Abstract—Support Vector Ordinal Regression (SVOR) is a popular method to tackle ordinal regression problems. However, until now there are no effective algorithms proposed to address incremental SVOR learning due to the complicated formulations of SVOR. Recently, an interesting accurate on-line algorithm was proposed for training λ-Support Vector Classification, which can handle a quadratic formulation with a pair of equality constraints. In this paper, we first present a modified SVOR formulation based on a sum-of-margins strategy, which has multiple constraints of the mixture of an equality and an inequality. Then, we extend the accurate on-line λ-Support Vector Classification algorithm to the modified formulation, and propose an effective incremental SVOR algorithm, which can handle a quadratic formulation with multiple constraints of the mixture of an equality and an inequality, and tackle the conflicts between the equality and inequality constraints. We also provide its finite convergence analysis. Numerical experiments on the several benchmark and real-world datasets show that the incremental algorithm can converge to the optimal solution in finite steps, and is faster than the existing batch and incremental SVOR algorithms. Meanwhile, the modified formulation has better accuracy than the existing incremental SVOR algorithm, and is as accurate as the sum-of-margins based formulation of Shasha and Levin.

Index Terms—Incremental learning, online learning, Support Vector Machine, ordinal regression.

I. INTRODUCTION

In conventional machine learning and data mining research, predictive learning has become a standard inductive learning, where different sub-problem formulations have been identified, for example, classification, metric regression, ordinal regression (OR) and so on. In OR problems, training samples are marked by a set of ranks, which exhibit an ordering among different categories. In contrast to metric regression problems, the ranks for OR are of finite types and the metric distances between the ranks are not defined; in comparison to classification problems, these ranks are also different from the labels of multiple classes due to ordering information [7]. Therefore, OR is a special case in predictive learning.

In practical OR tasks, such as information retrieval [14], collaborative filtering [5], flight delays forecasts [21] and so on, training data is usually provided one example at a time. This is a so called online scenario. We use flight delays forecasts as an example. The given flight delay data streams are non-stationary, meaning that data distributions vary over time. Batch algorithms will generally fail if such ambiguous information is present and is erroneously integrated by the batch algorithm. Incremental learning algorithms are more capable in this case, because they allow the incorporation of additional training data without re-training from scratch [17].

Ever since Vapnik’s influential work in statistical learning theory [25], Support Vector Machines (SVMs) have gained profound interest because of good generalization performance [1], [19]. There are also several Support Vector Ordinal Regression (SVOR) formulations proposed to tackle OR problems. For example, Herbrich et al. [14] gave a SVM formulation based on a loss function between pairs of ranks, which is called PSVM here. However, the problem size of PSVM is a quadratic function of the training data size. To address this problem, Shasha and Levin [5] proposed two SVM formulations by finding multiple parallel discrimination hyperplanes. One is a fixed-margin based formulation, and the other is a sum-of-margins based formulation (SMF). Chu and Keerthi [7] further improved the fixed-margin based SVOR formulation by explicitly and implicitly keeping ordinal inequalities on the thresholds, in which the explicit constraints based SVOR was called EXC. Cardoso and Pinto da Costa [22] proposed a data replication method and mapped it into SVM, which also implicitly used the fixed-margin strategy. The problem sizes of these SVOR formulations are all linear in the training data size. In addition, more recently, Seah et al. [16] presented a transductive SVM learning paradigm for OR, by taking benefits from the abundance of unlabeled patterns.

Although there exist several perceptron-like algorithms proposed for incremental OR learning (e.g. [2], [8], [9]), very little work has been done on incremental learning for SVOR. Previous works mostly focus on incremental learning for standard SVM, one-class SVM, Support Vector Regression (SVR), and so on. For example, Cauwenberghs and Poggio [6] proposed an exact incremental learning approach (the C&P algorithm) for SVM in 2001. Later, Martin [10] extended it to SVR and proposed an accurate incremental SVR algorithm.
Laskov et al. [17] implemented an accurate incremental one-class SVM algorithm. Karasuyama and Takeuchi [24] gave an extended algorithm that can handle multiple data samples simultaneously. Recently, Gu et al. [27] extended the C&P algorithm to ν-Support Vector Classification (ν-SVC) and proposed an effective accurate on-line ν-SVC algorithm (AONSVM), which can handle the conflict between a pair of equality constraints during the process of incremental learning. Gu and Sheng [28] proved the feasibility and finite convergence of AONSVM under two assumptions (i.e. Assumption 1 and 2 as mentioned in [28]).

