Optimization of an SVM QP Problem Using Mixed Variable Nonlinear Polynomial Kernel Map and Mixed Variable Unimodal Functions

Emre Tokgöz[1,2] and Theodore B. Trafalis[1]

[1]School of Industrial Engineering, University of Oklahoma, Norman, OK, 73019, U.S.A.
[2]Department of Mathematics, University of Oklahoma, Norman, OK, 73019, U.S.A.

Emre.Tokgoz-1@ou.edu and ttrafalis@ou.edu

Abstract: - Support Vector Machines (SVM) can be constructed with the selection of an appropriate kernel function to solve an optimization problem. Algorithmic approaches can be taken to solve problems related to SVM which are used for regression analysis and data classification of a data set. The (inhomogeneous) polynomial kernel $k(x,y) = \left(1 + x^T \cdot y\right)^d$ is useful for non-linear data set classification. In this work, the SVM QP problem with (inhomogeneous) polynomial kernel $k(x,y)$ is expressed as a mixed convex optimization problem with respect to the real variables $\alpha \in \mathbb{R}^l$ and $b \in \mathbb{R}$, and the integer variable $d \in \mathbb{Z}$. Several examples of mixed convexity and computational results are given. In addition, we introduce the definitions of unimodal and semi-unimodal mixed variable functions with the corresponding minimization results.

Key-Words: - SVM, Inhomogeneous Polynomial Kernel Map, Unimodal Function, Mixed Convex Function, Minimization.

1. Introduction

Support Vector Machines have important applications in practical optimization problems in several different research areas. Some of these areas include fault detection of motor failure in industrial motors within the technological improvement of artificial neural networks ([3]), the modelling of the multi-variable systems on the reproduction of the kernel Hilbert Space ([7]), and the control aspects of man-machine systems with brain machine interface where a kernel based approach is taken for the appropriate non-linear SVM ([4]).

A data set (x_i, l_i), $i = 1, 2, ..., l$, with $l_i = \pm 1$ is given with $x_i \in \mathbb{R}^n$ in the input space. This can be mapped to the feature space by a map Φ to obtain a separating hyperplane for two subsets of the given data set (positive and negative subsets.) The classification of the data sets is possible through the maximum-margin hyperplane in the feature space by using a map $k : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$

$$k(x,y) = \langle \Phi(x), \Phi(y) \rangle,$$

called the kernel map ([11]). In particular, nonlinear classification of the given data sets can be obtained by using the map

$$k(x,y) = \left(1 + x^T \cdot y\right)^d \quad (1)$$

called the (inhomogeneous) polynomial kernel map where $x, y \in \mathbb{R}^n$ are the input data vectors ([11]). Note that the hyperplane in the feature space has the form

$$w^T \Phi(x) + b = 0.$$

The SVM optimization problem can be expressed as

$$\min_{w\in\mathbb{R}^l, b\in\mathbb{R}} \frac{\|w\|^2}{2} \quad (2)$$

subject to

$$\begin{aligned} w^T\cdot\Phi(x_i)+b &\geq 0 \text{ for } x_i \text{ with } l_i=+1, \quad (3)\\ w^T\cdot\Phi(x_i)+b &\leq 0 \text{ for } x_i \text{ with } l_i=-1. \quad (4)\end{aligned}$$

Note that w can be written as

$$w=\sum_{i=1}^{l}\alpha_i\Phi(x_i),$$

indicating

$$\begin{aligned}\|w\|^2 &= \langle w,w\rangle\\ &= \left\langle \sum_{i=1}^{l}\alpha_i\Phi(x_i), \sum_{j=1}^{l}\alpha_j\Phi(x_j)\right\rangle\\ &= \alpha^T K\alpha,\end{aligned}$$

where $\alpha=(\alpha_1,\alpha_2,...,\alpha_l)\in\mathbb{R}^l$ and

$$k(x_i,x_j)=k_{ij}=\langle\Phi(x_i),\Phi(x_j)\rangle$$

forms the matrix K_d. Hence the SVM quadratic programming problem proposed in $(2)-(4)$ takes the form

$$\begin{aligned} &\min_{\alpha,d,b}\frac{1}{2}\alpha^T K_d\alpha\\ &\textit{subject to}\\ \left(k_i^+\right)^T\cdot\alpha &\leq -b,\ 1\leq i\leq m,\\ -\left(k_i^-\right)^T\cdot\alpha &\leq b,\ m+1\leq i\leq l \quad (5)\end{aligned}$$

where k_i^+ denotes the column of K_d referring to the points that lie above the hyperplane and k_i^- denotes the column of K_d referring to the points that lie below the hyperplane. This SVM QP problem is a well known minimization problem in the literature when d is considered to be a constant ([11]). By including $d\in\mathbb{Z}$ to this SVM QP problem as an integer variable our goal is to express (5) as a mixed convex minimization problem. We first change the constrained minimization problem given in (5) to the unconstrained minimization problem

$$\begin{aligned} f &: \mathbb{R}^l\times\mathbb{R}\times\mathbb{Z}\to\mathbb{R}\\ (\alpha,b,d) &\to f(\alpha,b,d)=\min_{\alpha,d,b}\frac{1}{2}\alpha^T K_d\alpha\\ &\quad+\sum_{i=1}^{l}\Big((k_i(d))^T\alpha+b\Big)^2, \quad (6)\end{aligned}$$

where $\alpha=(\alpha_1,\alpha_2,...,\alpha_l)\in\mathbb{R}^l$ and $b\in\mathbb{R}$ are the real variables, and $d\in\mathbb{Z}$ is the integer variable of the proposed problem in (6). We consider K as a function of d and express it as K_d.

