International Journal of Innovative
Computing, Information and Control
Volume **11**, Number **6**, December **2015** ISSN 1349-4198
pp. 2103–2117

EFFECTIVE ν-SOLUTION PATH FOR ν-SUPPORT VECTOR REGRESSION

Binjie Gu* and Feng Pan

Key Laboratory of Advanced Process Control for Light Industry, Ministry of Education
Jiangnan University
No. 1800, Lihu Road, Wuxi 214122, P. R. China
*Corresponding author: gubinjie1980@126.com

Received April 2015; revised August 2015

Abstract. *In this paper, we deal with the ν-solution path problem for ν-support vector regression (ν-SVR). An effective ν-solution path algorithm for ν-SVR is proposed which can avoid the infeasible updating path during the adiabatic incremental adjustment for ν. Compared with the existing ν-solution path algorithm, our algorithm first solves the two complications in the dual formulation of ν-SVR. Then a novel strategy is utilized to address the infeasible updating path problem. Moreover, we prove the feasibility and finite convergence of our algorithm. Finally, simulation experiment results further demonstrate that our algorithm can effectively avoid the infeasible updating path, and will converge to the optimal solution of minimization problem within finite steps.*
Keywords: Machine learning, Model selection, ν-solution path, Feasibility analysis, Finite convergence analysis

1. **Introduction.** The support vector machine (SVM) proposed by Vapnik is a machine learning algorithm based on statistical learning theory, which can solve small sample learning problem effectively [1]. To date, SVM has been widely used in many areas, such as forest fires burned area prediction [2], and fast predictors for large-scale time series [3]. However, there are still some open questions needed to be addressed. The central one is model selection, i.e., how to tune the parameter of SVM to achieve optimal generalization capacity [4-6].

The idea of general model selection approach is to select some candidate parameter values and then apply cross validation (CV) to choose the optimal parameter value among the candidates [4]. Unfortunately, if the search space is very large, the general model selection approach must train SVM many times under different parameter settings. This greatly limits its application in online scenarios.

Over the last decades, various approaches have been developed to address the problem mentioned above. Based on the piece-wise linear fashion, a novel approach is proposed for C-SVM (hereinafter referred to as the SvmPath), which can fit the entire solution path for the regularization parameter C and only needs to train SVM once [7]. In the work of [8-10], the SvmPath is extended to the solution path for ε-support vector regression (ε-SVR), hereinafter referred to as the SvrPath. The work of [11,12] focuses on the asymptotically optimal selection of parameter ν for ν-SVM. In the work of [13], a novel ν-solution path for ν-support vector regression (ν-SVR) is proposed; however, directly applying it will not guarantee that there always exists a feasible updating path [14]. To address this issue, an effective ν-solution path for ν-SVR (called the ν-SvrPath) is proposed in this paper, which can be viewed as an extension of the SvrPath.

The main contributions of this paper are summarized as follows. (1) We present an equivalent formulation of ν-SVR such that the constraints are independent of regularization parameter and the size of the training set, and then transform the inequality constraint of the equivalent formulation into equality constraint. (2) The ν-SvrPath ensures that there always exists a feasible updating path by utilizing a novel strategy. (3) We prove the feasibility and finite convergence of the ν-SvrPath, which ensures that the ν-SvrPath is reliable and will converge to the optimal solution of minimization problem within finite steps.

The rest of this paper is organized as follows. Section 2 provides a brief review of ν-SVR, and then discusses how to address the two complications in the dual problem of ν-SVR. The ν-SvrPath is presented in Section 3. In Section 4, we prove the feasibility and finite convergence of the ν-SvrPath. The simulation experiments are carried out in Section 5. Finally, conclusions are made in Section 6.

Notations: $\mathbf{R}^n$ denotes the n-dimensional Euclidean space; $(*)$ stands for a variable without and with $*$; Δ stands for the amount of the change of each variable; '*def*' above '=' means that the left side of the equal sign is defined as the right side; the superscript T stands for transposition; $\varnothing$ stands for the empty set; $\det(\cdot)$ stands for the determinant of a square matrix; $\mathbf{P}^{-1}$ denotes the inverse of the matrix $\mathbf{P}$; $\mathbf{e}_S$ denotes the all ones column vector indexed by the set S; $\mathbf{0}$ denotes the all zeros column vector with proper dimensions; $\mathbf{I}_m$ denotes the identity matrix with m dimensions; $\mathbf{Q}_{St}$ denotes the subvector of the matrix $\mathbf{Q}$ with the rows and columns indexed by the sets S and t, respectively; $\mathbf{Q}_{SS}$ denotes the submatrix of $\mathbf{Q}$ with the rows and columns indexed by the set S; M_{tt} denotes the i_tth row and the i_tth column of the matrix $\mathbf{M}$, where i_t stands for the corresponding index in $\mathbf{M}$; $\mathbf{M}_{\backslash tt}$ denotes the submatrix of $\mathbf{M}$ with deleting the i_tth row and i_tth column, where i_t stands for the corresponding index in $\mathbf{M}$.

2. Equivalent Formulation of $\boldsymbol{\nu}$-SVR.

2.1. Brief review of $\boldsymbol{\nu}$-SVR.

The ν-SVR proposed by Schölkopf et al. is an interesting type of SVM [11]. Given a training sample set $T = \{(\mathbf{x}_1, y_1), \cdots, (\mathbf{x}_l, y_l)\}$ such that $\mathbf{x}_i \in \mathbf{R}^n$ is an input and $y_i \in \mathbf{R}$ is a target output, the ν-SVR considers the following primal problem [15]:

$$\begin{aligned} \min_{\mathbf{w}, b, \varepsilon, \xi_i^{(*)}} \quad & P = \frac{1}{2} \|\mathbf{w}\|^2 + C \left(\nu\varepsilon + \frac{1}{l} \sum_{i=1}^{l} (\xi_i + \xi_i^*) \right) \\ \text{s.t.} \quad & \left(\mathbf{w}^T \phi(\mathbf{x}_i) + b\right) - y_i \le \varepsilon + \xi_i, \ y_i - \left(\mathbf{w}^T \phi(\mathbf{x}_i) + b\right) \le \varepsilon + \xi_i^*, \ \xi_i^{(*)} \ge 0, \\ & i = 1, \cdots, l, \ \varepsilon \ge 0 \end{aligned} \tag{1}$$

Here the training samples $\mathbf{x}_i$ are mapped into a high dimensional reproducing kernel Hilbert space (RKHS) by the transformation function ϕ, $\xi_i^{(*)}$ are nonnegative slack variables, b is bias, and the ε-insensitive loss function means that if $\mathbf{w}^T\phi(\mathbf{x}_i) + b$ is in the range of $y_i \pm \varepsilon$, no loss is considered. ν is the proportion parameter with $0 \le \nu \le 1$, which makes one control the number of support vectors and errors. To be more precise, ν is an upper bound on the fraction of margin errors, and a lower bound of the fraction of support vectors. In addition, with probability 1, asymptotically, ν equals both fractions. Therefore, it is easier to tune parameter ν than ε-SVR.

The corresponding dual problem of (1) is [15]:

$$\begin{aligned}
\min_{\alpha_i^{(*)}} \; & D = \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}(\alpha_i - \alpha_i^*)H_{ij}\left(\alpha_j - \alpha_j^*\right) + \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)y_i \\
\text{s.t. } & \sum_{i=1}^{l}(\alpha_i - \alpha_i^*) = 0;\; 0 \le \alpha_i^{(*)} \le C/l,\; i = 1, \cdots, l;\; \sum_{i=1}^{l}(\alpha_i + \alpha_i^*) \le C\nu
\end{aligned} \tag{2}$$

where $H_{ij} = K(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j)\rangle$, K is the kernel function, and $\langle \cdot, \cdot \rangle$ denotes the inner product in RKHS.

However, in comparison to the dual problem of ε-SVR [15], ν-SVR has two complications: the first one is that the constraints $0 \le \alpha_i^{(*)} \le C/l$, $i = 1, \cdots, l$ are related to regularization parameter and the size of the training set, and the second one is that (2) has an extra inequality constraint.

