

Rigorous Proof of Termination of SMO Algorithm for Support Vector Machines

Author(s): Norikazu Takahashi and Tetsuo Nishi

Journal: IEEE Transactions on Neural Networks

Volume: 16

Number: 3

Pages: 774–776

Month: May

Year: 2005

Published Version: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1427778

©2005 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Rigorous Proof of Termination of SMO Algorithm for Support Vector Machines

Norikazu Takahashi and Tetsuo Nishi

Abstract—Sequential minimal optimization (SMO) algorithm is one of the simplest decomposition methods for learning of support vector machines (SVMs). Keerthi and Gilbert have recently studied the convergence property of SMO algorithm and given a proof that SMO algorithm always stops within a finite number of iterations. In this paper, we point out the incompleteness of their proof and give a more rigorous proof.

Index Terms—Support vector machines (SVMs), sequential minimal optimization (SMO) algorithm, convergence, termination

I. INTRODUCTION

Due to their high generalization performance, support vector machines (SVMs) [1] have attracted great attention in the field of pattern recognition, machine learning, neural networks and so on. Learning of an SVM leads to a quadratic programming (QP) problem with the size being equal to the number of training samples. Since many techniques for solving QP problems are available (see [2] for example), SVM learning can be implemented with one of them. However, if the number l of training samples is very large, those conventional methods cannot be directly applied because it is impossible to store all elements of an $l \times l$ matrix in memory. To overcome this difficulty, so-called decomposition methods have been proposed [3], [4]. Given a QP problem, a decomposition method tries to find an optimal solution by solving QP subproblems with q ($\leq l$) variables iteratively. Since q is much smaller than l in general, decomposition methods can avoid the memory-related problem mentioned above.

It is very important especially from the theoretical point of view to make clear the condition under which a decomposition method converges to an optimal solution. Keerthi and Gilbert have recently studied the convergence property of sequential minimal optimization (SMO) algorithm and given a proof that it always stops within a finite number of iterations after finding an optimal solution [5]. SMO algorithm is a special type of decomposition methods such that the size q of subproblems is fixed to two. On the other hand, however, no general result on the convergence of decomposition methods have been obtained so far. Some significant results have recently been presented by Lin [6]–[8], but these results require some assumptions on convexity of QP problems and how to select q variables for updating.

The objective of this paper is to point out incompleteness of the convergence proof given by Keerthi and Gilbert [5] and make a more rigorous analysis in order to complete their proof. Making clear and rigorous analysis of the asymptotic behavior

This research was partially supported by the 21st Century COE Program ‘Reconstruction of Social Infrastructure Related to Information Science and Electrical Engineering’.

N. Takahashi and T. Nishi are with the Department of Computer Science and Communication Engineering, Kyushu University, Fukuoka, Japan email: norikazu@csce.kyushu-u.ac.jp; nishi@csce.kyushu-u.ac.jp

of SMO algorithm is very important because some techniques used for analyzing the simplest decomposition method may play important roles in convergence analysis for more general decomposition methods. We first introduce the QP problem arising in SVM learning and its optimality condition. We then describe a general form of SMO algorithm for solving the QP problem. Next we review the analytical results concerning the convergence properties of SMO algorithm given by Keerthi and Gilbert [5] and point out that their convergence proof is incomplete. Finally we give a rigorous proof that SMO algorithm stops in a finite number of iterations.

II. SVM DUAL PROBLEM AND ITS OPTIMALITY CONDITION

Given a set of training samples $\{(\mathbf{x}_i, y_i)\}_{i=1}^l$ where $\mathbf{x}_i \in \mathbb{R}^n$ is the i -th input pattern and $y_i \in \{1, -1\}$ the label of the class to which \mathbf{x}_i belongs, learning of an SVM with the kernel function $K(\cdot, \cdot)$ leads to the following quadratic programming (QP) problem.