2. Mixed Convexity

The convexity conditions of functions with integer and real (mixed) variables have been obtained by many researchers by either fixing the integer or the real variable. Examples of such results can be found in [4] for the Erlang delay and loss formulas in telecommunication network queueing systems and in [2] for an inventory model with time limit on backorders. The mixed convexity results obtained in [9] are obtained without fixing the integer or the real variable where a mixed function, a function with integer and real variables, $\Psi:\mathbb{Z}^n\times\mathbb{R}^m\to\mathbb{R}$ is defined to be 2-smooth strict mixed convex if it is obtained from a C^2 real variable strictly convex function $\overline{\Psi}:\mathbb{R}^{n+m}\to\mathbb{R}$ satisfying $\overline{\Psi}(d,h)=\Psi(d,h)$ for $\forall(d,h)\in\mathbb{Z}^n\times\mathbb{R}^m$. The mixed Hessian matrix corresponding to a 2-smooth strict mixed convex function Ψ is defined by

$$H_\Psi=\begin{bmatrix}\left[\nabla_{ij}\Psi(d,h)\right]_{n\times n} & \left[\nabla_i(\frac{\partial\Psi(d,h)}{\partial h_k}\right]_{n\times m}\\ \left[\frac{\partial}{\partial h_k}(\nabla_j\Psi(d,h)\right]_{m\times n} & \left[\frac{\partial^2\Psi(d,h)}{\partial h_k\partial h_t}\right]_{m\times m}\end{bmatrix}, \quad (7)$$

where $d\in\mathbb{Z}^n, h\in\mathbb{R}^m, 1\leq i,\ j\leq n$ and $1\leq k$, $t\leq m$. In H_Ψ,

$$\nabla_i(\Psi(d,h))=\Psi(d+e_i,h)-\Psi(d,h)$$

denotes the first difference and

$$\nabla_{ij}(\Psi(d,h)) = \Psi(d+e_i+e_j,h) - \Psi(d+e_j,h) - \Psi(d+e_i,h) + \Psi(d,h) \quad (8)$$

denotes the second difference of Ψ with respect to the discrete variable d where e_i is the unit integer vector at the i^{th} position while $\frac{\partial}{\partial h}$ and $\frac{\partial^2}{\partial h^2}$ denote the first and second partial derivatives of Ψ with respect to the real variable h, respectively. In addition, by applying the following theorem obtained in [9], the 2-smooth strict mixed convexity of a mixed function can be obtained.

Theorem 2.1: A function $\Psi : \mathbb{Z}^n \times \mathbb{R}^m \to \mathbb{R}$ is 2-smooth strict mixed convex if and only if the mixed Hessian matrix corresponding to Ψ is positive definite.

An example of a 2-smooth strict mixed convex function is given in [9] as follows: Consider an automated machine that can perform up to s operations in series where each step takes the same mean time, with the times distributed exponentially. If the arrival process is Poisson, we have an $M/E_k/1$ queueing system. Let A_1 be a linear cost associated with each operation, A_2 be a linear cost associated with each unit of service, and A_3 a linear cost associated with the length of the queue, L_q. Then the design problem is to minimize $f : \mathbb{Z} \times \mathbb{R} \to \mathbb{R}$ such that

$$f(k,\mu) = A_1 s + A_2\mu + A_3(\frac{\lambda}{\mu})^2 \frac{(s+1)}{2s(1-\frac{\lambda}{\mu})}$$

subject to $s \in \mathbb{Z}^+$ and $\rho = \frac{\lambda}{\mu}$ on the neighborhood $U = \{\mu \mid \mu > \lambda > 0\}$. The function f is shown to be a 2-smooth strict mixed convex function by using theorem 2.1 in [9]. Note that the determinant of the mixed Hessian matrix corresponding to f,

$$\begin{aligned} \det(H_f) &= \frac{(C_3)^2\lambda^4((2+k(7+4k))\lambda^2}{4k^2(k+1)^2(k+2)(\lambda-\mu)^4\mu^4} \\ &+ \frac{4(1+k(5+3k))\mu^2}{4k^2(k+1)^2(k+2)(\lambda-\mu)^4\mu^4} \\ &- \frac{4(1+k(5+3k))\lambda\mu)}{4k^2(k+1)^2(k+2)(\lambda-\mu)^4\mu^4}, \end{aligned}$$

is strictly positive with respect to the specified constraints.