2.2. **Equivalent formulation of ν-SVR.** To solve the first complication, we multiply the objective function P of (1) by the size of the training sample set, and consider the following primal problem:

$$\begin{aligned}
\min_{\mathbf{w}, b, \varepsilon, \xi_i^{(*)}} \; & P = \frac{l}{2}\|\mathbf{w}\|^2 + C\left(\nu l\varepsilon + \sum_{i=1}^{l}(\xi_i + \xi_i^*)\right) \\
\text{s.t. } & \left(\mathbf{w}^T\phi(\mathbf{x}_i) + b\right) - y_i \le \varepsilon + \xi_i,\; y_i - \left(\mathbf{w}^T\phi(\mathbf{x}_i) + b\right) \le \varepsilon + \xi_i^*,\; \xi_i^{(*)} \ge 0, \\
& i = 1, \cdots, l,\; \varepsilon \ge 0
\end{aligned} \tag{3}$$

It is easy to verify that (3) is equivalent to (1). The corresponding dual problem of (3) is:

$$\begin{aligned}
\min_{\alpha_i^{(*)}} \; & D = \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}(\alpha_i - \alpha_i^*)Q_{ij}\left(\alpha_j - \alpha_j^*\right) - \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)y_i' \\
\text{s.t. } & \sum_{i=1}^{l}(\alpha_i - \alpha_i^*) = 0;\; 0 \le \alpha_i^{(*)} \le 1,\; i = 1, \cdots, l;\; \sum_{i=1}^{l}(\alpha_i + \alpha_i^*) \le \nu l
\end{aligned} \tag{4}$$

where $Q_{ij} = H_{ij}/l = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j)\rangle / l$ and $y_i' = -y_i/C$.

Note that the original training sample set $T = \{(\mathbf{x}_1, y_1), \cdots, (\mathbf{x}_l, y_l)\}$ will be changed into the new training sample set $F = \{(\mathbf{x}_1, y_1'), \cdots, (\mathbf{x}_l, y_l')\}$.

Furthermore, we can solve the second complication based on Theorem 2.1.

Theorem 2.1. *For (4), if $0 \le \nu \le 1$, there are always optimal solutions which happen at the equality $\sum_{i=1}^{l}(\alpha_i + \alpha_i^*) = \nu l$.*

The detailed proof of Theorem 2.1 can be found in [15], and it is omitted here.

Based on Theorem 2.1, $\sum_{i=1}^{l}(\alpha_i + \alpha_i^*) \le \nu l$ can be replaced by $\sum_{i=1}^{l}(\alpha_i + \alpha_i^*) = \nu l$. Therefore, we can consider the following minimization problem instead of (4):

$$\begin{aligned}
\min_{\alpha_i^{(*)}} \; & D = \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}(\alpha_i - \alpha_i^*)Q_{ij}\left(\alpha_j - \alpha_j^*\right) - \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)y_i' \\
\text{s.t. } & \sum_{i=1}^{l}(\alpha_i - \alpha_i^*) = 0;\; 0 \le \alpha_i^{(*)} \le 1,\; i = 1, \cdots, l;\; \sum_{i=1}^{l}(\alpha_i + \alpha_i^*) = \nu l
\end{aligned} \tag{5}$$

Given the solution of (5), the regression function of (1) can be written as:

$$f(\mathbf{x}) = \sum_{j=1}^{l}\left(\alpha_j - \alpha_j^*\right)\left[K(\mathbf{x}_j, \mathbf{x})/l\right] + b \tag{6}$$

According to the convex optimization theory, the solution of (5) can be obtained by minimizing the following convex quadratic objective function under constraints:

$$\begin{aligned}\min_{0\le\alpha_i^{(*)}\le 1} W = & \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}(\alpha_i-\alpha_i^*)Q_{ij}\left(\alpha_j-\alpha_j^*\right)-\sum_{i=1}^{l}(\alpha_i-\alpha_i^*)y_i' \\ & + b\sum_{i=1}^{l}(\alpha_i-\alpha_i^*)+\rho\left(\sum_{i=1}^{l}(\alpha_i+\alpha_i^*)-\nu l\right)\end{aligned} \tag{7}$$

where b and ρ are Lagrange multipliers.

For simplicity, let $\theta_i = \alpha_i - \alpha_i^*$ denote coefficient difference. Then optimizing (7) leads to the following Karush_Kuhn_Tucker (KKT) conditions [16,17]:

$$g_i = \frac{\partial W}{\partial \alpha_i} = \sum_{j=1}^{l} Q_{ij}\theta_j - y_i' + b + \rho \tag{8}$$

$$g_i^* = \frac{\partial W}{\partial \alpha_i^*} = -\sum_{j=1}^{l} Q_{ij}\theta_j + y_i' - b + \rho = -g_i + 2\rho \tag{9}$$

$$\frac{\partial W}{\partial b} = \sum_{i=1}^{l}(\alpha_i - \alpha_i^*) = 0 \tag{10}$$

$$\frac{\partial W}{\partial \rho} = \sum_{i=1}^{l}(\alpha_i + \alpha_i^*) - \nu l = 0 \tag{11}$$

Combining (8) and (9), we have:

$$\begin{cases} g_i \ge 2\rho,\ g_i^* \le 0 & \theta_i = -1 & \alpha_i = 0,\ \alpha_i^* = 1 & \forall i \in E_R \\ g_i = 2\rho,\ g_i^* = 0 & -1 < \theta_i < 0 & \alpha_i = 0,\ 0 < \alpha_i^* < 1 & \forall i \in S_R \\ 0 \le g_i,\ g_i^* \le 2\rho & \theta_i = 0 & \alpha_i = \alpha_i^* = 0 & \forall i \in R \\ g_i = 0,\ g_i^* = 2\rho & 0 < \theta_i < 1 & 0 < \alpha_i < 1,\ \alpha_i^* = 0 & \forall i \in S_L \\ g_i \le 0,\ g_i^* \ge 2\rho & \theta_i = 1 & \alpha_i = 1,\ \alpha_i^* = 0 & \forall i \in E_L \end{cases} \tag{12}$$

According to the value of θ_i, the new training sample set F can be partitioned into three sets as shown in Figure 1.

(a) the set $S = S_L \cup S_R = \{i|0 < |\theta_i| < 1\}$, which includes margin support vectors strictly on the margins;

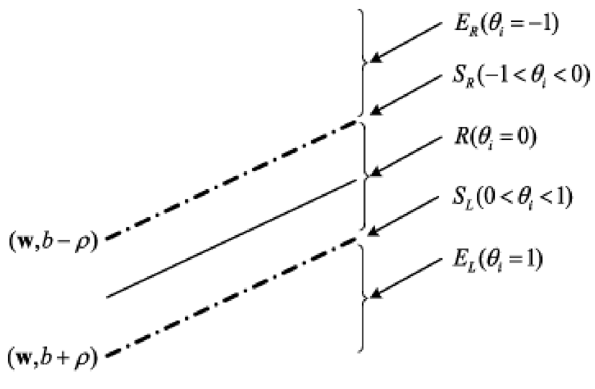

FIGURE 1. Partition of the new training samples F into three sets by the KKT conditions

(b) the set $E = E_L \cup E_R = \{i||\theta_i| = 1\}$, which includes error support vectors exceeding the margins;

(c) the set $R = \{i||\theta_i| = 0\}$, which includes the remaining vectors ignored by the margins.

For convenience, suppose the number of samples in the sets S_L, S_R and S is p, q and r, respectively. It is obvious that $r = p + q$.

3. The ν-SvrPath.

In this section, we will address the infeasible updating path problem during the adiabatic incremental adjustments for ν. The ν-SvrPath mainly includes two parts: the first part is to establish the initial solution for the minimization problem of (5), and the second part is to explore the solution path for all values of $0 \le \nu \le 1$.

3.1. Initialization.

Firstly, the ν-SvrPath needs to establish an initial solution for the minimization problem (5).

