Duals of the QCQP and SDP Sparse SVM

Antoni B. Chan, Nuno Vasconcelos, and Gert R. G. Lanckriet

Statistical Visual Computing Laboratory

SVCL UCSD

SVCL-TR 2007-02 v2

April 2007

Duals of the QCQP and SDP Sparse SVM

Antoni B. Chan, Nuno Vasconcelos, and Gert R. G. Lanckriet

Statistical Visual Computing Lab
Department of Electrical and Computer Engineering
University of California, San Diego

April 2007

Abstract

This is the technical report that accompanies the ICML 2007 paper "Direct Convex Relaxations of Sparse SVM" (Chan et al., 2007). In this report, we derive the dual problems for the SDP and QCQP relaxations of the sparse SVM.

Author email: abchan@ucsd.edu

SVCL Technical reports are available on the SVCL's web page at
http://www.svcl.ucsd.edu

University of California, San Diego
Statistical Visual Computing Laboratory
9500 Gilman Drive, Mail code 0407
EBU 1, Room 5512
La Jolla, CA 92093-0407

1 Dual of the QCQP Sparse SVM

The primal problem for the QCQP relaxation of the sparse SVM is

Problem 1

$$\begin{aligned}
\min_{w,\xi,b,t} \quad & \frac{1}{2}t + C\sum_{i=1}^{N}\xi_i \\
\text{s.t.} \quad & y_i(x_i^T w + b) \geq 1 - \xi_i, \quad i = 1,\dots,N, \\
& \xi_i \geq 0, \\
& \|w\|_1^2 \leq rt, \\
& \|w\|_2^2 \leq t.
\end{aligned}$$

The QCQP primal (Problem 1) can be transformed into an equivalent problem by setting $t = s^2$ and $w = u - v$, for $u, v \in \mathbb{R}_+^d$,

Problem 2

$$\begin{aligned}
\min_{u,v,\xi,b,s} \quad & \frac{1}{2}s^2 + C\sum_{i=1}^{N}\xi_i \\
\text{s.t.} \quad & y_i(x_i^T(u - v) + b) \geq 1 - \xi_i, \quad i = 1,\dots,N, \\
& \xi_i \geq 0, \\
& 1^T(u + v) \leq \sqrt{r}s, \\
& \frac{1}{2}\|u - v\|_2^2 \leq \frac{1}{2}s^2, \\
& u \geq 0, \quad v \geq 0.
\end{aligned}$$

Note that there is an implicit constraint that $s \geq 0$. Introducing the Lagrangian multipliers $(\alpha, \gamma) \in \mathbb{R}_+^N$, $(\mu, \eta) \in \mathbb{R}_+$, and $(\lambda, \theta) \in \mathbb{R}_+^d$, the Langrangian is

$$\begin{aligned}
&\mathcal{L}(u, v, \xi, b, s, \alpha, \mu, \gamma, \eta, \lambda, \theta) \\
&\quad = \frac{1}{2}s^2 + C\sum_i \xi_i + \sum_i \alpha_i\left(1 - \xi_i - y_i x_i^T(u - v) - y_i b\right) - \sum_i \gamma_i \xi_i \\
&\qquad + \mu\left(1^T u + 1^T v - \sqrt{r}s\right) + \frac{\eta}{2}\left((u - v)^T(u - v) - s^2\right) - \lambda^T u - \theta^T v \\
&\quad = \frac{1 - \eta}{2}s^2 - \sqrt{r}\mu s + \sum_i \xi_i(C - \alpha_i - \gamma_i) + \sum_i \alpha_i - b\sum_i \alpha_i y_i \\
&\qquad - \sum_i \alpha_i y_i x_i^T(u - v) + \mu\left(1^T u + 1^T v\right) + \frac{\eta}{2}(u - v)^T(u - v) - \lambda^T u - \theta^T v.
\end{aligned}$$

