CS281B/Stat241B (Spring 2008) Statistical Learning Theory **Lecture: 5**

Constrained Optimization

Lecturer: Peter Bartlett *Scribe: Nimar S. Arora*

The reference for this lecture is Chapter 5 of Boyd and Vanderberghe's *Convex Optimization*.

1 Primal

Consider the optimization problem *(primal problem)*:

$$p^* = \min_{x \in \mathbb{R}^n} \; f_0(x)$$
$$\text{s.t. } f_i(x) \le 0, \qquad i = 1, \dots, m$$

The optimal value is

$$p^* = f_0(x^*)$$

Define the Lagrangian:

$$L : \mathbb{R}^{n+m} \to \mathbb{R}$$
$$L(x, \lambda) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x)$$

The λ_is are called dual variables or Lagrange multipliers with $\lambda_i \ge 0$

2 Saddle Point

See Figure 1 for an example of a saddle point.

In a minimax problem, if the min player gets to play second he can achieve a lower value. Thus,

$$d^* = \sup_{\lambda \ge 0} \; \inf_x L(x, \lambda) \;\le\; \inf_x \; \sup_{\lambda \ge 0} \; L(x, \lambda)$$

Suppose there are x^* and λ^* s.t.,

$$L(x^*, \lambda) \le L(x^*, \lambda^*) \le L(x, \lambda^*)$$

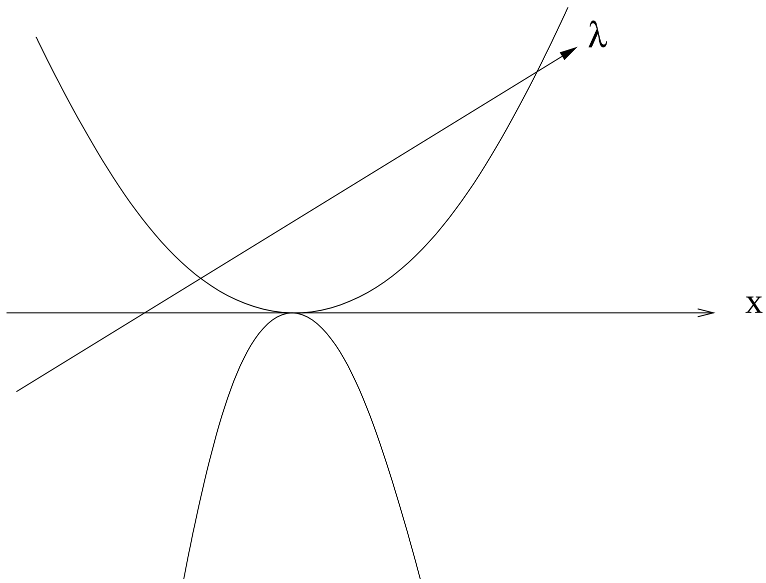

Figure 1: Saddle Point

for all feasible x and $\lambda \geq 0$. Then,

$$\begin{aligned}
\inf_x \sup_{\lambda \geq 0} L(x,\lambda) &\leq \sup_{\lambda \geq 0} L(x^*,\lambda) \qquad (\text{fix } x = x^*) \\
&= L(x^*,\lambda^*) \\
&= \inf_x L(x,\lambda^*) \\
&\leq \sup_{\lambda \geq 0} \inf_x L(x,\lambda) \qquad (\text{fixing } \lambda = \lambda^* \text{ we get previous}) \\
\text{So,} \qquad \inf_x \sup_{\lambda \geq 0} L(x,\lambda) &= \sup_{\lambda \geq 0} \inf_x L(x,\lambda)
\end{aligned}$$

3 Lagrange Dual Function

Define the Lagrange dual function:

$$\begin{aligned}
g(\lambda) &= \inf_x \; L(x,\lambda) \\
&= \inf_x \; (f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x))
\end{aligned}$$

