\mathbf{x}) \rangle + \bar{b}_2$ is equivalent to solve the following optimization problem:

$$\underset{\bar{w}_2, \bar{b}_2, \xi_{2i}}{\text{minimize}} \ \frac{1}{2}\|\bar{\mathbf{w}}_2\|^2 + \bar{b}_2 + C_1 \sum_{i=1}^{N} \xi_{2i} \quad (24)$$

$$\text{subject to } \langle \bar{\mathbf{w}}_2 \cdot \Phi(\mathbf{x}_i) \rangle + \bar{b}_2 \geq -y_i - \xi_{2i}$$

$$\text{and } \xi_{2i} \geq 0 \text{ for } i=1,\ldots,N.$$

Similar to the above Lagrange multipliers substituting procedure, we obtain its dual problem as

$$\max \ \frac{-1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_{2i}\alpha_{2j} \langle \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j) \rangle - \sum_{i=1}^{N} \alpha_{2i} y_i \quad (26)$$

$$\text{subject to } \sum_{i=1}^{N} \alpha_{2i} = 1, \text{ and } \alpha_{2i} \in [0, C_1].$$

After solving (26), we obtain the weight vector $\bar{\mathbf{w}}_2 = \sum_{i=1}^{N} \alpha_{2i} \Phi(\mathbf{x}_i)$. While parameter b_2 can be calculated from the KKT conditions:

$$\alpha_{2i}\left[\langle \bar{\mathbf{w}}_2 \cdot \Phi(\mathbf{x}_i) \rangle + \bar{b}_2 + y_i + \xi_{2i} \right] = 0, \quad (27)$$

$$(C_2 - \alpha_{2i})\xi_{2i} = 0 \quad (28)$$

For some $\alpha_{2i} \in (0, C_2)$, we have $\xi_{2i} = 0$ and moreover the second factor in (27) equals to zero. Hence, b_2 can be calculated as follows:

$$\bar{b}_2 = -y_i - \langle \bar{\mathbf{w}}_2 \cdot \Phi(\mathbf{x}_i) \rangle \text{ for some } \alpha_{2i} \in (0, C_2).$$

Finally, the lower bound function of the regression model is

$$f_2(\mathbf{x}) = -\bar{f}_2(\mathbf{x}) = -\sum_{i=1}^{N} \alpha_{2i} k(\mathbf{x}, \mathbf{x}_i) - \bar{b}_2. \quad (29)$$

After seeking the upper bound and lower bound function, the final regression function is obtained as follows

$$\begin{aligned} f(\mathbf{x}) &= \frac{1}{2}\left(f_1(\mathbf{x}) + f_2(\mathbf{x})\right) \\ &= \frac{1}{2}\sum_{i=1}^{N}(\alpha_{1i} - \alpha_{2i})k(\mathbf{x}, \mathbf{x}_i) + \frac{1}{2}(b_1 - \bar{b}_2) \end{aligned} \quad (30)$$

The training samples with corresponding α_{1i}, α_{2i}>0 are called support vectors since only those data vectors construct the final regression function. Noted, according to the KKT conditions (17), (18), (22), and (23), only points outside the insensitive zone (or lying on the upper or lower bound function of the insensitive zone) are captured as support vectors. Usually, the number of support vector is very few. Hence, pair-SVR significantly enhances the sparsity than that of TSVR.

B. GDOP approximation by pair-SVR

In a complex system, if the input–output values were most concerned, rather than how their relationships are organized, numeric regression for fitting input–output data provides a efficient solution. Previous studies have reported that numerical regression on GPS GDOP can yield satisfactory results and reduce many calculation steps [4-6]. This paper apply the proposed pair-SVR algorithm as a means for the approximation of GPS GDOP.