When $d \in \mathbb{Z}^+$ is assumed to be a constant, the solution to the minimization problem $\min \frac{1}{2}\|w\|^2$ is unique with respect to the real variables $\alpha \in \mathbb{R}^l$ and $b \in \mathbb{R}$ if the Hessian matrix K is positive definite where $\|w\|^2$ is a C^2 strict real convex function. In this paper we obtain theoretical mixed convexity and minimization results to solve the minimization problems (5) and (6) by considering the real and the integer variables simultaneously. The convexity and optimization results obtained in this paper are of closed form and the mixed convexity regions obtained are obtained with respect to both integer and real variables. We suggest an algorithmic approach to determine the mixed convexity conditions of a mixed function because of the complex nature of the calculations that arise while finding the positive definiteness conditions of the corresponding mixed Hessian matrix.

We first obtain theoretical mixed convexity results to solve the problems stated in (5) and (6).

Let

$$\begin{aligned} q_{ij} &= x_i^T \cdot x_j && (9) \\ k(d) &= [k_i(d)] && (10) \\ k_{ij}(d) &= (1+q_{ij})^d && (11) \\ k_{ij}^T(d) &= \left(1+q_{ij}^T\right)^d && (12) \\ p(d,\alpha) &= \frac{1}{2}\alpha^T K_d \alpha && (13) \\ h(\alpha,b,d) &= \sum_{i=1}^{l}\left((k_i(d))^T \cdot \alpha + b\right)^2 && (14) \end{aligned}$$

The first difference of $p(d,\alpha)$ given in (13) with respect to the integer variable d is

$$\nabla_1 p(d,\alpha) = \frac{1}{2}\alpha^T[K_{d+1} - K_d]\alpha = \frac{1}{2}\alpha^T A\alpha,$$

where A is an $l \times l$ matrix with the elements of the form

$$A_{ij}(d) = (1+q_{ij})^d q_{ij} = k_{ij}q_{ij}. \qquad (15)$$

Let

$$A_{ij}^T(d) = \left(1+q_{ij}^T\right)^d q_{ij}^T = k_{ij}^T q_{ij}^T. \qquad (16)$$

The first difference of $h(\alpha,b,d)$ is

$$\begin{aligned}\nabla_1 h(\alpha,b,d) &= \nabla_1\left(\sum_{i=1}^{l}\left(k_i^T(d)\alpha+b\right)^2\right)\\ &= \sum_{i=1}^{l}[(k_i^T(d+1)\alpha+b)^2\\ &\quad -(k_i^T(d)\alpha+b)^2]\\ &= B(\alpha,b,d), \qquad (17)\end{aligned}$$

where

$$B(\alpha,b,d) = \sum_{i,j=1}^{l} A_{ij}^T(d)\alpha_j(\left[2k_{ij}^T(d) + A_{ij}^T(d)\right]\alpha_i + 2b). \qquad (18)$$

The second difference of $p(\alpha,b,d)$ with respect to the integer variable d is

$$\begin{aligned}\nabla_{11}p(\alpha,d) &= \frac{1}{2}\alpha^T\left[k(d+2) - 2k(d+1) + k(d)\right]\alpha\\ &= \frac{1}{2}\alpha^T C\alpha, \qquad (19)\end{aligned}$$

where C is an $l \times l$ matrix with the components

$$C_{ij}(d) = \left[(1+q_{ij})^d (q_{ij})^2\right]. \qquad (20)$$

The second difference of h with respect to d is the first difference of $B(\alpha,b,d)$. Therefore

$$\begin{aligned}\nabla_{11}h(\alpha,b,d) &= \nabla_1 B(\alpha,b,d)\\ &= \sum_{i,j=1}^{l}\nabla_1(C_{ij}^2(d)\\ &\quad +2\alpha_j C_{ij}(d)k_{ij}^T(d)\\ &\quad +2C_{ij}(d)b)\\ &= \sum_{i,j=1}^{l}\{(A_{ij}^T(d+1)\alpha_j)\cdot\\ &\quad ([2k_{ij}^T(d+1)\\ &\quad +A_{ij}^T(d+1)]\alpha_i+2b)\\ &\quad -(A_{ij}^T(d)\alpha_j)([2k_{ij}^T(d)\\ &\quad +A_{ij}^T(d)]\alpha_i+2b)\}.\end{aligned}$$