Problem 1: Find $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \dots, \alpha_l]^T$ which minimizes

$$W(\boldsymbol{\alpha}) = \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l q_{ij} \alpha_i \alpha_j - \sum_{i=1}^l \alpha_i \quad (1)$$

under the constraints

$$\sum_{i=1}^l y_i \alpha_i = 0 \quad (2)$$

and

$$0 \leq \alpha_i \leq C, \quad i = 1, 2, \dots, l \quad (3)$$

where $q_{ij} = y_i y_j K(\mathbf{x}_i, \mathbf{x}_j)$ and C is a user-specified positive constant.

We assume hereafter that the kernel function $K(\cdot, \cdot)$ satisfies Mercer’s condition. In this case Problem 1 is a convex QP problem because the matrix $Q = [q_{ij}] \in \mathbb{R}^{l \times l}$ is positive semi-definite. Hence an $\boldsymbol{\alpha}$ is a solution of Problem 1 if and only if the KKT conditions are satisfied at $\boldsymbol{\alpha}$. Let us define two sets $I_{up}(\boldsymbol{\alpha}) \subseteq \{1, 2, \dots, l\}$ and $I_{low}(\boldsymbol{\alpha}) \subseteq \{1, 2, \dots, l\}$ as follows:

$$\begin{aligned} I_{up}(\boldsymbol{\alpha}) &= I_0(\boldsymbol{\alpha}) \cup I_1(\boldsymbol{\alpha}) \cup I_2(\boldsymbol{\alpha}) \\ I_{low}(\boldsymbol{\alpha}) &= I_0(\boldsymbol{\alpha}) \cup I_3(\boldsymbol{\alpha}) \cup I_4(\boldsymbol{\alpha}) \end{aligned}$$

where $I_0(\boldsymbol{\alpha}) = \{i \mid 0 < \alpha_i < C\}$, $I_1(\boldsymbol{\alpha}) = \{i \mid y_i = 1, \alpha_i = 0\}$, $I_2(\boldsymbol{\alpha}) = \{i \mid y_i = -1, \alpha_i = C\}$, $I_3(\boldsymbol{\alpha}) = \{i \mid y_i = 1, \alpha_i = C\}$ and $I_4(\boldsymbol{\alpha}) = \{i \mid y_i = -1, \alpha_i = 0\}$. Making use of these notations, Keerthi *et al.* [9] have shown that an $\boldsymbol{\alpha}$ is an optimal solution of Problem 1 if and only if the conditions (2), (3) and

$$\min_{i \in I_{up}(\boldsymbol{\alpha})} F_i(\boldsymbol{\alpha}) \geq \max_{i \in I_{low}(\boldsymbol{\alpha})} F_i(\boldsymbol{\alpha}) \quad (4)$$

are satisfied, where $F_i(\boldsymbol{\alpha})$ is defined by

$$F_i(\boldsymbol{\alpha}) = y_i \left(\sum_{j=1}^l q_{ij} \alpha_j - 1 \right). \quad (5)$$

Since it is usually not possible to achieve the optimality condition (4) exactly when solving Problem 1 numerically, the approximate condition

$$\min_{i \in I_{up}(\boldsymbol{\alpha})} F_i(\boldsymbol{\alpha}) \geq \max_{i \in I_{low}(\boldsymbol{\alpha})} F_i(\boldsymbol{\alpha}) - \tau \quad (6)$$

is employed instead of (4), where τ is a positive tolerance parameter [9]. In the following, any α satisfying the conditions (2), (3) and (6) is said to be a τ -optimal solution [5]. Also, any pair of indices (i, j) such that

$$i \in I_{up}(\alpha), j \in I_{low}(\alpha), F_i(\alpha) < F_j(\alpha) - \tau \quad (7)$$

is called a τ -violating pair at α . It is obvious from these definitions that a feasible solution α , that is, an α satisfying (2) and (3), is a τ -optimal solution if and only if there is no τ -violating pair at α .

Remark 1: Note that the definition of τ -violating pair given above differs from that in Reference [5]. Two statements “ (i, j) is a τ -violating pair” and “ (j, i) is a τ -violating pair” must be distinguished in this paper. In fact, it is obvious from our definition that if (i, j) is a τ -violating pair at some α then (j, i) cannot be a τ -violating pair at the same α . On the other hand, these two statements are regarded as equivalent in [5].