Lemma 3.1. *If $\nu = 0$, the optimal solution of (5) is $\alpha_i^{(*)} = 0$.*

Proof: If $\nu = 0$, then according to (11), we have $\sum\limits_{i=1}^{l} (\alpha_i + \alpha_i^*) = 0$. Note that $0 \le \alpha_i^{(*)} \le 1$, so we have $\alpha_i^{(*)} = 0$. This completes the proof.

In fact, the ν-SvrPath can be started at any intermediate solution, i.e., the solution of the minimization problem (5) for any feasible value of $0 \le \nu \le 1$.

Lemma 3.1 means $\theta_i = 0$, so according to (8) and (9), we have $g_i = -y_i' + b + \rho$ and $g_i^* = y_i' - b + \rho$. Then according to (12), we have:

$$\begin{cases} 0 \le -y_i' + b + \rho \le 2\rho \\ 0 \le y_i' - b + \rho \le 2\rho \end{cases} \tag{13}$$

Picking $i_+ = \arg\max\limits_{i \in R} y_i'$ and $i_- = \arg\min\limits_{i \in R} y_i'$, for simplicity, we assume that i_+ and i_- are unique. Obviously, the solutions of (13) can be formulated as follows:

$$\rho \ge (y_{i_+}' - y_{i_-}')/2 \tag{14}$$

$$b \in \left[y_{i_+}' - \rho, y_{i_-}' + \rho\right] \tag{15}$$

Furthermore, if $\rho = (y_{i_+}' - y_{i_-}')/2$, then we can easily obtain that $b = (y_{i_+}' + y_{i_-}')/2$, which constitutes the initial state of ρ and b. This also means that two or more samples start in the set S [8].

3.2. ν-Solution path.

After the initialization was completed, the ν-solution path will explore the solution path by gradually increasing ν from 0 to 1 under the condition of rigorously keeping all samples satisfying the KKT conditions, and will terminate when $\sum_{i=1}^{l} (\alpha_i + \alpha_i^*) = l$ is met. Furthermore, if the ν-SvrPath starts at an intermediate solution, we can also obtain the solution path similarly [8].

A. Adiabatic Incremental Adjustments for ν. During the adiabatic incremental adjustments for ν, in order to keep all the samples satisfying the KKT conditions, the θ_i in the set S, the Lagrange multipliers b and ρ should also be adjusted accordingly. Based on (8)-(11), we can obtain the following linear system:

$$\Delta g_i = \sum_{j \in S_L} \Delta\alpha_j Q_{ij} - \sum_{j \in S_R} \Delta\alpha_j^* Q_{ij} + \Delta b + \Delta\rho = 0, \ \forall i \in S_L \tag{16}$$

$$\Delta g_i^* = -\sum_{j \in S_L} \Delta\alpha_j Q_{ij} + \sum_{j \in S_R} \Delta\alpha_j^* Q_{ij} - \Delta b + \Delta\rho = 2\Delta\rho - \Delta g_i = 0, \quad \forall i \in S_R \tag{17}$$

$$\sum_{j\in S_L}\Delta\alpha_j-\sum_{j\in S_R}\Delta\alpha_j^*=0 \tag{18}$$

$$\sum_{j\in S_L}\Delta\alpha_j+\sum_{j\in S_R}\Delta\alpha_j^*=\Delta\nu\cdot l \tag{19}$$

Note that $\Delta\nu$ in (19) is unknown and $\Delta\nu>0$. Obviously, if $S_L=\varnothing$ or $S_R=\varnothing$, (18) and (19) cannot hold simultaneously. We call this the contradictions, as shown in Table 1.

Table 1. Two cases of contradiction during the adiabatic incremental adjustment for ν

S_L **and** S_R		**Contradiction (Yes or No)**
$S_L\neq\varnothing$	$S_R\neq\varnothing$	No
$S_L=\varnothing$	$S_R\neq\varnothing$	Yes
$S_L\neq\varnothing$	$S_R=\varnothing$	Yes

To avoid the contradictions, we change (19) into the following form:

$$\sum_{j\in S_L}\Delta\alpha_j+\sum_{j\in S_R}\Delta\alpha_j^*+\vartheta\Delta\rho+\Delta\eta=0 \tag{20}$$

where, $\Delta\eta$ is the introduced new variable for adjusting $\sum_{j\in S_L}\alpha_j+\sum_{j\in S_R}\alpha_j^*$; ϑ is any negative number; $\vartheta\Delta\rho$ is an extra term. The purpose of using $\vartheta\Delta\rho+\Delta\eta$ is to prevent the occurrence of contradictions as described in Table 1; meanwhile, it can preserve the KKT conditions of the path.

Define $\mathbf{e}_S=\begin{bmatrix}\mathbf{e}_{S_L}\\ \mathbf{e}_{S_R}\end{bmatrix}$, $\mathbf{m}_S=\begin{bmatrix}\mathbf{e}_{S_L}\\ -\mathbf{e}_{S_R}\end{bmatrix}$, $\mathbf{Q}_{SS}=\begin{bmatrix}\mathbf{Q}_{S_LS_L} & -\mathbf{Q}_{S_LS_R}\\ -\mathbf{Q}_{S_RS_L} & \mathbf{Q}_{S_RS_R}\end{bmatrix}$ and $\Delta\boldsymbol{\alpha}_S=\begin{bmatrix}\Delta\boldsymbol{\alpha}_{S_L}^T & \Delta\boldsymbol{\alpha}_{S_R}^{*T}\end{bmatrix}^T$, where $\Delta\boldsymbol{\alpha}_{S_L}=[\Delta\alpha_1,\cdots,\Delta\alpha_p]^T$ and $\Delta\boldsymbol{\alpha}_{S_R}^*=[\Delta\alpha_1^*,\cdots,\Delta\alpha_q^*]^T$. Then the linear systems (16)-(18) and (20) can be further rewritten as:

$$\underbrace{\begin{bmatrix}0 & 0 & \mathbf{m}_S^T\\ 0 & \vartheta & \mathbf{e}_S^T\\ \mathbf{m}_s & \mathbf{e}_s & \mathbf{Q}_{SS}\end{bmatrix}}_{\mathbf{P}}\cdot\begin{bmatrix}\Delta b\\ \Delta\rho\\ \Delta\boldsymbol{\alpha}_S\end{bmatrix}=-\begin{bmatrix}0\\ 1\\ \mathbf{0}\end{bmatrix}\Delta\eta \tag{21}$$

Letting $\mathbf{M}=\mathbf{P}^{-1}$, then we have:

$$\begin{bmatrix}\Delta b\\ \Delta\rho\\ \Delta\boldsymbol{\alpha}_S\end{bmatrix}=-\mathbf{M}\cdot\begin{bmatrix}0\\ 1\\ \mathbf{0}\end{bmatrix}\Delta\eta\overset{def}{=}\underbrace{\begin{bmatrix}\beta_b\\ \beta_\rho\\ \boldsymbol{\beta}_i^{(*)}\end{bmatrix}}_{\boldsymbol{\beta}}\Delta\eta,\quad\forall i\in S \tag{22}$$

where β_b stands for the dimension corresponding to b in the column vector $\boldsymbol{\beta}$, and β_ρ and $\boldsymbol{\beta}_i^{(*)}$ have the same meaning. From (20) and (22), we have $\sum_{j\in S_L}\Delta\alpha_j+\sum_{j\in S_R}\Delta\alpha_j^*=-(\vartheta\beta_\rho+1)\Delta\eta$, which also indicates that the control of the adjustment of $\sum_{j\in S_L}\alpha_j+\sum_{j\in S_R}\alpha_j^*$ can be realized by $\Delta\eta$.

Finally, substituting (22) into (16) and (17), we have:

$$\Delta g_i=\left(\sum_{j\in S_L}\beta_jQ_{ij}-\sum_{j\in S_R}\beta_j^*Q_{ij}+\beta_b+\beta_\rho\right)\Delta\eta\overset{def}{=}\tau_i\Delta\eta \tag{23}$$

$$\Delta g_i^*=\left(-\sum_{j\in S_L}\beta_jQ_{ij}+\sum_{j\in S_R}\beta_j^*Q_{ij}-\beta_b+\beta_\rho\right)\Delta\eta\overset{def}{=}\tau_i^*\Delta\eta=(2\beta_\rho-\tau_i)\Delta\eta \tag{24}$$

Obviously, we have $\tau_i = 0$, $i \in S_L$ and $\tau_i^* = 0$, $i \in S_R$.