To the best of our knowledge, the PSVM based incremental algorithm (IPSVM) [11] is the only work on incremental SVOR learning. As mentioned above, this approach is limited by the size of the problem, which is quadratic in the training data size. Therefore, it is highly desirable to design an effective incremental learning algorithm for the SVOR formulations, whose problem size is linear in the training data size. In this paper, we focus on the SMF of Shashua and Levin [5]. We first present a modified SMF (called MSMF), which has multiple constraints of the mixture of an equality and an inequality. Then, we extend AONSVM to MSMF, and propose an effective incremental SVOR algorithm (called ISVOR). The incremental algorithm includes two steps, i.e., relaxed adiabatic incremental adjustments (RAIA), and strict restoration adjustments (SRA). Based on the two steps, the incremental algorithm can handle inequality constraints, and can tackle the conflicts between the equality and inequality constraints. We also provide its finite convergence analysis. Numerical experiments show that ISVOR can converge to the optimal solution in finite steps, and is faster than the existing batch and incremental SVOR algorithms. Meanwhile, the modified formulation has better accuracy than the existing incremental SVOR algorithm, and is as accurate as the SMF of Shashua and Levin [5].

The main contributions of this paper are summarized as follows:

1) We propose an effective incremental SVOR algorithm (i.e., ISVOR), whose problem size is linear in the training data size. We also prove the finite convergence of ISVOR. Numerical experiments show that ISVOR is faster than the existing batch and incremental SVOR algorithms.

2) The existing incremental SVM algorithms can handle a quadratic formulation with a pair of equality constraints or an equality constraint for a binary classification problem. ISVOR can handle a quadratic formulation with multiple constraints of the mixture of an equality and an inequality for multiple binary classification problems. ISVOR can be viewed as a generalization of the existing incremental SVM algorithms.

The rest of this paper is organized as follows. Section II gives a modified SVOR formulation (i.e. MSMF), and its Karush-Kuhn-Tucker (KKT) conditions. The Incremental SVOR algorithm is presented in Section III. The experimental setup, results and discussions are presented in Section IV and V. The last section gives some concluding remarks.

Notation: In order to make notations easier to follow, we give a summary of the notations in the following list.

\( \alpha_i, g_i \): The \( i \)-th element of a vector \( \alpha \) and \( g \).

\( \alpha_e, y_e, J_e \): The weight, label of a candidate sample \((x_e, y_e)\), and the index of the two-class sample set \( S_l^j \) to which \((x_e, y_e)\) belongs.

\( \Delta \): The complement of the set \( J \), the contracted set of \( J \) by deleting \( j_c \), and the enlarged set of \( J \) by adding \( j_c \).

\( d_{j_c}^l, E_{j_c} \): The subvector of \( d \) by extracting the elements indexed by \( J \), and a submatrix of \( E \) by extracting the columns indexed by \( J \).

\( Q_{S_l^j} \): The submatrix of \( Q \) with the rows and columns indexed by \( S_l^j \).

\( \Delta d_{j_c}^l, \Delta \alpha_i \): The columns corresponding to \( \Delta d_{j_c}^l \) indexed by \( J_c \) and \( \Delta \alpha_i \) indexed by \( M \).

\( M^T \): The transpose of the matrix \( M \).

\( \mathbf{0}, \mathbf{1} \): A zero matrix with proper dimensions, and \( \mathbf{1} \times \mathbf{1} \) matrix with all zeroes except that \( O_{j_c j_c} = 1 \).

\( u_{j_c}, y_{j_c} \): A \((r-1)\)-dimensional column vector with all zeroes except that the \( j_c \)-th position is equal to \( y_{j_c} \) and one respectively.