In (8), GDOP is mainly evaluated by $(\mathbf{H}'\mathbf{H})^{-1}$. Nevertheless, it is not a good practice to invert the normal equations matrix. If the matrix is well-conditioned and positive definite, that is, it has full rank, the normal equations in (8) can be solved directly by using the Cholesky decomposition [10], $\mathbf{H}'\mathbf{H} = \mathbf{R}'\mathbf{R}$, where R is an upper triangular matrix. It gives

$$\mathbf{R} = \begin{bmatrix} h_1 & h_2 & h_3 & h_4 \\ & h_5 & h_6 & h_7 \\ & & h_8 & h_9 \\ \multicolumn{3}{l}{symmetrical} & h_{10} \end{bmatrix} \quad (31)$$

where $h_k = (\mathbf{H}'\mathbf{H})_{ij}$, $1 \leq i \leq j \leq 4$, $k = 1,\ldots, 10$. A regression problem is formed as a functional mapping $R^{10} \rightarrow R^1$ with 10 inputs from $h_1, h_2, \ldots, h_{10}$ and one output for GDOP. Based on these settings, pair-SVR can be trained with a set D of input-output pairs $d_i \in$ D, $d_i = ((h_1, h_2, \cdots h_{10})_i, GDOP_i)$ and expected to produce a functional approximation for GDOP. In each d_i, $(h_1, h_2, \cdots h_{10})_i$ is computed from $\mathbf{H}'\mathbf{H}$ and $GDOP_i$ is determined by (8).

IV. EXPERIMENTS

In experiment part, we first adopt a heteroscedastic dataset to verify the effectiveness of the proposed pairing support vector regression algorithm. The Gaussian kernel

$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma\|\mathbf{x}_i - \mathbf{x}_j\|^2)$$

is used here. The RBF kernel is widely used due to its simplicity and the ability to deal with nonlinear problems.

The optimal value of model-parameters was tuned using a grid search mechanism. For simplicity, we set C_1=C_2. The training data sets are generated by

$$y_k = 0.2\sin(2\pi x_k) + 0.2x_k^2 + 0.3 + (0.1x_k^2 + 0.05)e_k, \quad (32)$$
$$x_k = 0.02(k-1), \quad k = 1,2,....,51,$$

where e_k represents a real number randomly generated in the interval $[-1;\ 1]$. This dataset has heteroscedastic noise structure, i.e., the noise is strongly corrected with the input value **x**. This example was also used in [11]. Figure 1 shows the regression function estimated by classical SVR (ε -SVR), TSVR, and the proposed pair-SVR. The support vectors are marked with circles in ε -SVR and pair-SVR. In ε -SVR and pair-SVR, the number of basis function for estimating the regression function is equal to number of support vector. However, in TSVR, the number of basis function is equal to the number of training samples. Hence, the sparsity of TSVR is worst. The ε -SVR is based on the assumption that the noise level is uniform throughout the domain. The assumption of a uniform noise model, however, is not always satisfied. In many regression tasks, the spread of noise might depend on location. Due to the assumption that the ε -insensitive zone has a tube (or slab) structure, the test error (risk) in ε -SVR is sensitive toward the changes in ε on this heteroscedastic data. As shown in Figure 1 (a) and (b), parameter ε determines the trade-off between sparsity and accuracy. As seen from Figure 1 (c), the nonparallel bound functions of TSVR captures the characteristics of the data set well and yield satisfactory regression function. However, the major disadvantage of TSVR is it loses the sparsity. The prediction time cost of TSVR is worst among the three approaches. Figure 1 (d) shows that the proposed pair-SVR derives the satisfying solution to estimating the distribution of noise and captures well the characteristics of the data set. More importantly, our approach preserves the benefit of TSVR, i.e., fast speed in learning, and meanwhile has the benefit of sparsity of classic ε -SVR, that is, small prediction time complexity.

We then employ the proposed pair-SVR algorithm to the approximation of GPS GDOP problem. In order to fairly compare the performance of other support vector regression algorithm, the same data set from [12] is adopted. In the dataset, more than 2500 data pairs are obtained. For forming the geometry matrix $\mathbf{M} = \mathbf{H}^t\mathbf{H}$, signals from 4 different satellites are randomly selected, from which GDOP is computed according to (8). The input h_1, h_2, . . . , h_{10} are extracted by using the Cholesky decomposition on $\mathbf{H}^t\mathbf{H} = \mathbf{R}^t\mathbf{R}$. Finally, inputs are collected accordingly from (31). To avoid the biased results, we employ the ten-fold cross-validation for the estimation of regression performance. All results are presented in average. The effectiveness of the GDOP regression model is evaluated by the mean square errors (MSEs):