Let $\partial_\alpha := \left(\frac{\partial}{\partial\alpha_1},\frac{\partial}{\partial\alpha_2},...,\frac{\partial}{\partial\alpha_l}\right)$ denotes the gradient operator with respect to $\alpha \in \mathbb{R}^l$. By taking the gradient of $\nabla_1 f$ with respect to α we have the vector

$$\begin{aligned}\partial_\alpha\nabla_1 f(\alpha,b,d) &= A\alpha + \partial_\alpha\{\sum_{i=1}^{l}[(k_i^T(d+1)\\ &\quad -k_i^T(d))\alpha][(k_i^T(d+1)\\ &\quad +k_i^T(d))\alpha+2b)\}\\ &= A\alpha + \sum_{i=1}^{l}\{(k_i^T(d+1)\\ &\quad -k_i^T(d))^T[(k_i^T(d+1)\\ &\quad +k_i^T(d))\alpha+2b]\}\\ &\quad +\sum_{i=1}^{l}[(k_i^T(d+1)-k_i^T(d))]\cdot\\ &\quad \alpha(k_i^T(d+1)+k_i^T(d))^T\\ &\quad +k_i^T(d))^T.\\ &= A\alpha + \sum_{i=1}^{l}\left(\sum_{j=1}^{l}A_{ij}^T(d)\alpha_j\right)\cdot\\ &\quad [(k_i^T(d+1)+k_i^T(d))\alpha+2b]\\ &\quad +\sum_{i=1}^{l}\left(\sum_{j=1}^{l}A_{ij}^T(d)\alpha_j\right)(k_i^T(d+1).\end{aligned}$$

Rewriting $f(\alpha, b, d)$ we have

$$
\begin{aligned}
f(\alpha, b, d) &= \frac{1}{2}\alpha^T k(d)\alpha + \sum_{i=1}^{l}(k_i(d)\alpha + b)^2 \\
&= \frac{1}{2}\alpha^T k(d)\alpha + \sum_{i=1}^{l}[\alpha^T k_i^T(d) k_i(d)\alpha \\
&\quad +2bk_i(d)\alpha + b^2].
\end{aligned}
$$

Hence

$$
\begin{aligned}
\partial_\alpha f(\alpha, b, d) &= k(d)\alpha + 2\sum_{i=1}^{l}[\left(k_i^T(d) k_i(d)\right)\alpha \\
&\quad +2bk_i^T(d)] \\
\partial_\alpha \partial_\alpha f(\alpha, b, d) &= k^T(d) + 2\sum_{i=1}^{l} k_i^T(d) k_i(d).
\end{aligned}
$$

The elements of the mixed Hessian matrix H_f with respect to the real variable b are

$$\frac{\partial}{\partial b}\nabla_1 f(\alpha, b, d) = 2\sum_{i,j=1}^{l}\left(A_{ij}^T(d)\alpha_j\right) \quad (21)$$

$$\frac{\partial^2 f(\alpha, b, d)}{\partial b^2} = 2 \quad (22)$$

$$\frac{\partial}{\partial b}\partial_\alpha f(\alpha, b, d) = 2k_i^T(d). \quad (23)$$

By using (7), the mixed Hessian matrix $H_f(\alpha, b, d)$ corresponding to $f(\alpha, b, d)$ is

$$H_f(\alpha, b, d) = \begin{bmatrix} \nabla_{11} f & \frac{\partial}{\partial b}\nabla_1 f & [\partial_\alpha \nabla_1 f] \\ \nabla_1 \frac{\partial}{\partial b} f & \frac{\partial^2 f}{\partial b^2} & \left[\partial_\alpha \frac{\partial}{\partial b} f\right] \\ [\nabla_1 \partial_\alpha f] & \left[\frac{\partial}{\partial b}\partial_\alpha\right] & [\partial_\alpha \partial_\alpha f] \end{bmatrix}$$

by the symmetry property of the mixed Hessian matrix. By theorem 2.1, $f(\alpha, b, d)$ is strict mixed convex if and only if the mixed Hessian matrix H_f is positive definite. However it is a complicated task to compute the positive definiteness of H_f for large values of the dimension l. Therefore we provide the following algorithm for symbolic calculations.

2.1 An Algorithm

The next algorithm can be useful to observe the strict mixed convexity conditions of $f(\alpha, b, d)$ with respect to b, d, and α for large values of l and n. The proposed algorithm can be used to find the region where H_f is positive definite by symbolic programming.

An Algorithm:

Introduce dimensions l and n
Define the numerical input data x, and y
Introduce the symbols α, d, and b
Calculate the upper triangular elements of the matrix $H_f(l+2, l+2)$.
Use symmetry property of the mixed Hessian matrix to define the lower triangular portion of H_f
// Checking the positive definiteness of the mixed Hessian Matrix.
Calculate the principal minor matrix determinants of H_f. Call them h_f.
for all h_f
 Calculate symbolically $det(h_f)$
 If $det(h_f) > 0$
  { Call this region D_f^1 in the domain,
  print "f is strict mixed convex in D_f^1"}
else if $det(h_f) < 0$
  { Call this region D_f^2 in the domain,
  print "f is not strict mixed convex in D_f^2"}
else if $det(h_f) = 0$
  { Call this region D_f^3 in the domain,
  print "f is not strict mixed convex in D_f^3"}
end

3. Optimization Results

In this section we introduce new theoretical global optimization results for a set of unconstrained nonlinear mixed optimization problems after introducing increasing and decreasing mixed function concepts. The theoretical results obtained in this section are for unconstrained nonlinear mixed optimization problems.