\( c_{S_l^j} \): A \(|S_l^j|\)-dimensional column vector with all zeroes except that the positions corresponding to the samples \((x_i, y_i)\) of \( S_l^j \) are equal to \(-1\) and \( y_i \) respectively.

II. A MODIFIED SVOR FORMULATION

In this section, we first review SMF, then present MSMF and its dual problem. Finally, we present the KKT conditions for the solution of the dual problem.

A. Review of SMF

Without loss of generality, we consider an OR problem with \( r \) ordered categories and denote these categories as consecutive integers \( Y = \{1, 2, \cdots, r\} \) to keep the known ordering information. The number of training samples in the \( j \)-th category \((j \in Y)\), is denoted as \( n_l^j \), and the \( i \)-th training sample is denoted as \( x_i = (x_i^1, \cdots, x_i^r) \) where \( X_i \in \mathbb{R}^d \).

To learn a mapping function \( r() : X \rightarrow Y \), Shashua and Levin [5] considered \( r-1 \) parallel discrimination hyperplanes, i.e., \( \langle w, x \rangle - b_j with b_j \leq \cdots \leq b_{r-1} \), where \( b_j \) is the threshold of the \( j \)-th discrimination hyperplane. Supposing \( b_r = \infty \), the decision mapping function \( r() \) can be denoted as

\[
 r(x) = \min_{j \in Y} \{ j : \langle w, x \rangle - b_j < 0 \} \tag{1}
\]

Let \( d_j \geq 0 \) be the shortest distance from the \( j \)-th discrimination hyperplane to the closest sample in the \( j \)-th or \((j+1)\)-th category, which is the margin of the \( j \)-th discrimination
The key of such an approach is to keep the thresholds \( f \) with unit magnitude in optimizing the objective function. Furthermore, the data slack variable measuring the degree of misclassification of \( x \) is to use the popular implicit approach to achieve the ordinal thresholds. Thus, a modified formulation of (2) is used here by discarding the constraint of \( b_j + d_j \leq b_{j+1} - d_{j+1} \). After discarding the constraint, our proposed OR formulation (i.e. MSMF) is more favorable to design an incremental SVOR algorithm, because the primal variables \( b_j \) and \( d_j \) can be induced directly in KKT conditions (see Section II-C).

To present the dual function of the modified formulation in a compact form, we introduce some new notations:

1. Based on the reduction framework of [19], OR can be regarded as \( r - 1 \) binary classification. Thus, we define the two-class training sample set \( S^j = \{ (x_i^j, y_i^j = -1) \}_{i=1}^{n_j} \cup \{ (x_i^j, y_i^j = +1) \}_{i=n_j+1}^{r n_j} \}, \) and the extended training sample set \( S = \bigcup_{j=1}^r S^j = \{ (x_i, y_i) , \cdots , (x_{i}, y_i) \}, \) where \( l = 2 \times \sum_{j=1}^r n_j - n^j - n \).

2. We let \( \lambda_j = [ \lambda_j^1, \cdots, \lambda_j^{n_j} ] \) and \( \delta_j = [ \delta_j^1, \cdots, \delta_j^n ] \), where \( \lambda_j^i \) and \( \delta_j^i \) are the Lagrangian multipliers corresponding to the first and third inequality constraints in (2), respectively. Thus, \( \alpha = [ \lambda^1, \delta^1, \cdots, \lambda^{r-1}, \delta^{r-1} ] \) is defined to be the row vector holding all the \( \lambda_j^i \) and \( \delta_j^i \) Lagrangian multipliers.

3. We define the kernel matrix \( Q \) as \( Q_{ik} = y_i y_k K(x_i, x_k) \) for all \( 1 \leq i, k \leq l \).

Based on the above notations, the dual problem can be formulated as follows:

\[
\min_{\alpha} \quad \frac{1}{2} \alpha Q \alpha^T \\
\text{s.t.} \quad \sum_{i \in S^j} \alpha_i = 0, \quad \sum_{i \in S^j} \alpha_i \geq 2, \quad j = 1, \cdots, r - 1, \quad 0 \leq \alpha_i \leq C, \quad i = 1, \cdots, l,
\]

where \( i \in S^j \) is the abbreviated form of \( (x_i, y_i) \in S^j \).