B. Compute the Minimal Adjustment Quantity $\boldsymbol{\Delta\eta^{\min}}$. The samples will migrate between the sets S, E, and R during the adiabatic incremental adjustment for ν, and this will change the composition of the sets S, E, and R. To address this problem, the strategy is to compute the minimal adjustment quantity $\Delta\eta^{\min}$ such that only a certain sample migrates among the sets S, E, and R. Four cases should be considered to account for such membership changes.

Case 1: A certain sample is added to the set S from the set E or R. Firstly, compute the sets $\mathbf{I}^{S_L} = \{\beta_i \neq 0, i \in S_L\}$ and $\mathbf{I}^{S_R} = \{\beta_i^* \neq 0, i \in S_R\}$, where the samples with $\beta_i^{(*)} = 0$ are ignored due to their insensitivity to $\Delta\eta$.

Then the minimal possible adjustment is

$$\Delta\eta^{Case1} = \min_{i \in \mathbf{I}^{S_L} \cup \mathbf{I}^{S_R}} \left(\alpha_i^{(*)} - 1 \Big/ \beta_i^{(*)}, \alpha_i^{(*)} \Big/ \beta_i^{(*)} \right).$$

Case 2: A certain sample is removed from the set S to the set E. Firstly, compute the sets $\mathbf{I}_+^{E_L} = \{\tau_i > 0, i \in E_L\}$ and $\mathbf{I}_+^{E_R} = \{\tau_i^* > 0, i \in E_R\}$, where the samples with $\tau_i^{(*)} = 0$ are similarly ignored. Then the minimal possible adjustment is $\Delta\eta^{Case2} = \min_{i \in I_+^{E_L} \cup I_+^{E_R}} g_i^{(*)} \Big/ \tau_i^{(*)}$.

Case 3: A certain sample is removed from the set S to the set R. Similar to Case 2, compute the set $\mathbf{I}_-^R = \left\{\tau_i^{(*)} < 0, i \in R\right\}$. Then the minimal possible adjustment is $\Delta\eta^{Case3} = \min_{i \in \mathbf{I}_-^R} g_i^{(*)} \Big/ \tau_i^{(*)}$.

Case 4: $\sum_{i=1}^{l} (\alpha_i + \alpha_i^*) = l$, i.e., the termination condition is met. Then the minimal adjustment is $\Delta\eta^{Case4} = \left(\sum_{i=1}^{l} (\alpha_i + \alpha_i^*) - l\right) \Big/ (\vartheta\beta_\rho + 1)$.

Finally, the largest of the four values

$$\Delta\eta^{\min} = \max\left\{\Delta\eta^{Case1}, \Delta\eta^{Case2}, \Delta\eta^{Case3}, \Delta\eta^{Case4}\right\} \tag{25}$$

will constitute the minimal adjustment quantity of $\Delta\eta$.

C. Update ν, b, ρ, $\alpha_i^{(*)}$, $g_i^{(*)}$, S, E, and R. After the minimal adjustment quantity of $\Delta\eta^{\min}$ is determined, ν can be updated from (19), (20), and (22) as $\nu \leftarrow \nu - (\vartheta\beta_\rho + 1)\Delta\eta^{\min}/l$. Similarly, b, ρ, $\alpha_i^{(*)}$ and $g_i^{(*)}$ can be updated from (22), (23), and (24) as $b \leftarrow b + \beta_b \Delta\eta^{\min}$, $\rho \leftarrow \rho + \beta_\rho \Delta\eta^{\min}$, $\alpha_i^{(*)} \leftarrow \alpha_i^{(*)} + \beta_i^{(*)} \Delta\eta^{\min}$, and $g_i^{(*)} \leftarrow g_i^{(*)} + \tau_i^{(*)} \Delta\eta^{\min}$, respectively.

After the minimal adjustment quantity of $\Delta\eta^{\min}$ is calculated, if $\Delta\eta^{\min} = \Delta\eta^{Case4}$, the ν-SvrPath has to terminate. Otherwise, the index of the sample yielding the maximum in (25) can be obtained, which is denoted as t. Then the sets S, E, and R can be updated accordingly as follows: if $\Delta\eta^{\min} = \Delta\eta^{Case1}$, then t should be added to the set S from the set E or R; if $\Delta\eta^{\min} = \Delta\eta^{Case2}$, then t should be removed from the set S to the set E; if $\Delta\eta^{\min} = \Delta\eta^{Case3}$, then t should be removed from the set S to the set R.

D. Update the Inverse Matrix M. Once a sample is either added to or removed from the set S, there will also exist changes in matrix $\mathbf{P}$ and its inverse matrix $\mathbf{M}$ accordingly. Fortunately, based on Lemma 3.2, we can update the inverse matrix $\mathbf{M}$ effectively.

Lemma 3.2. *Suppose a $(s+1) \times (s+1)$ matrix $\mathbf{B}$ can be partitioned into a block form:*

$$\mathbf{B} = \begin{bmatrix} \mathbf{A} & \boldsymbol{\eta}_t \\ \boldsymbol{\eta}_t^T & Q_{tt} \end{bmatrix}$$

where $\mathbf{A}$ *is an* $s \times s$ *matrix and* $\mathbf{A}$ *can be inverted,* $\boldsymbol{\eta}_t = [Q_{1t}, \cdots, Q_{st}]^T$*, and* $Q_{tt} \neq 0$ *is a constant.*

Then, the inverse matrix of $\mathbf{B}$ *can be expanded as follows:*

$$\mathbf{B}^{-1} = \begin{bmatrix} \mathbf{A}^{-1} & \mathbf{0} \\ \mathbf{0}^{\mathrm{T}} & 0 \end{bmatrix} + \frac{1}{k} \begin{bmatrix} \boldsymbol{\beta}_t \\ 1 \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{\beta}_t \\ 1 \end{bmatrix}^{\mathrm{T}}$$

where $\boldsymbol{\beta}_t = -\mathbf{A}^{-1}\boldsymbol{\eta}_t$ *and* $k = \boldsymbol{\eta}_t^T \boldsymbol{\beta}_t + Q_{tt}$*.*

Furthermore, if $\mathbf{B}$ *can be inverted and* $(\mathbf{B}^{-1})_{tt} \neq 0$*,* $t = s+1$*, then the inverse of matrix of* $\mathbf{A}$ *can be contracted as follows:*

$$\mathbf{A}^{-1} = \left(\mathbf{B}^{-1}\right)_{\backslash tt} - \frac{\left((\mathbf{B}^{-1})_{*t} \cdot (\mathbf{B}^{-1})_{t*}\right)_{\backslash tt}}{(\mathbf{B}^{-1})_{tt}}$$

where $* \neq t$*.*

It can be easily verified that $\mathbf{B}\mathbf{B}^{-1} = \mathbf{I}_{s+1}$ and $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}_s$. The detailed proof of Lemma 3.2 can be found in [18], and it is omitted here.

Based on Lemma 3.2, if a sample $(\mathbf{x}_t, y'_t)$ is added to the set S, then the inverse matrix $\mathbf{M}$ can be expanded as follows:

$$\mathbf{M} \leftarrow \begin{bmatrix} \mathbf{M} & \mathbf{0} \\ \mathbf{0}^{\mathrm{T}} & 0 \end{bmatrix} + \frac{1}{\bar{\tau}_t} \begin{bmatrix} \boldsymbol{\gamma}_t \\ 1 \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{\gamma}_t \\ 1 \end{bmatrix}^T \tag{26}$$

where $\boldsymbol{\gamma}_t = \begin{bmatrix} \bar{\beta}_b \\ \bar{\beta}_\rho \\ \bar{\boldsymbol{\beta}}_t \\ \bar{\boldsymbol{\beta}}_t^* \end{bmatrix} = -\mathbf{M} \begin{bmatrix} z_t \\ 1 \\ z_t \mathbf{Q}_{S_L t} \\ -z_t \mathbf{Q}_{S_R t} \end{bmatrix}$, $\bar{\tau}_t = z_t \left(\sum_{j \in S_L} \bar{\beta}_j Q_{ij} - \sum_{j \in S_R} \bar{\beta}_j^* Q_{ij} + \bar{\beta}_b \right) + \bar{\beta}_\rho + Q_{tt}$, $z_t = +1$ or $z_t = -1$, which corresponds to the sample $(\mathbf{x}_t, y'_t)$ is added to the set S_L or S_R, respectively.