Let

$$\mathbb{Z}_{[a,\infty)}^n = \prod_{i=1}^{n}[a_i, \infty) \subset \mathbb{Z}^n$$

and

$$\mathbb{R}^m_{[b,\infty]} = \prod_{j=1}^{m} [b_j, \infty) \subset \mathbb{R}^m$$

where $a = (a_i)_{i=1}^n \in \mathbb{Z}^n$ and $b = (b_j)_{j=1}^m \in \mathbb{R}^m$. We assume all the functions considered in this section are C^1 unless stated otherwise.

Definition 3.1: A mixed function $f : D \to \mathbb{R}$ is strictly increasing on $D \subseteq \mathbb{Z}^n \times \mathbb{R}^m$ if $\nabla_i f(x, y) > 0$ and $\frac{\partial f(x,y)}{\partial y_j} > 0$ hold for all $1 \le i \le n$, $1 \le j \le m$ and $(x, y) \in D$. A mixed function $g : D \to \mathbb{R}$ is strictly decreasing in D if $-g$ is strictly increasing in D.

Let

$$D_1 = \mathbb{Z}^n_{[a,\infty)} \times \mathbb{R}^m_{[b,\infty)}$$

and

$$D_2 = \mathbb{Z}^n_{(-\infty,c]} \times \mathbb{R}^m_{(-\infty.d]}.$$

Theorem 3.1: If $f : D_1 \to \mathbb{R}$ is a strictly increasing mixed convex function in D_1 then the unconstrained mixed optimization problem $\min\limits_{(x,y)\in D_1} f(x, y)$ has a unique solution attained at (a, b).

Proof: Let $f : D_1 \to \mathbb{R}$ be a strictly increasing mixed convex function in D_1. Therefore, by definition 3.1 f satisfies the inequalities

$$\nabla_i f(x, y) > 0 \Rightarrow f(x + e_i, y) > f(x, y),$$

for all $1 \le i \le n$ and all $(x, y) \in D_1$, and

$$\frac{\partial f(x, y)}{\partial y_j} > 0,$$

for all $1 \le j \le m$ and all $(x, y) \in D_1$. Suppose on the contrary to the assumption that there does not exist a unique global minimum point of f. Hence assume the existence of the global minimum points $(x_1, y_1) \in D_1$ and $(x_2, y_2) \in D_1$ of $f(x, y)$ in the domain D_1. Therefore the point $(x_1, y_1) \in D_1$ satisfies

$$f(x + e_i, y) > f(x_1, y_1), \tag{24}$$

for all $1 \le i \le n$ and all $(x, y) \in D_1$. Considering the point $(x_2, y_2) \in D_1$,

$$f(x + e_k, y) > f(x_2, y_2) \tag{25}$$

holds for all $k = 1, ..., n$ and all $(x, y) \in D_1$. By (24) and (25), the inequalities $f(x_1, y_1) > f(x_2, y_2)$ and $f(x_2, y_2) > f(x_1, y_1)$ hold. This gives a contradiction to the strictly increasing mixed convexity of f. This indicates that the minimal value is attained when $(x, y) = (a, b)$ since it is the boundary value of a strictly increasing mixed convex function.

Corollary 3.1: If $g : D_2 \to \mathbb{R}$ is a strictly decreasing mixed convex function in D_2 then the unconstrained mixed optimization problem $\max g(x, y)$ for $(x, y) \in D_2$ has a unique solution attained at (c, d).

Proof: The proof follows from theorem 3.1 where we have $f = -g$ in D_2.

Corollary 3.2: If $f : D_1 \to \mathbb{R}$ is a 2-smooth mixed function with positive definite mixed Hessian matrix for all $(x, y) \in D_1$, and f is a strictly increasing mixed function in D_1 then f has a unique global minimum point in D_1 which is attained when $(x, y) = (a, b)$.

Proof: We have $f : D_1 \to \mathbb{R}$, $D_1 \subseteq \mathbb{Z}^n \times \mathbb{R}^m$, as a 2-smooth mixed function with positive definite mixed Hessian matrix for all $(x, y) \in D_1$ therefore it is a strict mixed convex function by theorem 2.1. By the assumption f is a strictly increasing mixed convex function. Therefore, by theorem 3.1 f has a unique global minimum point.

The proof of the following corollary follows similar to the proof of corollary 3.2.

Corollary 3.3: If $g : D_2 \to \mathbb{R}$ is a 2-smooth mixed function with negative definite mixed Hessian matrix for all $(x, y) \in D_2$, and g is a strictly decreasing mixed function in D_2 then g has a unique global maximum point in D_2 which is attained when $(x, y) = (c, d)$.