To minimize $\mathcal{L}$ with respect to (u, v, ξ, b, s), we take the partial derivatives and set them to zero. For s we have

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial s} \quad &= \quad (1 - \eta)s - \mu\sqrt{r} = 0 \\
&\Rightarrow \quad s = \frac{\mu\sqrt{r}}{1 - \eta}, \quad \eta \leq 1
\end{aligned}$$

where the constraint on η follows from the implicit non-negative constraint on s. For b and ξ we have

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial b} &= -\sum_i \alpha_i y_i = 0 \\ &\Rightarrow \sum_i \alpha_i y_i = 0\end{aligned}$$

and

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \xi_i} &= C - \alpha_i - \gamma_i = 0 \\ &\Rightarrow \alpha_i + \gamma_i = C \\ &\Rightarrow \alpha_i \leq C.\end{aligned}$$

The partial derivatives with respect to the hyperplane parameters u and v are

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial u} &= -\sum_i \alpha_i y_i x_i + \mu 1 + \eta(u - v) - \lambda = 0, \\ \frac{\partial \mathcal{L}}{\partial v} &= \sum_i \alpha_i y_i x_i + \mu 1 - \eta(u - v) - \theta = 0.\end{aligned}$$

Adding the two partial derivatives gives a constraint on λ and θ,

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial u} + \frac{\partial \mathcal{L}}{\partial v} &= 2\mu 1 - \lambda - \theta = 0 \\ &\Rightarrow \lambda + \theta = 2\mu 1 \\ &\Rightarrow \lambda \leq 2\mu 1 \\ &\Rightarrow \theta \leq 2\mu 1,\end{aligned}$$

and subtracting the two partial derivatives gives the equation for the hyperplane $w = u - v$,

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial u} - \frac{\partial \mathcal{L}}{\partial v} &= -2\sum_i \alpha_i y_i x_i + 2\eta(u - v) - \lambda + \theta = 0 \\ &\Rightarrow (u - v) = \frac{1}{\eta}\left(\sum_i \alpha_i y_i x_i + \frac{1}{2}(\lambda - \theta)\right) = \frac{1}{\eta} q \\ &\Rightarrow u = \frac{1}{\eta} q + v\end{aligned}$$

where

$$q = \sum_i \alpha_i y_i x_i + \frac{1}{2}(\lambda - \theta).$$

Substituting these conditions back into $\mathcal{L}$, we obtain the g objective function of the dual

$$
\begin{aligned}
&g(\alpha,\mu,\gamma,\eta,\lambda,\theta) = \min_{u,v,\xi,b,s} \mathcal{L}(u,v,\xi,b,s,\alpha,\mu,\gamma,\eta,\lambda,\theta)\\
&= \min_v \frac{r\mu^2}{2(1-\eta)} - \frac{r\mu^2}{1-\eta} + \sum_i \alpha_i - \frac{1}{\eta}\sum_i \alpha_i y_i x_i^T q\\
&\quad + \mu\left(1^T\left(\frac{1}{\eta}q+v\right)+1^T v\right) + \frac{\eta}{2\eta^2}q^T q - \lambda^T\left(\frac{1}{\eta}q+v\right) - \theta^T v\\
&= \frac{-r\mu^2}{2(1-\eta)} + \sum_i \alpha_i - \frac{1}{\eta}\left(q-\frac{1}{2}(\lambda-\theta)\right)^T q + \frac{\mu}{\eta}1^T q + \frac{1}{2\eta}q^T q - \frac{1}{\eta}\lambda^T q\\
&\quad + \min_v (2\mu 1^T v - \lambda^T v - \theta^T v)\\
&= \frac{-r\mu^2}{2(1-\eta)} + \sum_i \alpha_i - \frac{1}{2\eta}q^T q + \frac{1}{2\eta}\lambda - \frac{1}{2\eta}\theta + \frac{\mu}{\eta}1^T q - \frac{1}{\eta}\lambda^T q\\
&\quad + \min_v (2\mu 1 - \lambda - \theta)^T v\\
&= \frac{-r\mu^2}{2(1-\eta)} + \sum_i \alpha_i - \frac{1}{2\eta}q^T q + \frac{1}{2\eta}(2\mu 1 - \lambda - \theta) + \min_v (2\mu 1 - \lambda - \theta)^T v\\
&= \frac{-r\mu^2}{2(1-\eta)} + \sum_i \alpha_i - \frac{1}{2\eta}q^T q,
\end{aligned}
$$