Similarly, if a sample $(\mathbf{x}_t, y'_t)$ is removed from the set S, then the inverse matrix $\mathbf{M}$ can be contracted as follows:

$$\mathbf{M} \leftarrow \mathbf{M}_{\backslash tt} - \frac{(\mathbf{M}_{*t} \cdot \mathbf{M}_{t*})_{\backslash tt}}{M_{tt}} \tag{27}$$

E. The $\boldsymbol{\nu}$-SvrPath Procedure. The ν-SvrPath procedure is presented in Algorithm 1.

4. Feasibility and Finite Convergence Analysis.

4.1. Feasibility analysis.

The feasibility analysis ensures that each adiabatic incremental adjustment for ν is reliable.

Assumption 4.1. *The matrix* $\mathbf{Q}_{SS}$ *is positive definite.*

It is easy to prove that if and only if $\{y'_1\phi(\mathbf{x}_1), \cdots, y'_r\phi(\mathbf{x}_r)\}$ in RKHS is linearly independent, the matrix $\mathbf{Q}_{SS}$ is positive definite. For example, if radial basis function is used as kernel function and where $\mathbf{x}_i \neq \pm\mathbf{x}_j$ for $i \neq j$, then the matrix $\mathbf{Q}_{SS}$ is positive definite. In practice, the size of matrix $\mathbf{Q}_{SS}$ is a very small number in comparison to the dimension of RKHS. Therefore, Assumption 4.1 always holds.

Lemma 4.1. *Suppose that* $\mathbf{C}$*,* $\mathbf{D}$*,* $\mathbf{E}$*,* $\mathbf{F}$ *are* $n \times n$*,* $n \times m$*,* $m \times n$*,* $m \times m$ *matrices, respectively, and* $\mathbf{F}$ *has the inverse matrix. Then*

$$\det\left(\begin{bmatrix} \mathbf{C} & \mathbf{D} \\ \mathbf{E} & \mathbf{F} \end{bmatrix}\right) = \det(\mathbf{F}) \cdot \det(\mathbf{C} - \mathbf{D}\mathbf{F}^{-1}\mathbf{E}).$$

Algorithm 1 The ν-SvrPath procedure: high-level summary.

Inputs: the new training set F and $\Delta\nu$
Outputs: $\alpha_i^{(*)}$, $g_i^{(*)}$, and ν

Initialize ν, b, ρ, $\alpha_i^{(*)}$, $g_i^{(*)}$, S, E, R // see Section 3.1
repeat
  set exception$\leftarrow$false.
  while $\nu > 0$ and exception$\leftarrow$false
    compute $\boldsymbol{\beta}$ and $\tau_i^{(*)}$ // see Section 3.2.A
    compute $\Delta\eta^{\min}$ // see Section 3.2.B
    update ν, b, ρ, $\alpha_i^{(*)}$, $g_i^{(*)}$, S, E and R // see Section 3.2.C
    update the inverse matrix $\mathbf{M}$ // see Section 3.2.D
    if $S_L = \varnothing \& E_L = \varnothing$ or $S_R = \varnothing \& E_R = \varnothing$
      set exception$\leftarrow$true
      update $\nu \leftarrow \nu + 0.001$
    endif
  endwhile
until $\nu < 1$

Proof: It is easy to verify that

$$\begin{bmatrix} \mathbf{C} & \mathbf{D} \\ \mathbf{E} & \mathbf{F} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_n & \mathbf{D} \\ \mathbf{0} & \mathbf{F} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{C} - \mathbf{D}\mathbf{F}^{-1}\mathbf{E} & \mathbf{0} \\ \mathbf{F}^{-1}\mathbf{E} & \mathbf{I}_m \end{bmatrix}.$$

Therefore, we have $\det\left(\begin{bmatrix} \mathbf{C} & \mathbf{D} \\ \mathbf{E} & \mathbf{F} \end{bmatrix}\right) = \det\left(\begin{bmatrix} \mathbf{I}_n & \mathbf{D} \\ \mathbf{0} & \mathbf{F} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{C} - \mathbf{D}\mathbf{F}^{-1}\mathbf{E} & \mathbf{0} \\ \mathbf{F}^{-1}\mathbf{E} & \mathbf{I}_m \end{bmatrix}\right) = \det(\mathbf{F}) \cdot \det(\mathbf{C} - \mathbf{D}\mathbf{F}^{-1}\mathbf{E})$. This completes the proof.

Theorem 4.1. *During the adiabatic incremental adjustments for ν, if $\vartheta < 0$, then the determinant of $\mathbf{P}$ is always greater than zero.*

Proof: Define the matrix $\mathbf{N} = \begin{bmatrix} 0 & 0 & \mathbf{m}_S^T \\ 0 & 0 & \mathbf{e}_S^T \\ \mathbf{m}_s & \mathbf{e}_s & \mathbf{Q}_{SS} \end{bmatrix}$. From Assumption 4.1, $\mathbf{Q}_{SS}$ is positive definite. This means $\mathbf{Q}_{SS}$ can be inverted, and $\det(\mathbf{Q}_{SS}) > 0$. From Lemma 4.1, we have:

$$\begin{aligned} \det(\mathbf{N}) &= \det(\mathbf{Q}_{SS}) \cdot \det\begin{pmatrix} -\mathbf{m}_S^T\mathbf{Q}_{SS}^{-1}\mathbf{m}_s & -\mathbf{m}_S^T\mathbf{Q}_{SS}^{-1}\mathbf{e}_s \\ -\mathbf{e}_S^T\mathbf{Q}_{SS}^{-1}\mathbf{m}_s & -\mathbf{e}_S^T\mathbf{Q}_{SS}^{-1}\mathbf{e}_s \end{pmatrix} \\ &= \det(\mathbf{Q}_{SS}) \cdot \left(\mathbf{m}_S^T\mathbf{Q}_{SS}^{-1}\mathbf{m}_s \cdot \mathbf{e}_S^T\mathbf{Q}_{SS}^{-1}\mathbf{e}_s - \mathbf{m}_S^T\mathbf{Q}_{SS}^{-1}\mathbf{e}_s \cdot \mathbf{e}_S^T\mathbf{Q}_{SS}^{-1}\mathbf{m}_s\right). \end{aligned}$$

According to the Cauchy-Schwarz inequality, we have:

$$\mathbf{m}_S^T\mathbf{Q}_{SS}^{-1}\mathbf{m}_s \cdot \mathbf{e}_S^T\mathbf{Q}_{SS}^{-1}\mathbf{e}_s - \mathbf{m}_S^T\mathbf{Q}_{SS}^{-1}\mathbf{e}_s \cdot \mathbf{e}_S^T\mathbf{Q}_{SS}^{-1}\mathbf{m}_s \geq 0.$$

Then we have $\det(\mathbf{N}) \geq 0$.

Similarly, we can also prove that $\det(\mathbf{P}_{\backslash\rho\rho}) = \det(\mathbf{Q}_{SS}) \cdot \det(-\mathbf{m}_S^T\mathbf{Q}_{SS}\mathbf{m}_S) < 0$.

By comparison of determinant expansion of the second row for $\mathbf{P}$ and $\mathbf{N}$, and note that $\vartheta < 0$, then we have $\det(\mathbf{P}) = \det(\mathbf{N}) + \vartheta \det(\mathbf{P}_{\backslash\rho\rho}) > 0$. This completes the proof.