3.1 Mixed Convexity Examples

We first consider the minimization problem given in (5) without constraints. Considering equation (13), our goal is to solve $\min_{\alpha,d} p(d,\alpha)$. As a function of $\alpha \in \mathbb{R}^l$ when $d \in \mathbb{Z}^+$ is a fixed constant, it is well known that $p(\alpha)$ is a strict real convex function when the corresponding mixed Hessian matrix $H = \left[\left(1 + x_i^T \cdot x_j\right)^d\right]_{l \times l}$ is positive definite. Therefore $p(\alpha)$ has a unique minimum when H_p is positive definite in $\mathbb{R}^l$. When $d \in \mathbb{Z}^+$ is a variable we have

$$\begin{aligned} p \quad &: \quad \mathbb{R}^l \times \mathbb{Z}^+ \to \mathbb{R} \\ (\alpha, d) \quad &\mapsto \quad \frac{1}{2}\alpha^T \left[(1 + q_{ij})^d\right] \alpha \end{aligned}$$

with the first difference

$$\begin{aligned} \nabla_1 p(\alpha, d) &= \nabla_1 \left(\frac{1}{2}\alpha^T \left[(1 + q_{ij})^d\right] \alpha\right) \\ &= \frac{1}{2}\alpha^T A \alpha. \end{aligned}$$

$\nabla_1 p > 0$ holds in particular when $q_{ij} > 0$ and $(\alpha, d) \in (\mathbb{R}^+)^l \times \mathbb{Z}^+$ where A has elements of the form (15). The gradient vector of p is $\partial_\alpha p = \left(1 + x^T \cdot y\right)^d \alpha > 0$ for all $(\alpha, d) \in (\mathbb{R}^+)^l \times \mathbb{Z}^+$. Therefore by definition 3.1, $p(\alpha, d)$ is an increasing function in particular in $D = (\mathbb{R}^+)^l \times \mathbb{Z}^+$ if $q_{ij} > 0$. Now we consider the mixed Hessian matrix of p,

$$H_p = \begin{bmatrix} \frac{1}{2}\alpha^T C \alpha & \left[(1 + q_{ij})^d q_{ij}\right]^T \alpha \\ \left[(1 + q_{ij})^d q_{ij}\right]^T \alpha & [k(d)] \end{bmatrix},$$

where the components of H_p are calculated in the previous section. Note that H_p is positive definite when $\frac{1}{2}\alpha^T C \alpha > 0$ and $\det(h_p) > 0$ for all h_p. We let $D \subseteq (\mathbb{R}^+)^l \times \mathbb{Z}^+$ to be the region of strict mixed convexity of p which is obtained by positive definite mixed Hessian matrix H_p. Noting that $p(\alpha, d)$ is a 2-smooth mixed function, corollary 3.2 indicates the existence of a unique minimum value of p in the region D when C is positive definite and all the determinants of the principal minor matrices (i.e. $\det(h_p)$ stated in the algorithm above) are positive definite. In the case when $\frac{1}{2}\alpha^T C \alpha > 0$ and $\det(h_p)$ are positive definite for all $(\alpha, d) \in (\mathbb{R}^+)^l \times \mathbb{Z}^+$, the minimum value of $p(\alpha, d)$ is obtained for $d = 2$ in the nonlinear case.

Now we consider $\min_{\alpha,b,d} f(\alpha, b, d)$ given in (6) with constraints. We first observe whether the first difference and the gradient vector of f are positive or not. It is easy to see from the previous section that the first difference of f is

$$\nabla_1 f(\alpha, b, d) = B(\alpha, b, d) + \frac{1}{2}\alpha^T A \alpha$$

for all $(\alpha, b, d) \in \mathbb{R}^{l+1} \times \mathbb{Z}$. The gradient of f is

$$\begin{aligned} \partial_\alpha f(\alpha, b, d) &= k(d)\alpha + 2\sum_{i=1}^{l} [\left(k_i^T(d) k_i(d)\right)\alpha \\ &\quad + 2b k_i^T(d)] \\ \frac{\partial f}{\partial b} &= \sum_{i=1}^{l} \left(2k_i^T(d) \cdot \alpha + 2b\right) \end{aligned}$$

for all $(\alpha, b, d) \in \mathbb{R}^{l+1} \times \mathbb{Z}^+$. Considering $\nabla_1 f$, $\partial_\alpha f$ and $\frac{\partial f}{\partial b}$ stated above, it is not an easy task to determine whether f is a strictly increasing mixed function or not. If we consider $(\alpha, b, d) \in (\mathbb{R}^+)^l \times \mathbb{R}^+ \times \mathbb{Z}^+$, it is easy to see that f is a strictly increasing mixed function when $q_{ij} > 0$. If the mixed Hessian matrix H_f corresponding to f is positive definite in $D = (\mathbb{R}^+)^l \times \mathbb{R}^+ \times \mathbb{Z}^+$ and since f is an increasing 2-smooth mixed function in D, corollary 3.2 indicates the existence of a unique solution of $\min_{\alpha,b,d} f(\alpha, b, d)$ in $(\mathbb{R}^+)^l \times \mathbb{R}^+ \times \mathbb{Z}^+$. In this case the minimal value is obtained when $d = 2$ for the nonlinear case.

4. Computational Results

The problem of identifying an optimal kernel for a specific class of data with respect to the real variables is an active research area in the machine learning community ([10]). The polynomial kernel optimization with respect to the integer variable (dimension d) is not obtained by any other researcher to our knowledge. The objective of this section is to obtain computational results for some of the matrices used in [10] where SVM with inhomogeneous polynomial kernel map is used. Our main goal in this section is to find the dimension d where the obtained kernel matrix is satisfied by using the theoretical results obtained previously in this work.