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(f(\vec{x}_i) - GDOP_i\right)^2$$

The model parameters of classical SVR, TSVR and the proposed pair-SVR are tuned by grid-search strategy and ten-fold cross-validation procedure. Table I reports a comparison of the regression performance of classical SVR, TSVR, and the proposed pair-SVR in terms of MSEs, training time, and sparsity (the number of basis function (BF) used in the final regression model) on the GPS GDOP approximation problem. Noted, the number of basis function used in the regression model estimated by TSVR equals the number of training samples, while the number of basis functions in classical SVR and the proposed pair-SVR equals the number of support vectors (SVs). In general, the number of SVs is much fewer than training samples. Because the number of basis function is the main factor that effects the prediction time, the prediction speed of TSVR is the slowest due to the sparsity of TSVR is the worst. As shown in Table I, the training speed of classical SVR is the slowest because the learning of SVR needs to solve a large dense QPP. The computational complexity for solving the SVM-type QPP is O(N^3), where N is the number of training samples. On other hand, the TSVR and the proposed pair-SVR employ the strategy that solves two smaller-sized QPP rather than a single larger QPP. This strategy makes the learning speed of TSVR and the pair-SVR is approximately four times faster than classical SVR, as shown in Table I. Another disadvantage of TSVR is it considers only the training error instead of the generalization performance in the primal problem. In other words, TSVR performs the empirical risk minimization principle. However, it is well known that the superior generalization capability of SVM is achieved by performing of the structure risk minimization principle. Because both classical SVR and the proposed pair-SVR perform the structure risk minimization principle by introducing a regularization term that captures the characteristics of model complexity, they yield better MSEs than TSVR, as shown in Table I. Experimental results have demonstrated the effectiveness of the proposed method.

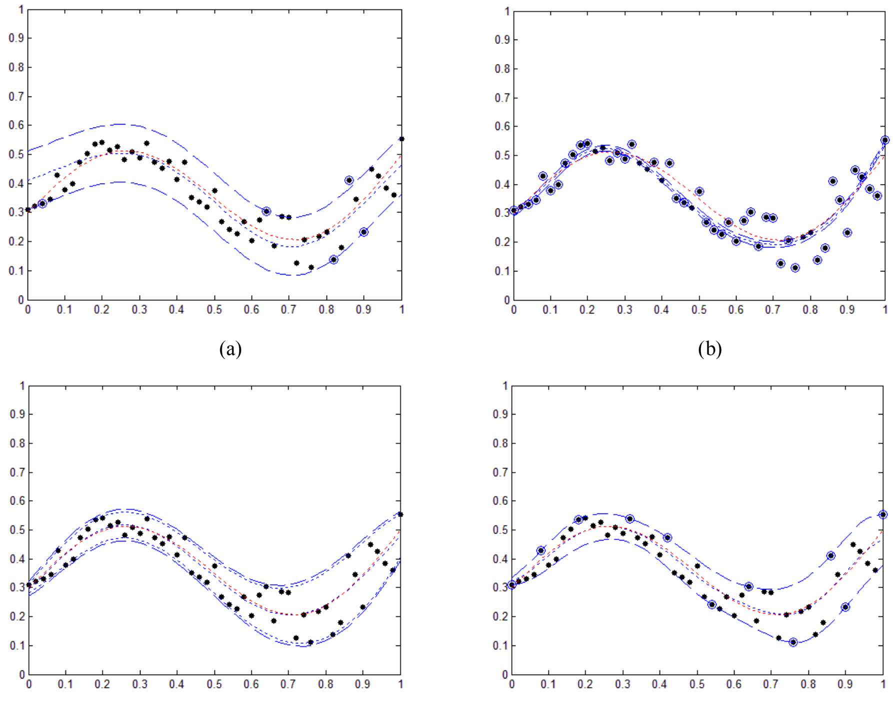
TABLE I. PERFORMANCE COMPARISONS.