with the constraints

$$
\begin{aligned}
&\sum_i \alpha_i y_i = 0,\\
&\lambda + \theta = 2\mu 1,\\
&0 \le \eta \le 1,\\
&0 \le \alpha_i \le C,\\
&0 \le \theta \le 2\mu 1,\\
&0 \le \lambda \le 2\mu 1,\\
&0 \le \mu.
\end{aligned}
$$

The second constraint and the non-negative constraint on θ and λ implies that

$$-\mu 1 \le \frac{1}{2}(\lambda-\theta) \le \mu 1.$$

Hence, we define $\nu = \frac{1}{2}(\lambda-\theta)$ which leads to the dual problem of the QCQP-SSVM,

Problem 3

$$\max_{\alpha,\eta,\mu,\nu} \qquad \sum_{i=1}^N \alpha_i - \frac{1}{2\eta}\left\|\sum_{i=1}^N \alpha_i y_i x_i + \nu\right\|_2^2 - \frac{r\mu^2}{2(1-\eta)}$$

$$
\begin{aligned}
\text{s.t.} \quad & \sum_{i=1}^{N} \alpha_i y_i = 0, \quad 0 \le \alpha_i \le C, \quad i = 1, \ldots, N, \\
& -\mu \le \nu_j \le \mu, \quad j = 1, \ldots, d, \\
& \mu \ge 0, \quad 0 \le \eta \le 1.
\end{aligned}
$$

1.1 Dual of QCQP-SSVM in RQC form

The dual can be rewritten as a rotated-quadratic-cone (RQC) program by bounding each of the quadratic terms,

$$
\begin{aligned}
t &\ge \frac{1}{\eta} \left\| \sum_{i=1}^{N} \alpha_i y_i x_i + \nu \right\|_2^2 \\
\Rightarrow t\eta &\ge \left\| \sum_{i=1}^{N} \alpha_i y_i x_i + \nu \right\|_2^2
\end{aligned}
$$

and

$$
\begin{aligned}
s &\ge \frac{\mu^2}{(1-\eta)} \\
\Rightarrow s(1-\eta) &\ge \mu^2
\end{aligned}
$$

Hence,

Problem 4

$$
\begin{aligned}
\min_{\alpha,\eta,\mu,\nu,s,t} \quad & \frac{1}{2} t + \frac{r}{2} s - \sum_{i=1}^{N} \alpha_i \\
\text{s.t.} \quad & \sum_{i=1}^{N} \alpha_i y_i = 0, \quad 0 \le \alpha_i \le C, \quad i = 1, \ldots, N, \\
& t\eta \ge \left\| \sum_{i=1}^{N} \alpha_i y_i x_i + \nu \right\|_2^2 \\
& s(1-\eta) \ge \mu^2 \\
& -\mu \le \nu_j \le \mu, \quad j = 1, \ldots, d, \\
& \mu \ge 0, \quad 0 \le \eta \le 1.
\end{aligned}
$$