Corollary 4.1. *During the adiabatic incremental adjustments for ν, there always exists the inverse matrix $\mathbf{M}$ for $\mathbf{P}$.*

Proof: According to the necessary and sufficient condition of an inverse matrix, the corollary can be easily derived from Theorem 4.1.

Lemma 4.2. *During the adiabatic incremental adjustments for ν, if $S = \{(\mathbf{x}_t, y_t)\}$, then $\beta_t^{(*)} = 0$.*

Proof: From (22), we have $\beta_t^{(*)} = -\mathbf{M}_{t\rho}$. According to the definition of an inverse matrix, we have $\mathbf{M}_{t\rho} = (-1)^{i_t+i_\rho}\det(\mathbf{P}_{\backslash\rho t})/\det(\mathbf{P})$. If $S_L = \{(\mathbf{x}_t, y_t)\}$, it is easy to verify that $\det(\mathbf{P}_{\backslash\rho t}) = 0$, and then based on Theorem 4.1, we have $\beta_t = 0$. Similarly, if $S_R = \{(\mathbf{x}_t, y_t)\}$, we can prove that $\beta_t^* = 0$. This completes the proof.

Theorem 4.2. *During the adiabatic incremental adjustments for ν, the set S will always be nonempty.*

Proof: From Lemma 4.2, if $S = \{(\mathbf{x}_t, y_t)\}$, then $\beta_t^{(*)} = 0$, so the sample t will not be removed from the set S. This completes the proof.

4.2. **Finite convergence analysis.** The finite convergence analysis ensures that the ν-SvrPath will converge to the optimal solution of minimization problem within finite steps.

Lemma 4.3. *During the adiabatic incremental adjustments for ν, if $\vartheta < 0$, then $\vartheta\beta_\rho + 1 \geq 0$ with equality if and only if $S_L = \varnothing$ or $S_R = \varnothing$.*

Proof: Based on (22) and the proof in Theorem 4.1, we have:

$$\vartheta\beta_\rho + 1 = -\vartheta \cdot \frac{(-1)^{i_\rho+i_\rho}\det(\mathbf{P}_{\backslash\rho\rho})}{\det(\mathbf{P})} + 1 = \frac{\det(\mathbf{P}) - \vartheta\det(\mathbf{P}_{\backslash\rho\rho})}{\det(\mathbf{P})} = \frac{\det(\mathbf{N})}{\det(\mathbf{P})} \geq 0.$$

with equality if and only if $\det(\mathbf{N}) = 0$.

This requires that $\mathbf{m}_S^T\mathbf{Q}_{SS}^{-1}\mathbf{m}_s \cdot \mathbf{e}_S^T\mathbf{Q}_{SS}^{-1}\mathbf{e}_s - \mathbf{m}_S^T\mathbf{Q}_{SS}^{-1}\mathbf{e}_s \cdot \mathbf{e}_S^T\mathbf{Q}_{SS}^{-1}\mathbf{m}_s = 0$, i.e., $\mathbf{m}_s = \mathbf{e}_S$ or $\mathbf{m}_s = -\mathbf{e}_S$, which means that $S_L = \varnothing$ or $S_R = \varnothing$. This completes the proof.

Corollary 4.2. *For each adiabatic incremental adjustment for ν, we have $\Delta\eta^{\min} < 0$.*

Proof: According to the presentation in 3.2.B, based on Lemma 4.3 and (25), it is easy to prove that $\Delta\eta^{\min} = \max\left\{\Delta\eta^{Case1}, \Delta\eta^{Case2}, \Delta\eta^{Case3}, \Delta\eta^{Case4}\right\} = \max\{<0, <0, <0, <0\} < 0$.

Lemma 4.4. *During the adiabatic incremental adjustments for ν, if $S_L = \varnothing$, then we have $\beta_i^* = 0$, $\tau_i = -2/\vartheta$, $\tau_i^* = 0$, $\forall i \in F$, if $S_R = \varnothing$, and then we have $\beta_i^* = 0$, $\tau_i = 0$, $\tau_i^* = -2/\vartheta$, $\forall i \in F$.*

Proof: If $S_L = \varnothing$, then the matrices $\mathbf{P}$ and $\mathbf{N}$ change into $\mathbf{P} = \begin{bmatrix} 0 & 0 & -\mathbf{e}_{S_R}^T \\ 0 & \vartheta & \mathbf{e}_{S_R}^T \\ -\mathbf{e}_{S_R} & \mathbf{e}_{S_R} & \mathbf{Q}_{S_RS_R} \end{bmatrix}$ and $\mathbf{N} = \begin{bmatrix} 0 & 0 & -\mathbf{e}_{S_R}^T \\ 0 & 0 & \mathbf{e}_{S_R}^T \\ -\mathbf{e}_{S_R} & \mathbf{e}_{S_R} & \mathbf{Q}_{S_RS_R} \end{bmatrix}$, respectively.

According to (22) and the definition of the inverse matrix, we have

$$\beta_i^* = -(-1)^{i_i+i_\rho}\det(\mathbf{P}_{\backslash\rho i})/\det(\mathbf{P}) = \det(\mathbf{P}_{\backslash\rho i})/\det(\mathbf{P}), \forall i \in F$$

$$\beta_b = -(-1)^{i_b+i_\rho}\det(\mathbf{P}_{\backslash\rho b})/\det(\mathbf{P}) = \det(\mathbf{P}_{\backslash\rho b})/\det(\mathbf{P})$$

$$\beta_\rho = -(-1)^{i_\rho+i_\rho}\det(\mathbf{P}_{\backslash\rho\rho})/\det(\mathbf{P}) = -\det(\mathbf{P}_{\backslash\rho\rho})/\det(\mathbf{P})$$

It is easy to verify that $\det(\mathbf{P}_{\backslash\rho i}) = 0$, and then based on Theorem 4.1, we have $\beta_i^* = 0$, $\forall i \in F$.

It is also easy to verify that $\det(\mathbf{N}) = 0$, and then from the proof of Theorem 4.1, we have $\det(\mathbf{P}) = \det(\mathbf{N}) + \vartheta\det(\mathbf{P}_{\backslash\rho\rho}) = \vartheta\det(\mathbf{P}_{\backslash\rho\rho})$, which means $\beta_\rho = -1/\vartheta$. Furthermore, we can prove that $\det(\mathbf{P}_{\backslash\rho b}) = q$ and $\det(\mathbf{P}) = \vartheta\det(\mathbf{P}_{\backslash\rho\rho}) = -\vartheta q$, so we have $\beta_b = -1/\vartheta$.

Then according to (23) and (24), we have $\tau_i = \beta_b + \beta_\rho = -2/\vartheta$ and $\tau_i^* = -\beta_b + \beta_\rho = 0$, $\forall i \in F$, respectively.

Similarly, if $S_R = \varnothing$, then we can also prove that $\beta_i^* = 0$, $\tau_i = 0$, $\tau_i^* = -2/\vartheta$, $\forall i \in F$. This completes the proof.

Assumption 4.2. *If $S_L = \varnothing$, then $E_L \neq \varnothing$; if $S_R = \varnothing$, then $E_R \neq \varnothing$.*

Simply speaking, Assumption 4.2 assumes that there does not exist the case of $S_L = \varnothing$ and $E_L = \varnothing$ or $S_R = \varnothing$ and $E_R = \varnothing$ during the adiabatic incremental adjustments for ν. We call these two cases the exceptions. In fact, the occurrence of exceptions is infrequent during the adiabatic incremental adjustments for ν, which is verified by the experimental results in Section 5.

If an exception occurs, from Lemma 4.4 and the presentation in 3.2.B, the minimal adjustment quantity of $\Delta\eta^{\min} = \max\{-\infty, -\infty, -\infty, -\infty\} = -\infty$. This implies that there does not exist any change in the composition of the set S, i.e., the corresponding solution will not vary with $\Delta\eta$. Therefore, the ν-SvrPath must terminate.