Model	SVR	TSVR	Pair-SVR
MSEs	0.808	0.811	0.808
Training time	273.1	0.947	2.562
Num of BFs	1032.1	2250	956.4

In summary, the proposed approach not only has the superiority in faster training speed, but also owns better generalization ability and faster prediction speed.

V. CONCLUSION

One of the more common behavioral manifestations of dementia-related disorders is severe problems with out-of-home mobility. Various efforts have been attempted to attain a better understanding of mobility behavior, but most



(a) (b) (c) (d)

Figure 1. Regression models obtained by (a) ε-SVR (ε=0.1), (b) ε-SVR (ε=0.01), (c) TSVR, and (d) pair- SVR

studies are based on institutionalized patients and the assessment usually relies on reports of caregivers and institutional staff, using observational approaches, activity monitoring, or behavioral checklists. Previous studies have reported that GPS is an advanced research tool able to understand out-of-home behavior better than was possible with previous methods. In this paper, the new pair-SVR algorithm is proposed to evaluate nonlinear regression models for the approximation of GPS GDOP, which can improve the use of GPS and advanced tracking technologies for the analysis of mobility in cognitive diseases. Motivated by TSVR, the pair-SVR estimates indirectly the regression model through a pair of nonparallel insensitive upper bound and lower bound functions solved by two smaller sized SVM- type problems, which makes the pair-SVR not only yield the faster learning speed than the classical SVR, but also be suitable for many cases, especially when the noise is heteroscedastic, that is, the noise is strongly corrected with the input. Besides, we improved the sparsity of TSVR by introducing an insensitive zone constructed by a pair of nonparallel upper bound and lower bound function. Only points outside the zone are captured as SVs, and only those SVs construct the final regression function. The experimental results validate that the pair-SVR not only has small training cost, but also owns good generalization ability and sparsity.

REFERENCES

[1] N. Shoval, et. al. The use of advanced tracking technologies for the analysis of mobility in Alzheimer's disease and related cognitive diseases, BMC Geriatrics,2008,8:7.

[2] N. Shoval, et. al. Use of the Global Positioning System to Measure the Out-of-Home Mobility of Older Adults with Differing Cognitive Functioning, ARTICLE in AGEING AND SOCIETY 31(05): · JULY 2011, pp. 849 - 869

[3] V. Ashkenazi, "Coordinate systems: How to get your position very precise and completely wrong," J. Navig., vol. 39, no. 2, May 1986, pp. 269–278.

[4] D. J. Jwo and K. P. Chin, "Applying back-propagation neural networks to GDOP approximation," J. Navig., vol. 55, no. 1, Jan. 2002, pp. 97–108.

[5] D. Simon and H. El-Sherief, "Navigation satellite selection using neural networks," Neurocomputing,7(3), 1995, pp. 247–258.

[6] C.-H. Wu, W.-H. Su, and Y.-W. Ho, "A Study on GPS GDOP Approximation Using Support-Vector Machines," IEEE Transactions on Instrumentation & Measurement, 60(1), 2011, pp 137 – 145.

[7] V. Vapnik, The Nature of Statistical Learning Theory. Springer, New York, 2000.

[8] Jayadeva, R. Khemchandani, and S. Chandra, "Twin support vector machines for pattern classification," IEEE Tran. on Pattern Analysis and Machine Intellegence 29 (5) (2007) pp. 905–910.

[9] X. Peng, "TSVR: an efficient twin support vector machine for regression," Neural Networks 23 (3) (2010) pp. 365–372.

[10] A. H. Roger and R. J. Charles, Matrix Analysis. Cambridge, U.K.: Cambridge Univ. Press, 1982.

[11] P.-Y. Hao, "New Support Vector Algorithms with Parametric Insensitive/Margin Model," Neural Networks, vol. 23, no. 1, January 2010, pp. 60-73.

[12] C.-H. Wu and W-H. Su, "A comparative study on regression models of gps gdop using soft-computing techniques," in Proceedings of the 2009 IEEE International Conference on Fuzzy Systems (Fuzz-IEEE 2009), Korea, Aug. 2009, pp. 1513-1516.