III. TERMINATION OF SMO ALGORITHM

SMO algorithm is a special class of decomposition methods for solving Problem 1. It was first proposed by Platt [4] and then improved by Keerthi *et al.* [9]. SMO algorithm tries to find an optimal solution of Problem 1 by iteratively solving QP problems having only two variables, which are chosen based on some criterion. A general form of SMO algorithm can be described as follows:

Algorithm 1 (SMO algorithm): Given a set of training samples $\{(x_i, y_i)\}_{i=1}^l$, kernel function $K(\cdot, \cdot)$ and a positive constant C , execute the following steps.

- 1) Set $k = 0$ and $\alpha(0) = [\alpha_1(0), \alpha_2(0), \dots, \alpha_l(0)]^T = \mathbf{0}$.
- 2) If $\alpha(k)$ satisfies (6) then stop.
- 3) Choose a τ -violating pair $(i(k), j(k))$ and solve Problem 1 under the additional constraints $\alpha_i = \alpha_i(k), \forall i \notin \{i(k), j(k)\}$. Set $\alpha(k+1) = \alpha^*$ where α^* is a solution of the above problem. Add 1 to k and go to Step 2).

Since each QP problem arising in Step 3) has only two variables, it can be solved analytically [4]. This means that SMO algorithm does not need any QP problem solver. In this sense, SMO algorithm is considered as one of the simplest decomposition methods for solving Problem 1.

One can easily see that $\alpha(k)$ generated by Algorithm 1 belongs to the feasible region of Problem 1 for all k and that $W(\alpha(k))$ is monotone decreasing with respect to k . This implies that the sequence $\{W(\alpha(k))\}_{k=0}^{\infty}$ converges to a certain value since $W(\alpha)$ is bounded from below in the feasible region of Problem 1. However, on the other hand, it is not clear whether SMO algorithm always stops within a finite number of iterations, that is, the sequence $\{\alpha(k)\}_{k=0}^{\infty}$ always converges to a τ -optimal solution of Problem 1. Therefore making clear the convergence property of the sequence $\{\alpha(k)\}_{k=0}^{\infty}$ is the most important and fundamental problem for SMO algorithm.

In the following, we will review the convergence analysis made by Keerthi and Gilbert [5] and point out that their proof for termination of SMO algorithm is not complete. Before doing so, we introduce here some notations which will be needed in later discussions. For any execution of SMO algorithm, the set of integers $L(p, q)$ is defined as

$$L(p, q) = \{k \mid (i(k), j(k)) = (p, q)\}$$

where (p, q) is any pair of indices. Let I_{∞} be the set of pairs (p, q) such that $|L(p, q)| = \infty$ where $|L(p, q)|$ represents the cardinality of the set $L(p, q)$. Obviously an execution of the SMO algorithm stops within a finite number of iterations if and only if $I_{\infty} = \emptyset$. The square region $[0, C] \times [0, C] \subset \mathbb{R}^2$ is denoted by S . The interior and boundary of S are denoted by $\text{int}S$ and ∂S , respectively. Four edges of S are represented as $E_n = (0, C) \times [C, C]$, $E_w = [0, 0] \times (0, C)$, $E_s = (0, C) \times [0, 0]$ and $E_e = [C, C] \times (0, C)$.

Keerthi and Gilbert [5] first analyzed the properties of a solution of the QP problem in Step 3) of SMO algorithm in detail, and derived the following lemma.

Lemma 1: The following statements hold true for all $k \geq 1$.

- (a) $\alpha(k+1) \neq \alpha(k)$.
- (b) Neither $(i(k), j(k))$ nor $(j(k), i(k))$ is a τ -violating pair at $\alpha(k+1)$.
- (c) If $(\alpha_{i(k)}(k+1), \alpha_{j(k)}(k+1)) \in \text{int}S$ then $F_{i(k)}(\alpha(k+1)) = F_{j(k)}(\alpha(k+1))$.
- (d) $W(\alpha(k)) - W(\alpha(k+1)) \geq \frac{\tau}{2\sqrt{2}} \|\alpha(k+1) - \alpha(k)\|$.