Once the optimal solution \( \alpha \) is obtained, the part \( \langle w, \phi(x) \rangle \) of the rank-monotonic mapping function \( f(x, j) \) in RKHS can be obtained as follows:

\[
\langle w, \phi(x) \rangle = \sum_{i=1}^l y_i \alpha_i K(x_i, x) \sqrt{\alpha Q \alpha^T}
\]

And \( b_j \) can be obtained by solving the following linear equations:

\[
\sum_{i=1}^l y_i \alpha_i K(x_i, x) = b_j + d_j = 0
\]

\[
\lambda_j = \sum_{i=1}^l y_i \alpha_i K(x_i, x) = b_j - d_j = 0 \quad (6)
\]

where \( \{ (x_i, y_i), (x_{i+1}, y_{i+1}) \} \subseteq S^j \) with \( y_1 = -1 \) and \( y_{n_j} = +1 \), and \( x_i, x_{i+1} \) are also support vectors with their weights \( 0 < \alpha_i < C, 0 < \alpha_{i+1} < C \).
C. KKT Conditions

According to convex optimization theory [3], the solution of the dual problem (3) can be obtained by the following min-max problem:

$$\min_{0 \leq \alpha_i \leq C} \max_{b'_j, d'_j \geq 0} W = \frac{1}{2} \sum_{i,k=1}^{l} \alpha_i \alpha_k Q_{ik} + \sum_{j=1}^{r-1} b'_j \left( \sum_{i \in S_j} y_i \alpha_i \right) + \sum_{i \in S_j} d'_j \left( 2 - \sum_{i \in S_j} \alpha_i \right)$$  \hspace{1cm} (7)

where $b'_j \in \mathbb{R}$ and $d'_j \in \mathbb{R}^+$ are Lagrangian multipliers.

From the KKT theorem [4], we obtain the following KKT conditions:

$$\sum_{i \in S_j} y_i \alpha_i = 0 \quad (8)$$

$$p_j \text{ def} = \sum_{i \in S_j} \alpha_i \begin{cases} \geq 2 & \text{for } d'_j = 0 \\ = 2 & \text{for } d'_j > 0 \end{cases}$$  \hspace{1cm} (9)

$$\forall i \in S^j : g_i \text{ def} = \frac{\partial W}{\partial \alpha_i} = \sum_{k=1}^{l} \alpha_k Q_{ik} + y_i b'_j - d'_j \begin{cases} \geq 0 & \text{for } \alpha_i = 0 \\ = 0 & \text{for } 0 < \alpha_i < C \\ \leq 0 & \text{for } \alpha_i = C \end{cases}$$  \hspace{1cm} (10)

According to the value of the function $g_i$, a two-class training sample set $S^j$ associated with the $j$-th binary classification is partitioned into three independent sets (see Fig. 2):

1) $S^S_j = \{ i \in S^j : g_i = 0, 0 < \alpha_i < C \}$, the set $S^S_j$ includes margin support vectors strictly on the margins;
2) $S^E_j = \{ i \in S^j : g_i \leq 0, \alpha_i = C \}$, the set $S^E_j$ includes error support vectors exceeding the margins;
3) $S^R_j = \{ i \in S^j : g_i \geq 0, \alpha_i = 0 \}$, the set $S^R_j$ includes the remaining vectors ignored by the margins.

Thus, the extended training sample set $S$ is partitioned into three independent sets, i.e., $S_S = \bigcup_{j=1}^{r-1} S^S_j$, $S_E = \bigcup_{j=1}^{r-1} S^E_j$, and $S_R = \bigcup_{j=1}^{r-1} S^R_j$.

In addition, according to the value of $p_j$ in (9), we can define an active set $J \subseteq \{1, \ldots, r-1\}$ with $p_j = 2$ and $d'_j > 0$ for all $j \in J$.