Theorem 4.3. *During the adiabatic incremental adjustments for ν, the convex quadratic objective function W in (7) is strictly monotonically increasing.*

Proof: Suppose that the previous adjustment is indexed by $k-1$, the immediate next adjustment is indexed by k, note that $\boldsymbol{\beta}_E^{(*)} = 0$ and $\boldsymbol{\beta}_R^{(*)} = 0$, according to (7), (10), (12), (18) and (22)-(24), we have:

$$
\begin{aligned}
W^{[k]} = & W^{[k-1]} + \sum_{i\in F} g_i^{[k-1]} \left(\beta_i^{[k-1]} - \beta_i^{*[k-1]}\right) \Delta\eta^{[k-1]} \\
& + \frac{1}{2}\sum_{i\in F} \tau_i^{[k-1]} \left(\beta_i^{[k-1]} - \beta_i^{*[k-1]}\right) \left(\Delta\eta^{[k-1]}\right)^2 - \frac{1}{2}\beta_\rho^{[k-1]} \left(\vartheta + \beta_\rho^{[k-1]}\right) \left(\Delta\eta^{[k-1]}\right)^2 \\
& + \beta_\rho^{[k-1]} \left(\sum_{i\in F} \left(\alpha_i^{[k-1]} + \alpha_i^{*[k-1]}\right) - l \right) \Delta\eta^{[k-1]} \\
= & W^{[k-1]} - \frac{1}{2}\beta_\rho^{[k-1]} \left(\vartheta + \beta_\rho^{[k-1]}\right) \left(\Delta\eta^{[k-1]}\right)^2 \\
& + \beta_\rho^{[k-1]} \left(\sum_{i\in F} \left(\alpha_i^{[k-1]} + \alpha_i^{*[k-1]}\right) - l \right) \Delta\eta^{[k-1]}.
\end{aligned}
$$

In other words,

$$
W^{[k]} - W^{[k-1]} = \beta_\rho^{[k-1]} \Delta\eta^{[k-1]} \left(\sum_{i\in F} \left(\alpha_i^{[k-1]} + \alpha_i^{*[k-1]}\right) - l - \frac{1}{2}\left(\vartheta + \beta_\rho^{[k-1]}\right) \Delta\eta^{[k-1]} \right).
$$

Then according to Corollary 4.2 and (25), it is easy to verify that

$$
\Delta\eta^{[k-1]} \left(\sum_{i\in F} \left(\alpha_i^{[k-1]} + \alpha_i^{*[k-1]}\right) - l - \frac{1}{2}\left(\vartheta + \beta_\rho^{[k-1]}\right) \Delta\eta^{[k-1]} \right) > 0.
$$

Furthermore, according to the proof of Lemma 4.3, it is easy to prove that $\beta_\rho^{[k-1]} > 0$. Therefore, we have $W^{[k]} - W^{[k-1]} > 0$, i.e., the convex quadratic objective function W in (7) is strictly monotonically increasing. This completes the proof.

Theorem 4.4. *During the adiabatic incremental adjustments for ν, the convex quadratic objective function W in (7) will converge to the optimal solution of* $\min_{0\le\alpha_i^{(*)}\le 1} W$ *within finite steps.*

Based on Theorem 4.3 and the strong duality theorem of the convex quadratic programming problem [19], it is easy to verify that the conclusion in Theorem 4.4 is correct. The detailed proof of Theorem 4.4 is omitted here, and a similar analysis can be found in [20].

5. **Simulation Experiments.** The simulation experiments include two parts: the first part is the verification experiments of the ν-SvrPath, and the second part is comparison with the SvrPath.

Table 2 summarizes the three benchmark datasets used in our experiments, which can be downloaded from http://archive.ics.uci.edu/ml/datasets.html. The datasets are randomly partitioned into the training set and the test set, and some are selected as the validation test. For each dataset, the validation test is used in a 5-fold CV procedure to obtain the optimal ν, and then the minimum regression error (MRE) can be obtained using the optimal parameter on the test set.

TABLE 2. Three benchmark datasets used in our experiments

Dataset	**Attributes**	**Maximal training set**	**Validation set**	**Test set**
Triazines	60	186	60	60
MPG	8	392	80	80
Housing	13	506	120	120

All experiments are performed on a 3.1 GHz Inter® Core$^{\text{TM}}$ i5-2400 with 4GB RAM and MATLAB 2010a platform. According to the description after Assumption 4.1, the radial basis function, $K(\mathbf{x}_i, \mathbf{x}_j) = \exp(||\mathbf{x}_i - \mathbf{x}_j||^2/2\sigma^2)$, is used in our experiment, where the kernel width parameter σ is set as 0.7071, 2.2361 and 7.0711, respectively. The function of the regularization parameter C is to transform the original training set T into the new training set F, so for simplicity, C is set as 1. The parameter ϑ is fixed at -1, because it is easy to verify that ϑ only depends on $\det(\mathbf{Q}_{SS})$, and ϑ is independent of the membership changes of the sets S, E, and R. Moreover, if an exception occurs, we reestablish the initial solution for a larger value $\nu \leftarrow \nu + 0.001$.

5.1. **The verification experiments of the ν-SvrPath.** In order to demonstrate the effectiveness of the ν-SvrPath, we count the numbers of "contradictions", "exceptions" and "steps", where "contradictions" represents the two cases of contradiction as shown in Table 1, "exceptions" stands for the events as described in Assumption 4.2, and "steps" denotes the iterations of adiabatic incremental adjustments for ν.

A. Triazines Dataset. The task of triazines dataset is to predict the qualitative structure activity relationships. The dataset has 186 instances with 60 continuous attributes. Table 3 presents the numbers of contradictions, exceptions and steps over 50 trials with the training data size of 40, 80, 120 and 160, respectively. From Table 3, it is obvious that the occurrences of contradictions and exceptions are infrequent, which means the ν-SvrPath can effectively avoid the infeasible updating path during the adiabatic incremental adjustments for ν. In addition, the number of steps shows that the ν-SvrPath will converge to the optimal solution of minimization problem within finite steps.

B. MPG Dataset. The MPG dataset was taken from the StatLib Library which is maintained at Carnegie Mellon University. The data concerns city-cycle fuel consumption in miles per gallon, to be predicted in terms of 3 multivalued discrete and 5 continuous attributes. Table 4 presents the numbers of contradictions, exceptions and steps over 50 trials with the training data size of 80, 160, 240 and 320, respectively. From Table 4,

Table 3. Results of the ν-SvrPath on the Triazines dataset

Dataset size		**40**	**80**	**120**	**160**
$\sigma = 0.7071$	contradictions	0.4	1.3	2.4	2.5
	exceptions	0.1	0.3	0.2	0.0
	steps	72.2	125.6	176.4	157.3
$\sigma = 2.2361$	contradictions	0.5	1.3	2.3	2.4
	exceptions	0.3	0.1	0.5	0.1
	steps	73.2	85.4	148.3	186.5
$\sigma = 7.0711$	contradictions	0.7	1.8	2.0	2.3
	exceptions	0.2	0.3	0.1	0.1
	steps	75.2	88.6	132.6	188.5

Table 4. Results of the ν-SvrPath on the MPG dataset

Dataset size		**80**	**160**	**240**	**320**
$\sigma = 0.7071$	contradictions	15.9	26.3	14.3	1.2
	exceptions	1.1	1.3	0.7	0.6
	steps	36.5	82.3	134.2	6.5
$\sigma = 2.2361$	contradictions	19.4	40.2	32.1	2.3
	exceptions	1.2	1.0	1.1	0.7
	steps	45.3	135.6	189.6	17.6
$\sigma = 7.0711$	contradictions	20.6	18.2	18.5	1.8
	exceptions	1.2	1.0	0.8	0.8
	steps	52.3	80.2	145.3	36.4

it is clear that the number of exceptions is much less than that of contradictions and the occurrences of exceptions are infrequent. This proves that the ν-SvrPath can avoid the infeasible updating path as far as possible. Furthermore, from the number of steps in Table 4, we can draw the conclusion that the ν-SvrPath will converge to the optimal solution of minimization problem within finite steps.