Combining Lemma 1 (d) and the fact that the sequence $\{W(\alpha(k))\}_{k=0}^{\infty}$ always converges to a certain value, the following lemma can be obtained [5].

Lemma 2: The sequence $\{\alpha(k)\}_{k=0}^{\infty}$ converges to a certain point in $[0, C]^l$ as $k \rightarrow \infty$ even though an execution of SMO algorithm does not stop.

In the following, the limit point of the sequence $\{\alpha(k)\}_{k=0}^{\infty}$ is denoted by $\alpha^* = [\alpha_1^*, \alpha_2^*, \dots, \alpha_l^*]^T$. Then we can easily derive the following lemma.

Lemma 3: If (p, q) belongs to I_{∞} then (q, p) does not belong to I_{∞} .

Proof: Assume that both (p, q) and (q, p) belong to I_{∞} . Then we have $F_p(\alpha(k)) < F_q(\alpha(k)) - \tau, \forall k \in L(p, q)$ and $F_q(\alpha(k')) < F_p(\alpha(k')) - \tau, \forall k' \in L(q, p)$. Letting $k \rightarrow \infty$ and $k' \rightarrow \infty$, we have $F_p(\alpha^*) \leq F_q(\alpha^*) - \tau$ and $F_q(\alpha^*) \leq F_p(\alpha^*) - \tau$ which lead to a contradiction. It is thus impossible that both (p, q) and (q, p) belong to I_{∞} . ■

The following lemma follows from Lemmas 1 and 2.

Lemma 4: Let (p, q) be any pair of indices belonging to I_{∞} . Then there exists a \bar{k} such that both $(\alpha_p(k), \alpha_q(k))$ and $(\alpha_p(k+1), \alpha_q(k+1))$ belong to ∂S for all $k \in L(p, q)$ satisfying $k \geq \bar{k}$. Moreover, the limit point (α_p^*, α_q^*) is given by the following equation.

$$(\alpha_p^*, \alpha_q^*) = \begin{cases} (0, 0) \text{ or } (C, C) & \text{if } y_p y_q = 1 \\ (0, C) \text{ or } (C, 0) & \text{if } y_p y_q = -1 \end{cases} \quad (8)$$

Figure 1 shows all possible transitions from $(\alpha_p(k), \alpha_q(k))$ to $(\alpha_p(k+1), \alpha_q(k+1))$ where $(p, q) \in I_{\infty}$ and $k \in L(p, q)$ is sufficiently large so that transition is from ∂S to ∂S .

The proof of termination of SMO algorithm given by Keerthi and Gilbert is summarized as follows: Assume first that an execution of SMO algorithm does not stop. Then I_{∞} is not empty. Let (p, q) be any pair in I_{∞} and consider for example the case where $y_p = y_q = 1$ and $(\alpha_p^*, \alpha_q^*) = (0, 0)$. Then $(\alpha_p(k), \alpha_q(k))$ approaches to $(0, 0)$ while visiting E_s and E_w alternatively. In order for the point $(\alpha_p(k), \alpha_q(k))$ moves from E_s to E_w infinitely many times, (q, p) must be a τ -violating pair infinitely many times, that is, (q, p) must belong to I_{∞} .