![Fig. 2: Partitioning a two-class training sample set $S^j$, which is associated with the $j$-th binary classification, into three independent sets by KKT-conditions: (a) $S^S_j$, (b) $S^E_j$, (c) $S^R_j$.](image)

III. INCREMENTAL SVOR LEARNING

In this section, we consider the incremental SVOR learning algorithm for the dual problem (3). When a sample $x_{new}$ is added into the $j$-th category, there correspondingly exists increments in $S$ and $S^j$ (see Table I). We define the increments as $S_{new}$ and $S^j_{new}$, respectively. Initially, we set the weight $\alpha_c$ of each sample $(x_c, y_c)$ in $S_{new}$ to zero. If this assignment satisfies the KKT conditions, adjustments are not needed. However, if this assignment violates the KKT conditions, additional adjustments become necessary (see Fig. 1(b)). The goal of the incremental SVOR algorithm is to find an effective method for updating the weights without re-training from scratch, when a sample in $S_{new}$ violates the KKT conditions.

**TABLE I: Three cases of the change of the extended training sample set $S$, when a sample $x_{new}$ is added into the $j$-th category.** $S^j_{new}$ denotes the increment in $S^j$, and $S_{new}$ denotes the increment in $S$, where $S_{new} = S^j_{new} \cup S^j_{new}$.

<table>
<thead>
<tr>
<th>$x_{new}$</th>
<th>$S_{new}$</th>
<th>$S^j_{new}$</th>
<th>$S^j_{new}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_{new}$</td>
<td>$\emptyset$</td>
<td>${x_{new}, -1}$</td>
<td>${x_{new}, -1}$</td>
</tr>
<tr>
<td>$x_{new}$</td>
<td>${x_{new}, +1}$</td>
<td>$\emptyset$</td>
<td>${x_{new}, +1}$</td>
</tr>
</tbody>
</table>

**TABLE II: Two cases of conflicts between $\sum_{i \in S_{jc}} \Delta \alpha_i + \Delta \alpha_c = 0$ and $\sum_{i \in S_{jc}} y_i \Delta \alpha_i + y_c \Delta \alpha_c = 0$, when $|\sum_{i \in S_{jc}} y_i| = |S_{jc}^j|$ and $d'_j > 0$ with a small increment $\Delta \alpha_c$.**

| $\sum_{i \in S_{jc}} \Delta \alpha_i + \Delta \alpha_c = 0$ | $\sum_{i \in S_{jc}} y_i \Delta \alpha_i + y_c \Delta \alpha_c = 0$ | $|\sum_{i \in S_{jc}} y_i| = |S_{jc}^j|$ | $d'_j > 0$ with a small increment $\Delta \alpha_c$ |
|-------------------------------|-------------------|-----------------|------------------|
| $\alpha_c > 0$ | $\alpha_c < 0$ | $\alpha_c = 0$ | $\alpha_c = 0$ |

Compared with the formulations of standard SVM, one-class SVM, SVR, and $\nu$-SVC, our SVOR formulation (3) has the following challenges, which prevent us from directly using the existing incremental SVM algorithms, including the C&P algorithm and AONSVM, on (3).

1) If $\sum_{i \in S_{jc}} y_i = |S_{jc}^j|$, $d'_j > 0$, and the label of an added sample $(x_c, y_c)$ in $S_{jc}^j$ is different from those of the margin support vectors in $S_{jc}^j$, there exists a conflict (referred to as Conflict-1) between (8) and (9) with a small increment of $\alpha_c$ (see Table II). Conflict-1 is different to the one in $\nu$-SVC, because an additional condition $d'_j > 0$ must be considered here.

2) The SVOR formulation (3) has multiple constraints of the mixture of an equality and an inequality, which is more complicated than a pair of equality constraints in $\nu$-SVC, and an equality constraint in standard SVM, one-class SVM, and SVR.

To address these challenges, we propose an incremental SVOR algorithm (i.e. ISVOR, see Algorithm 1), which includes two steps, similarly to AONSVM.

The first step is RAIA. Because there may exist Conflict-1 between (8) and (9) as shown in Table II, the feasible updating path leading to the eventual satisfaction of the KKT conditions will not be guaranteed. To overcome this problem, the limitation on the enlarged $j$-th two-class training samples imposed by inequality (9) is removed from this step, similarly to AONSVM. And our basic idea is gradually increasing $\alpha_c$ under the condition of rigorously keeping all the samples satisfying the KKT conditions, except that the inequality...