The determinants of the principle minor matrices of an $n \times n$ matrix H will be denoted by $H^{k\times k}$ for all $1 \le k \le n$, and the corresponding set will be denoted by $\overline{H} = \{H^{1\times 1}, H^{2\times 2}, ..., H^{n\times n}\}$.

A kernel matrix can be constructed as a combination of several kernel maps. In [10] , the kernel matrix K is generated by using quadratic, radial-basis, and linear kernel maps with their corresponding kernel matrices K_1, K_2, and K_3, respectively. The regression kernel matrix is computed by using semidefinite programming in [10] . In this paper we first consider the kernel matrix

$$K_1 = \begin{bmatrix} 25 & 9 & 1 & 1 & 9 \\ 9 & 4 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 4 & 9 \\ 9 & 1 & 1 & 9 & 25 \end{bmatrix}$$

constructed in [10]. Since the set of the determinants of the minor matrices corresponding to K is $\overline{H} = \{25, 19, 8, 0, 0\}$, K_1 is a positive semi definite matrix. Therefore, in the non-linear case d can be chosen as 2 which indicates confirms the choice of the dimension in [10].

Suppose we consider the following matrix K obtained in [10] as a combination of the radial-basis, linear and quadratic kernels, and the test set given by

$$\begin{bmatrix} 0.3773 & 0.111 & 0.0222 & 0.111 & 0.3773 & 0.2219 \\ 0.111 & 0.0444 & 0.0222 & 0.0444 & 0.111 & 0.0721 \\ 0.0222 & 0.0222 & 0.0222 & 0.0222 & 0.0222 & 0.0222 \\ 0.111 & 0.0444 & 0.0222 & 0.0444 & 0.111 & 0.0721 \\ 0.3773 & 0.111 & 0.0222 & 0.111 & 0.3773 & 0.2219 \\ 0.2219 & 0.0721 & 0.0222 & 0.0721 & 0.2219 & 0.1345 \end{bmatrix}.$$

Using this matrix we can test whether the combination of all the kernels can be replaced with a polynomial kernel for a certain dimension or not. The set of the determinants of the minor matrices of K is $\overline{H} = \{0.3773,\ 0.0044,\ -4.9284 \times 10^{-8}, 0,\ 0\}$. This indicates that the determinant of the 3×3 minor matrix is not positive definite therefore the polynomial kernel for any dimension is not a good method to use.

5. Mixed Variable Unimodal Functions

The classical definition of a real variable unimodal function indicates $h(t)$ to be unimodal if for some value s, h is monotonically increasing for $t \ge s$ and f is monotonically decreasing for $t \le s$. In that case, the minimum value of h is $h(s)$ and there are no other local minima.

A function $h(t)$ is a weakly unimodal function if for some value s, it satisfies $h(t) \ge h(s)$ for $t \le s$ and $h(t) \le h(s)$ for $t \ge s$. In that case, the minimum value $h(s)$ can be reached for a continuous range of values of t ([8]).

Quadratic programming problems with inhomogeneous polynomial kernel can have a natural unimodal structure for both integer and real variables; therefore, a natural question arises: What is the unimodality of mixed variable functions? In this section, we introduce the mixed variable unimodal and semi-unimodal function definitions where increasing and decreasing function definitions follow from definition 3.1. In addition, we obtain mixed convexity and optimization results for mixed variable unimodal functions.

Definition 5.1: A mixed variable function $\varphi(z)$ is unimodal for some value q if it is strictly decreasing for all $z \leq q$ and it is strictly increasing for all $z \geq q$.

Definition 5.2: A mixed variable function $\varphi(z)$ is semi-unimodal for some value q if it is decreasing for $z \leq q$ and it is increasing for $z \geq q$.

The following result holds similar to the corresponding real variable optimization result in real convex analysis.

Theorem 5.1: A mixed variable unimodal function φ has a unique minimum point.

Proof: Suppose φ is a mixed variable unimodal function. By the definition of unimodal function there exists a vector q_0 such that φ is strictly decreasing for all $z \leq q_0$ indicating

$$\nabla_i \varphi(x, y) < 0 \Rightarrow \varphi(x + e_i, y) < \varphi(x, y),$$

for all $1 \leq i \leq n$, and

$$\frac{\partial \varphi(x, y)}{\partial y_j} < 0,$$

for all $1 \leq j \leq m$. Therefore, since φ is strictly decreasing, there exists a particular value of e_i such that $q_0 = (x + e_i, y)$ and $z = (x, y)$ satisfying

$$\varphi(z) > \varphi(q)$$

Let $q_0 = (x_0, y_0)$. By the definition of the unimodal function, φ is strictly increasing for all $z \geq q_0$, therefore we have

$$\nabla_i \varphi(x, y) > 0 \Rightarrow \varphi(x + e_i, y) > \varphi(x_0, y_0),$$

for all $1 \leq i \leq n$, and

$$\frac{\partial \varphi(x, y)}{\partial y_j} > 0,$$

for all $1 \leq j \leq m$. Since φ is strictly decreasing for all $z \geq q_0$ we have

$$\varphi(z) > \varphi(q).$$

This indicates that the only possible minimum point of φ is $\varphi(q)$ since the existence of any other minimum point would cause a contradiction with the definition of the unimodal function.