C. Housing Dataset. The housing dataset was taken from the StatLib Library which is maintained at Carnegie Mellon University. The dataset has 506 instances with 13 continuous attributes and 1 binary-valued attribute, which concerns housing values in suburbs of Boston. Table 5 presents the number of contradictions, exceptions and steps over 50 trials with the training data size of 120, 240, 360 and 480, respectively. Table 5 shows that the number of exceptions is much less than that of contradictions and the occurrences of exceptions are infrequent. Therefore, the ν-SvrPath can avoid the infeasible updating path as far as possible. Moreover, the number of steps in Table 5 demonstrates that the ν-SvrPath will converge to the optimal solution of minimization problem within finite steps.

Furthermore, we can also verify that the set S will always be nonempty and there always exists the inverse matrix $\mathbf{M}$ for $\mathbf{P}$ on three benchmark datasets, which will ensure that the feasible updating path for ν is reliable.

5.2. **Comparison with the SvrPath.** In order to demonstrate the superiority of the ν-SvrPath, we compare the ν-SvrPath with the SvrPath proposed in [13]. The parameter of insensitive loss function ε in the SvrPath is fixed at 1. The size of validation set is shown in Table 2. According to GCV standard, model selection was done based on the ν-solution path obtained from the SvrPath and the ν-SvrPath, respectively. Table 6 presents the

Table 5. Results of the ν-SvrPath on the Housing dataset

Dataset size		120	240	360	480
$\sigma = 0.7071$	contradictions	16.1	25.8	14.5	1.1
	exceptions	1.1	1.0	0.8	0.7
	steps	36.2	82.01	133.4	6.5
$\sigma = 2.2361$	contradictions	17.2	42.3	28.3	1.6
	exceptions	0.9	1.0	0.9	0.7
	steps	42.5	134.2	185.3	17.6
$\sigma = 7.0711$	contradictions	21.3	18.2	18.6	1.8
	exceptions	1.1	1.0	0.8	0.8
	steps	53.2	79.8	134.6	36.1

Table 6. MRE between the SvrPath and the ν-SvrPath

Dataset set		Triazines	MPG	Housing
$\sigma = 0.7071$	SvrPath	12.12	8.24	9.63
	ν-SvrPath	9.23	6.52	8.42
$\sigma = 2.2361$	SvrPath	10.65	7.64	9.72
	ν-SvrPath	9.12	6.42	8.43
$\sigma = 7.0711$	SvrPath	15.32	9.64	12.68
	ν-SvrPath	13.12	7.46	10.13

MRE based on the size of test set as shown in Table 2. It is clear that the ν-SvrPath has smaller MRE, which can be interpreted by the fact that the parameter of ν-SVR is easier to be tuned than ε-SVR. This means that the ν-SvrPath is more effective than the SvrPath.

In summary, there always exists a feasible updating path for the ν-SvrPath, and the ν-SvrPath will converge to the optimal solution of minimization problem within finite steps. Moreover, the MRE of the ν-SvrPath is smaller than the SvrPath, which further verifies the effectiveness and advantage of the ν-SvrPath.

6. **Conclusions.** This paper investigates the effective ν-solution path problem for ν-SVR. Based on the equivalent formulation of ν-SVR and a novel strategy, we propose the ν-SvrPath and present its ν-solution path. Theoretical analysis and simulation experiment results verify that the ν-SvrPath is effective. Furthermore, the ν-SvrPath is superior to the SvrPath.

In fact, the ν-SvrPath can be directly applied to a broader class of learning machines which have several equality constraints, such as incremental learning for ν-support vector classification [21], incremental learning for ν-SVR [22], and incremental learning for support vector ordinal regression [23].

Unfortunately, we need Assumptions 4.1 and 4.2 for the proofs. In addition, we just assume that only one sample can migrate from set to set at any given time. Can multiple samples migrate from set to set at the same time? There is no apparent reason that this is impossible. We hope this question can be answered sometime in the future.

Acknowledgement. This work was supported by the National Natural Science Foundation of China under Grant No. 61273131. The authors give special thanks to anonymous reviewers for their constructive comments and suggestions on the earlier version of this paper.

REFERENCES

[1] V. Vapnik, *Statistical Learning Theory*, John Wiley and Sons, Inc., New York, 1998.

[2] J. H. Ruan, Y. Shi and J. Yang, Forest fires burned area prediction based on support vector machines with feature selection, *ICIC Express Letters*, vol.5, no.8(A), pp.2597-2603, 2011.

[3] J. H. Ruan, X. P. Wang and Y. Shi, Developing fast predictors for large-scale time series using fuzzy granular support vector machines, *Applied Soft Computing*, vol.13, no.9, pp.3981-4000, 2013.

[4] D. Anguita, S. Ridella and F. Rivieccio, K-fold generalization capability assessment for support vector classifiers, *Proc. of the International Joint Conf. on Neural Networks*, Montreal, Canada, pp.855-858, 2005.

[5] O. Bousquet and A. Elisseeff, Stability and generalization, *Journal of Machine Learning Research*, vol.2, no.3, pp.499-526, 2002.

[6] S. Keerthi, V. Sindhwani and O. Chapelle, An efficient method for gradient-based adaptation of hyperparameters in SVM models, *Advances in Neural Information Processing Systems 19*, Cambridge, USA, pp.673-680, 2007.

[7] T. Hastie, S. Rosset, R. Tibshirani et al., The entire regularization path for the support vector machine, *Journal of Machine Learning Research*, vol.5, pp.1391-1415, 2004.

[8] L. Gunter and J. Zhu, Computing the solution path for the regularized support vector regression, *Advances in Neural Information Processing Systems 18*, Cambridge, USA, pp.483-490, 2005.

[9] L. Gunter and J. Zhu, Efficient computation and model selection for the support vector regression, *Neural Computation*, vol.19, no.6, pp.1633-1655, 2007.

[10] G. Wang, D. Y. Yeung and F. H. Lochovsky, A new solution path algorithm in support vector regression, *IEEE Trans. Neural Networks*, vol.19, no.10, pp.1753-1767, 2008.

[11] B. Schölkopf, A. J. Smola, R. C. Williamson et al., New support vector algorithms, *Neural Computation*, vol.12, no.5, pp.1207-1245, 2000.

[12] I. Steinwart, On the optimal parameter choice for the ν-support vector machines, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.25, no.10, pp.1274-1284, 2003.

[13] G. Loosli, G. Gasso and S. Canu, Regularization paths for ν-SVM and ν-SVR, *Advances in Neural Networks*, Heidelberg, Berlin, pp.486-496, 2007.

[14] B. Gu, J. D. Wang, G. S. Zheng et al., Regularization path for ν-support vector classification, *IEEE Trans. Neural Networks and Learning Systems*, vol.23, no.5, pp.800-811, 2012.

[15] C. C. Chang and C. J. Lin, Training ν-support vector regression: Theory and algorithms, *Neural Computation*, vol.14, no.8, pp.1959-1977, 2002.

[16] M. Martin, On-line support vector machines for function approximation, *Proc. of the 13th European Conf. on Machine Learning*, London, UK, pp.282-294, 2002.

[17] J. Ma, J. Theiler and S. Perkins, Accurate on-line support vector regression, *Neural Computation*, vol.15, no.11, pp.2683-2703, 2003.

[18] A. S. Householder, *The Theory of Matrices in Numerical Analysis*, Blaisdell Publishing Company, New York, 1974.

[19] D. P. Bertsekas, *Convex Optimization Theory*, Tsinghua University Press, Beijing, 2011.

[20] B. Gu, G. S. Zheng and J. D. Wang, Analysis for incremental and decremental standard support vector machine, *Journal of Software*, vol.24, no.7, pp.1601-1613, 2013.

[21] B. Gu, J. D. Wang, Y. C. Yu et al., Accurate on-line ν-support vector learning, *Neural Networks*, vol.27, pp.51-59, 2012.

[22] B. Gu, V. S. Sheng, Z. J. Wang et al., Incremental learning for ν-support vector regression, *Neural Networks*, vol.67, pp.140-150, 2015.

[23] B. Gu, V. S. Sheng, K. Y. Tay et al., Incremental support vector learning for ordinal regression, *IEEE Trans. Neural Networks and Learning Systems*, vol.26, no.7, pp.1403-1416, 2015.