1.2 Dual of QCQP-SSVM in SDP form

Problem 3 can be rewritten into an SDP by bounding each of the two quadratic terms and using the Schur complement lemma:

$$
t - \frac{1}{\eta} \left(\sum_{i=1}^{N} \alpha_i y_i x_i + \nu \right)^T \left(\sum_{i=1}^{N} \alpha_i y_i x_i + \nu \right) \quad \ge \quad 0
$$

$$\Rightarrow \begin{bmatrix} \eta I & q \\ q^T & t \end{bmatrix} \succeq 0$$

where $q = \sum_i \alpha_i y_i x_i + \nu$, and

$$\begin{aligned} s - \frac{\mu^2}{(1-\eta)} &\geq 0 \\ \Rightarrow \begin{bmatrix} (1-\eta) & \mu \\ \mu & s \end{bmatrix} &\succeq 0 \end{aligned}$$

The dual QCQP-SSVM can then be written as an SDP,

Problem 5

$$\begin{aligned} \min_{\alpha,\eta,\mu,\nu,s,t,q} \quad & \frac{1}{2}t + \frac{r}{2}s - \sum_{i=1}^{N} \alpha_i \\ \text{s.t.} \quad & \sum_{i=1}^{N} \alpha_i y_i = 0, \quad q = \sum_i \alpha_i y_i x_i + \nu, \\ & \begin{bmatrix} \eta I & q \\ q^T & t \end{bmatrix} \succeq 0, \quad \begin{bmatrix} (1-\eta) & \mu \\ \mu & s \end{bmatrix} \succeq 0, \\ & 0 \leq \alpha_i \leq C, \quad i = 1, \ldots, N, \\ & -\mu \leq \nu_j \leq \mu, \quad j = 1, \ldots, d, \\ & \mu \geq 0, \quad 0 \leq \eta \leq 1. \end{aligned}$$

2 Dual of the SDP sparse SVM

The primal problem for the SDP relaxation of the sparse SVM is

Problem 6

$$\begin{aligned} \min_{W,w,b,\xi} \quad & \frac{1}{2}\mathrm{tr}(W) + C \sum_{i=1}^{N} \xi_i \\ \text{s.t.} \quad & y_i(x_i^T w + b) \geq 1 - \xi_i, \quad i = 1, \ldots, N, \\ & \xi_i \geq 0, \\ & 1^T |W| 1 \leq r \mathrm{tr}(W), \\ & W - ww^T \succeq 0 \end{aligned}$$

By introducing two matrices $(U, V) \in \mathbb{R}_+^{d \times d}$ ($d \times d$ matrices with non-negative entries) such that $W = U - V$, the SDP-SSVM primal (Problem 6) is equivalent to

Problem 7

$$\min_{U,V,w,b,\xi} \quad \frac{1}{2}\mathrm{tr}(U - V) + C \sum_{i=1}^{N} \xi_i$$

$$
\begin{aligned}
\text{s.t.} \quad & y_i(x_i^T w + b) \geq 1 - \xi_i, \quad i = 1, \ldots, N, \\
& \xi_i \geq 0, \\
& \frac{1}{2} 1^T (U + V) 1 \leq \frac{r}{2} \operatorname{tr}(U - V), \\
& \frac{1}{2}(U - V) - \frac{1}{2} w w^T \succeq 0 \\
& U_{j,k} \geq 0, \quad V_{j,k} \geq 0, \quad j, k = 1, \ldots, d.
\end{aligned}
$$

Introducing the Lagrangian multipliers $(\alpha, \gamma) \in \mathbb{R}_+^N$, $\mu \in \mathbb{R}_+$, the matrix $\Lambda \in \mathbb{R}^{d \times d} \succeq 0$, and the matrices $(\eta, \theta) \in \mathbb{R}_+^{d \times d}$, the Lagrangian is