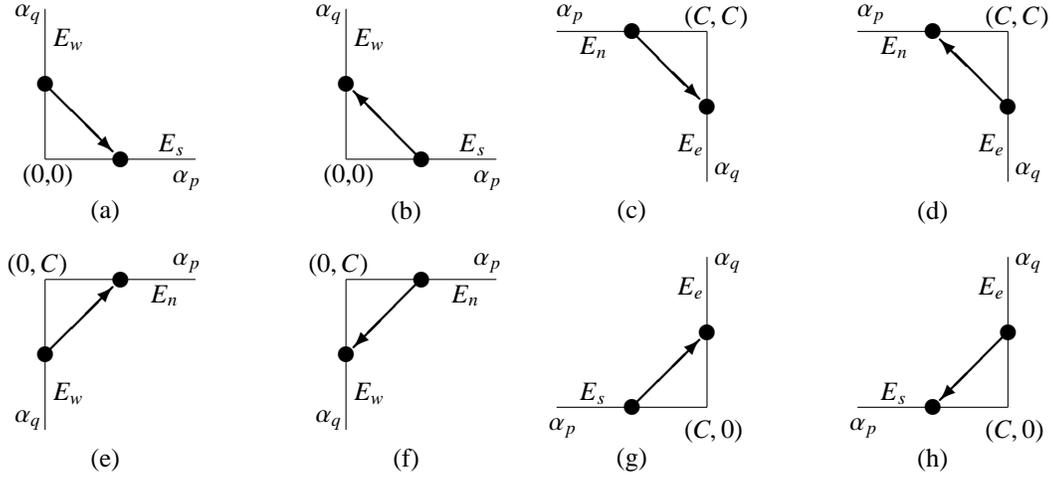


Fig. 1. Possible transitions from $(\alpha_p(k), \alpha_q(k))$ to $(\alpha_p(k+1), \alpha_q(k+1))$ for sufficiently large $k \in L(p, q)$: (a) $y_p = y_q = 1$, $(\alpha_p^*, \alpha_q^*) = (0, 0)$, (b) $y_p = y_q = -1$, $(\alpha_p^*, \alpha_q^*) = (0, 0)$, (c) $y_p = y_q = 1$, $(\alpha_p^*, \alpha_q^*) = (C, C)$, (d) $y_p = y_q = -1$, $(\alpha_p^*, \alpha_q^*) = (C, C)$, (e) $y_p = 1$, $y_q = -1$, $(\alpha_p^*, \alpha_q^*) = (0, C)$, (f) $y_p = -1$, $y_q = 1$, $(\alpha_p^*, \alpha_q^*) = (0, C)$, (g) $y_p = 1$, $y_q = -1$, $(\alpha_p^*, \alpha_q^*) = (C, 0)$, (h) $y_p = -1$, $y_q = 1$, $(\alpha_p^*, \alpha_q^*) = (C, 0)$.

On the other hand, it follows from Lemma 3 that (q, p) cannot belong to I_∞ . This leads to a contradiction. Therefore, SMO algorithm stops in a finite number of iterations.

The above proof is not correct because the point $(\alpha_p(k), \alpha_q(k))$ may be able to visit two edges of S alternatively even though one of (p, q) and (q, p) does not belong to I_∞ . Let us consider for example the case where $(p, q) \in I_\infty$, $y_p = y_q = 1$ and $(\alpha_p^*, \alpha_q^*) = (0, 0)$. Figure 2 shows a possible way that $(\alpha_p(k), \alpha_q(k))$ returns from E_s to E_w . First $(\alpha_p(k), \alpha_q(k))$ moves from E_w to E_s when k increases from k_1 to $k_1 + 1$. Second $(\alpha_s(k), \alpha_q(k))$ ($s \neq p$) moves from E_s to E_e when k increases from k_2 to $k_2 + 1$ ($k_2 > k_1$). Finally $(\alpha_p(k), \alpha_r(k))$ ($r \neq q$) moves from E_s to E_w when k increases from k_3 to $k_3 + 1$ ($k_3 > k_2$). Then the point $(\alpha_p(k), \alpha_q(k))$ is on the each E_w for $k = k_3 + 1$. More strictly speaking, the point $(\alpha_p(k), \alpha_q(k))$ may return from E_s to E_w infinitely many times if there exist an $r (\neq q)$ such that $(\alpha_p(k), \alpha_r(k)) \in E_n \cup E_s$ and $(\alpha_p(k+1), \alpha_r(k+1)) \in E_w$ hold for infinitely many k , and an $s (\neq p)$ such that $(\alpha_s(k), \alpha_q(k)) \in E_s$ and $(\alpha_s(k+1), \alpha_q(k+1)) \in E_e \cup E_w$ hold for infinitely many k . Furthermore it is obvious that the point $(\alpha_p(k), \alpha_q(k))$ cannot return from E_s to E_w infinitely many times unless this condition is satisfied.