Theorem 5.2: A mixed variable semi-unimodal function φ has a unique minimum value however it does not necessarily have a unique minimum point.

Proof: Suppose φ is a mixed variable semi-unimodal function. The proof follows similar to the proof of theorem 5.1 where we replace strict inequalities with inequalities. This indicates that there exists a unique minimum value with possible minimum points of φ where

$$\varphi(z_n) = \varphi(q)$$

for some n with z_n in the domain of φ.

5.1 Examples of Mixed Variable Unimodal Functions

In this section we give simple examples of unimodal and semi-unimodal functions that are condense mixed convex functions.

We will consider $\varphi_i : \mathbb{Z}^2 \times \mathbb{R}^2 \to \mathbb{R}$ for $i = 1$ and $i = 2$. Letting

$$\begin{aligned} \varphi_1(x_1, x_2, y_1, y_2) &= (x_1 - 0.5)^2 + (x_2 + 0.4)^2 \\ &\quad + (y_1 - 1)^2 + (y_2 + 2)^2, \end{aligned}$$

it is easy to see that φ_1 is a semi-unimodal function. There are two points which give the same minimum value of φ_1 which is 0.41.

Letting

$$\begin{aligned} \varphi_2(x_1, x_2, y_1, y_2) &= (x_1 - 0.6)^2 + (x_2 + 0.3)^2 \\ &\quad + (y_1 - 0.34)^2 + (y_2 + 1.7)^2, \end{aligned}$$

φ_2 is a unimodal function therefore it has a unique minimum point $(1, 0, 0.34, -1.7)$.

In the case when the mixed Hessian matrix corresponding to the SVM QP problem with inhomogeneous polynomial kernel is positive definite, it is evident that the model given in (5) without constraints is a mixed unimodal function.

6. Conclusion

In this work we first considered the SVM QP problem (5) with the (inhomogeneous) polynomial kernel

$$k(x, y) = \left(x^T \cdot y + 1\right)^d$$

which is expressed as a mixed convex optimization problem with respect to the real variables $\alpha \in \mathbb{R}^l$ and $b \in \mathbb{R}$, and the integer variable $d \in \mathbb{Z}$. We derived several theoretical mixed convexity and optimization results to determine the conditions to find the minimal dimension d which is considered as an integer variable. Throughout this work several examples of mixed convex functions are given. Several computational results are obtained to determine whether there is an appropriate dimension of the kernel map for the given kernel matrix or not. In the future we plan to use this formulation for large scale data sets.

In addition, we introduced the unimodal and semi-unimodal definitions of mixed variable functions. The minimization results for unimodal and semi-unimodal mixed variable functions are also proven.

References:

[1] Aizerman, M., Braverman E., and Rozonoer L., Theoretical foundations of the potential function method in pattern recognition learning, Automation and Remote Control, Vol. 25, 1964, pp. 821–837.

[2] Das, C., The (S-1,S) Inventory Model under Time Limit on Backorders, Journal of Operations Research, Vol. 25, No. 5, 1977, pp. 835-850.

[3] Ghate V. N., Dudul S. V., Induction machine fault detection using support vector machine based classifier, WSEAS Transactions on Systems, Vol. 8,No. 5, 2009, pp. 591-603.

[4] Harel, A., Sharp and simple bounds for the Erlang delay and loss formulae, Queueing Systems, Vol. 64, No. 2, 2010, pp. 119-143.

[5] Joachims, T., Transductive Inference for Text Classification using Support Vector Machines, Proceedings of the 1999 International Conference on Machine Learning, ICML 1999, pp. 200-209.

[6] Kawabe T., A Control Method with Brain Machine Interface for Man-machine Systems, Proceedings of the 11th WSEAS International Conference on Applied Mathematics, March 22-24, 2007.

[7] Okba T., Ilyes E., Tarek G., Hassani M., A new approach for identification of MIMO non linear system with RKHS model, WSEAS Transactions on Information Science and Applications, Vol. 7, No. 7, 2010, pp. 945-954.

[8] Reklaitis, G.V., Ravindran A., Raqsdell, Engineering Optimization: Methods and Applications, John Wiley & Sons, 2010.

[9] Tokgöz, E., Maalouf, M., Kumin, H., A Hessian matrix for functions with integer and continuous variable, International Journal of Pure and Applied Mathematics, Vol. 57, No. 2, 2009, pp. 209-218.

[10] Trafalis, T.B., Malyscheff, A.M., Computation of the regression kernel matrix using semidefinite programming, TOM 11, Vol. 2, 2006 pp. 3-14.

[11] Vapnik, V, The Nature of Statistical Learning Theory, Springer-Verlag, 1995.