$$
\begin{aligned}
&\mathcal{L}(U, V, w, b, \xi, \alpha, \gamma, \eta, \theta, \mu, \Lambda) \\
&\quad = \frac{1}{2} \operatorname{tr}(U - V) + C \sum_i \xi_i + \sum_i \alpha_i (1 - \xi_i - y_i x_i^T w - y_i b) - \sum_i \gamma_i \xi_i - \operatorname{tr}(\eta U) \\
&\qquad - \operatorname{tr}(\theta V) - \frac{1}{2} \operatorname{tr}(\Lambda (U - V - w w^T)) + \frac{\mu}{2} \left(1^T (U + V) 1 - r \operatorname{tr}(U - V) \right) \\
&\quad = \frac{1}{2} \operatorname{tr}\left((I - \Lambda - \mu r I)(U - V) \right) - \operatorname{tr}(\eta U) - \operatorname{tr}(\theta V) + \frac{\mu}{2} \operatorname{tr}((U + V) 1\, 1^T) \\
&\qquad + \sum_i \xi_i (C - \alpha_i - \gamma_i) + \sum_i \alpha_i - \sum_i \alpha_i y_i x_i^T w + \frac{1}{2} w^T \Lambda w - \sum_i \alpha_i y_i b
\end{aligned}
$$

The Lagrangian is minimized with respect to (U, V, w, b, ξ) by taking the partial derivatives and setting them to zero. For b, ξ, and w we have

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial b} &= -\sum_i \alpha_i y_i = 0 \\
&\Rightarrow \sum_i \alpha_i y_i = 0
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \xi_i} &= C - \alpha_i - \gamma_i = 0 \\
&\Rightarrow \alpha_i + \gamma_i = C \\
&\Rightarrow \alpha_i \leq C
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial w} &= -\sum_i \alpha_i y_i x_i + \Lambda w = 0 \\
&\Rightarrow w = \Lambda^{-1} \sum_i \alpha_i y_i x_i.
\end{aligned}
$$

The partial derivatives with respect to the matrices U and V are

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial U} &= \frac{1}{2}(I - \Lambda - \mu r I) - \eta + \frac{\mu}{2} 1\, 1^T = 0 \\
\frac{\partial \mathcal{L}}{\partial V} &= -\frac{1}{2}(I - \Lambda - \mu r I) - \theta + \frac{\mu}{2} 1\, 1^T = 0
\end{aligned}
$$

Adding the partial derivatives yields a constraint on θ and η,

$$\begin{aligned}\frac{\partial\mathcal{L}}{\partial U}+\frac{\partial\mathcal{L}}{\partial V} &= -\eta-\theta+\mu 1\, 1^T=0\\ &\Rightarrow \eta+\theta=\mu 1\, 1^T\end{aligned}$$

Subtracting the partial derivatives yields a formula for the weight matrix,

$$\begin{aligned}\frac{\partial\mathcal{L}}{\partial U}-\frac{\partial\mathcal{L}}{\partial V} &= (I-\Lambda-\mu rI)-\eta+\theta=0\\ &\Rightarrow \Lambda=(1-\mu r)I-\eta+\theta\end{aligned}$$

The dual objective function is obtained by substituting these conditions into $\mathcal{L}$,

$$\begin{aligned}
&g(\alpha,\gamma,\eta,\theta,\mu,\Lambda)=\min_{U,V,w,b,\xi}\mathcal{L}(U,V,w,b,\xi,\alpha,\gamma,\eta,\theta,\mu,\Lambda)\\
&= \min_{U,V}\frac{1}{2}\mathrm{tr}\left((\eta-\theta)(U-V)\right)-\mathrm{tr}(\eta U)-\mathrm{tr}(\theta V)+\frac{1}{2}\mathrm{tr}((U+V)(\eta+\theta))\\
&\quad +\sum_i\alpha_i-\sum_{i,j}\alpha_i\alpha_jy_iy_jx_i^T\Lambda^{-1}x_j+\frac{1}{2}\sum_{i,j}\alpha_i\alpha_jy_iy_jx_i^T\Lambda^{-1}x_j\\
&= \sum_i\alpha_i-\frac{1}{2}\sum_{i,j}\alpha_i\alpha_jy_iy_jx_i^T\Lambda^{-1}x_j\\
&\quad +\min_{U,V}\frac{1}{2}\mathrm{tr}\left(\eta U-\theta U-\eta V+\theta V-2\eta U-2\theta V\right)+\frac{1}{2}\mathrm{tr}((U+V)(\eta+\theta))\\
&= \sum_i\alpha_i-\frac{1}{2}\sum_{i,j}\alpha_i\alpha_jy_iy_jx_i^T\Lambda^{-1}x_j
\end{aligned}$$