By exploring this situation in detail for all possible cases, we can derive the following lemma.

Lemma 5: If $(p, q) \in I_\infty$ then there must be an $r (\neq q)$ such that $(r, p) \in I_\infty$, and an $s (\neq p)$ such that $(q, s) \in I_\infty$.

Proof: Let us consider the case where $y_p = y_q = 1$ and $\alpha_p^* = \alpha_q^* = 0$. In this case, it follows from Lemma 4 that the point $(\alpha_p(k), \alpha_q(k))$ moves from E_w to E_s infinitely many times. In order for this to occur the point $(\alpha_p(k), \alpha_q(k))$ must return from E_s to E_w infinitely many times, that is, 1) $0 < \alpha_p(k) < C$ and $\alpha_p(k+1) = 0$ for infinitely many k , and 2) $\alpha_q(k) = 0$ and $0 < \alpha_q(k+1) < C$ for infinitely many k . Since $(q, p) \notin I_\infty$ from Lemma 3, the condition 1) is satisfied only if there exists an $r (\neq p, q)$ such that either i) $(r, p) \in I_\infty$, $y_r = 1$, $\alpha_r^* = 0$ or ii) $(r, p) \in I_\infty$, $y_r = -1$, $\alpha_r^* = C$. This is easily verified from Fig.1. Similarly, the condition 2) is satisfied only if there exists an $s (\neq p, q)$ such that either i)

TABLE I
POSSIBLE VALUES OF (y_r, α_r^*) AND (y_s, α_s^*) .

Case	(y_r, α_r^*)	(y_s, α_s^*)
(a)	$(1, 0)$ or $(-1, C)$	$(1, 0)$ or $(-1, C)$
(b)	$(1, C)$ or $(-1, 0)$	$(1, C)$ or $(-1, 0)$
(c)	$(1, C)$ or $(-1, 0)$	$(1, C)$ or $(-1, 0)$
(d)	$(1, 0)$ or $(-1, C)$	$(1, 0)$ or $(-1, C)$
(e)	$(1, 0)$ or $(-1, C)$	$(1, 0)$ or $(-1, C)$
(f)	$(1, C)$ or $(-1, 0)$	$(1, C)$ or $(-1, 0)$
(g)	$(1, C)$ or $(-1, 0)$	$(1, C)$ or $(-1, 0)$
(h)	$(1, 0)$ or $(-1, C)$	$(1, 0)$ or $(-1, C)$

$(q, s) \in I_\infty$, $y_s = 1$, $\alpha_s^* = 0$ or ii) $(q, s) \in I_\infty$, $y_s = -1$, $\alpha_s^* = C$. Therefore the statement holds true. For each of the remaining seven cases, we can easily show in a similar way that for (p, q) to belong to I_∞ there must be an $r (\neq p, q)$ and an $s (\neq p, q)$ such that $(r, p) \in I_\infty$ and $(q, s) \in I_\infty$. Possible values of (y_r, α_r^*) and (y_s, α_s^*) are summarized in Table I. ■

We are now ready for giving a complete proof.

Theorem 1: SMO algorithm stops in a finite number of iterations for any $\tau > 0$.

Proof: The proof will be done by contradiction. Assume that SMO algorithm does not stop. Then I_∞ is not empty. Let (p, q) be any pair in I_∞ . It follows from Lemma 5 that there must be an $r (\neq q)$ such that $(r, p) \in I_\infty$ and an $s (\neq p)$ such that $(q, s) \in I_\infty$. Since $(i, j) \in I_\infty$ implies

$$F_i(\alpha^*) \leq F_j(\alpha^*) - \tau,$$

we have

$$F_r(\alpha^*) < F_p(\alpha^*) < F_q(\alpha^*) < F_s(\alpha^*)$$

which means that the indices p, q, r and s are different from each other. Next, by applying the above argument to the conditions $(r, p) \in I_\infty$ and $(q, s) \in I_\infty$, we see that there must be a $t (\neq p)$ such that $(t, r) \in I_\infty$ and a $u (\neq q)$ such that $(s, u) \in I_\infty$. Then we have

$$F_t(\alpha^*) < F_r(\alpha^*) \quad \text{and} \quad F_s(\alpha^*) < F_u(\alpha^*)$$