with constraints

$$\begin{aligned}
&\sum_i\alpha_iy_i=0,\\
&\eta+\theta=\mu 1\, 1^T,\\
&\Lambda=(1-\mu r)I-\eta+\theta\succeq 0,\\
&0\le\alpha_i\le C,\quad i=1,\dots,N,\\
&0\le\mu,\\
&0\le\eta_{j,k},\quad 0\le\theta_{j,k},\quad j,k=1,\dots,d.
\end{aligned}$$

The second constraint and the non-negative constraints on η and θ implies that

$$-\mu\le\theta_{j,k}-\eta_{j,k}\le\mu,\quad j,k=1,\dots,d.$$

Finally, defining $\nu=\theta-\eta$, the dual problem of the SDP-SSVM is

Problem 8

$$\max_{\alpha,\Lambda,\nu,\mu}\qquad \sum_{i=1}^N\alpha_i-\frac{1}{2}\sum_{i=1}^N\sum_{j=1}^N\alpha_i\alpha_jy_iy_jx_i^T\Lambda^{-1}x_j$$

$$
\begin{aligned}
\text{s.t.} \quad & \sum_{i=1}^{N} \alpha_i y_i = 0, \\
& \Lambda = (1 - \mu r) I + \nu \succeq 0, \\
& 0 \le \alpha_i \le C, \quad i = 1, \dots, N, \\
& -\mu \le \nu_{j,k} \le \mu, \quad j,k = 1, \dots, d, \\
& \mu \ge 0.
\end{aligned}
$$

2.1 Dual in SDP form

The dual can be written into SDP form by bounding the quadratic term by t, and using the Schur complement lemma (Boyd & Vandenberghe, 2004).

$$
\begin{aligned}
t & \ge \sum_{i,j} \alpha_i \alpha_j y_i y_j x_i^T \Lambda^{-1} x_j \\
t - \left(\sum_i \alpha_i y_i x_i \right)^T \Lambda^{-1} \left(\sum_i \alpha_i y_i x_i \right) & \ge 0 \\
\begin{bmatrix} \Lambda & q \\ q^T & t \end{bmatrix} & \succeq 0
\end{aligned}
$$

where $q = \sum_i \alpha_i y_i x_i$. The dual can then be written as an SDP,

Problem 9

$$
\begin{aligned}
\min_{\alpha,\mu,\Lambda,\nu,q} \quad & \frac{1}{2} t - \sum_{i=1}^{N} \alpha_i \\
\text{s.t.} \quad & \sum_{i=1}^{N} \alpha_i y_i = 0, \quad q = \sum_{i=1}^{N} \alpha_i y_i x_i, \\
& \Lambda = (1 - \mu r) I + \nu, \\
& \begin{bmatrix} \Lambda & q \\ q^T & t \end{bmatrix} \succeq 0, \\
& 0 \le \alpha_i \le C, \quad i = 1, \dots, N, \\
& -\mu \le \nu_{j,k} \le \mu, \quad j,k = 1, \dots, d, \\
& \mu \ge 0.
\end{aligned}
$$

References

Boyd, S., & Vandenberghe, L. (2004). *Convex optimization*. Cambridge University Press.

Chan, A. B., Vasconcelos, N., & Lanckriet, G. R. G. (2007). Direct convex relaxations of sparse SVM. *International Conference on Machine Learning*.

SVCL-TR 2007-02 v2

April 2007

Duals of the QCQP and SDP Sparse SVM

Antoni B. Chan