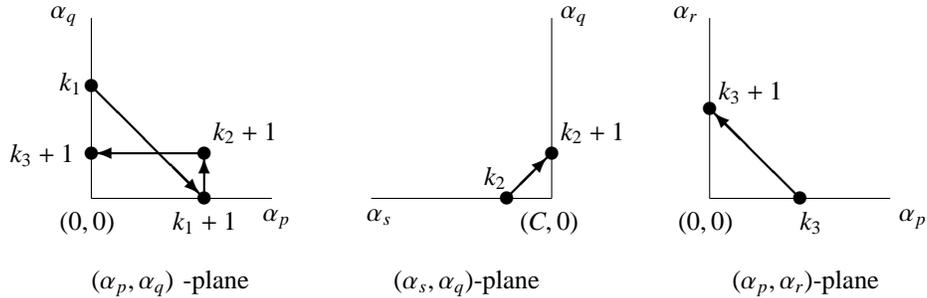


Fig. 2. An example of the way $(\alpha_p(k), \alpha_q(k))$ returns from E_s to E_w .

which mean that the indices p, q, r, s, t and u are different from each other. Since this process can be repeated infinitely many times, we reach the conclusion that if $(p, q) \in I_\infty$ then I_∞ must contain infinitely many pairs. However, this is impossible because the number of pairs cannot exceed $l(l-1)$. Therefore, SMO algorithm necessarily stops in a finite number of iterations. ■

IV. CONCLUSION

Convergence property of SMO algorithm was studied in this paper. We have given a complete proof that SMO algorithm always stops in a finite number of iterations. Although only QP problems in the form of Problem 1 were dealt with in this paper, it is straightforward to extend our result to more general QP problems like that considered in [5]. Convergence analysis of other decomposition methods for SVM learning is a future problem.

REFERENCES

- [1] V. N. Vapnik, *Statistical Learning Theory*. New York: John Wiley & Sons, 1998.
- [2] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty, *Nonlinear Programming: Theory and Algorithm*. New York: John Wiley & Sons, 1993.
- [3] T. Joachims, "Making large-scale support vector machine learning practical," in *Advances in Kernel Methods: Support Vector Machines*, B. Schölkopf, C. Burges, and A. Smola, Eds. Cambridge, MA: MIT Press, 1998.
- [4] J. C. Platt, "Fast training of support vector machines using sequential minimal optimization," in *Advances in Kernel Methods: Support Vector Machines*, B. Schölkopf, C. Burges, and A. Smola, Eds. Cambridge, MA: MIT Press, 1998.
- [5] S. S. Keerthi and E. G. Gilbert, "Convergence of a generalized SMO algorithm for SVM classifier design," *Machine Learning*, vol. 46, pp. 351–360, 2002.
- [6] C.-J. Lin, "On the convergence of the decomposition method for support vector machines," *IEEE Transactions on Neural Networks*, vol. 12, pp. 1288–1298, Nov. 2001.
- [7] —, "Asymptotic convergence of an SMO algorithm without any assumption," *IEEE Transactions on Neural Networks*, vol. 13, pp. 248–250, Jan. 2002.
- [8] —, "A formal analysis of stopping criteria of decomposition methods for support vector machines," *IEEE Transactions on Neural Networks*, vol. 13, pp. 1045–1052, Sept. 2002.
- [9] S. S. Keerthi, S. K. Shevade, C. Bhattacharyya, and K. R. K. Murthy, "Improvements to Platt's SMO algorithm for SVM classifier design," *Neural Computation*, vol. 13, pp. 637–649, 2001.