Note,

1. $g(\lambda)$ is concave (point-wise minima of concave functions)
2. If $\lambda_i \geq 0$ and x is primal feasible (i.e. $f_i(x) \leq 0$) then,

$$g(\lambda) \;\leq\; f_0(x)$$

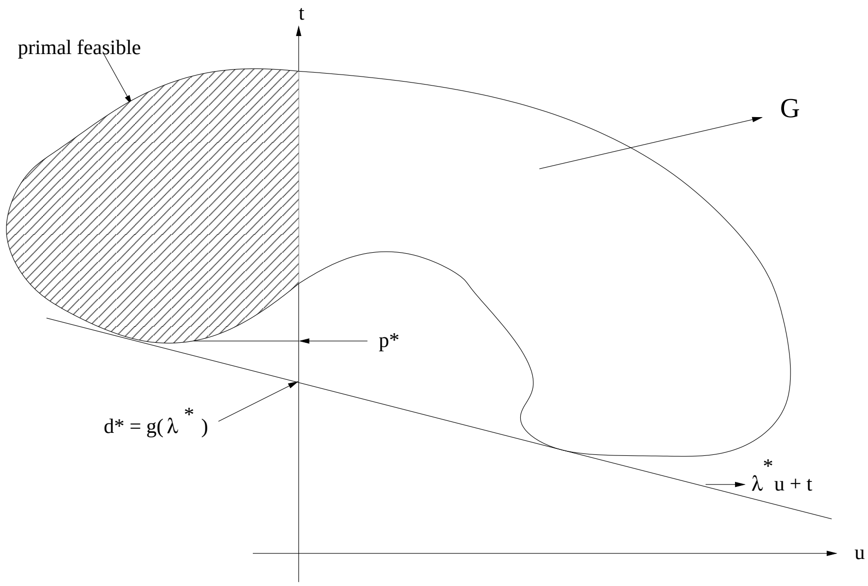

Figure 2: Geometric Interpretation of Duality

3. In particular, $\forall\lambda \geq 0$, $g(\lambda) \leq p^*$

4 Dual

$$\begin{aligned} \text{Dual:} &\quad \max\ g(\lambda) \\ &\quad \text{s.t. } \lambda \geq 0 \\ \text{Optimal Value:} &\quad g(\lambda^*) = d^* \end{aligned}$$

Note,

1. The dual is always a maximization of a concave function with convex constraints
2. Weak duality implies that $d^* \leq p^*$
3. The optimal duality gap is $p^* - d^*$

5 Geometric Interpretation

Define,

$$\mathcal{G} = \{(u,t) : \ \exists x \ f_i(x) = u_i; \ f_0(x) = t\}$$

$$
\begin{aligned}
g(\lambda) &= \inf_x \{f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x)\} \\
&= \inf_{(u,t)\in\mathcal{G}} (t + \sum_{i=1}^{m} \lambda_i u_i) \\
&= \left[\lambda^T \ 1\right]^T \begin{bmatrix} u \\ t \end{bmatrix}
\end{aligned}
$$

For m=1, the set

$$\{(u,t) : (\lambda \ 1)^T \begin{pmatrix} u \\ t \end{pmatrix} = c\}$$

is a line with slope λ and intercept $t = c = g(\lambda)$. See Figure 2 for an illustration of the set G and the Lagrange Dual.

6 Strong Duality

Weak duality states that $d^* \leq p^*$. Strong duality states $d^* = p^*$. Strong duality holds if f_0 and f_i are convex and there is a suitable qualification on the constraint. For example, Slater's condition requires that the primal is strictly feasible:

$$\exists x \qquad f_i(x) < 0 \qquad i = 1 \dots m$$

7 Complementary Slackness

If there is zero duality gap,

$$
\begin{aligned}
f_0(x^*) &= g(\lambda^*) \\
&= \inf_x \left(f_0(x) + \sum_{i=1}^{m} \lambda_i^* f_i(x) \right) \\
&\leq f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* f_i(x^*) \qquad \text{(fixing } x = x^*\text{)}
\end{aligned}
$$

$$
\begin{aligned}
\text{Hence,} \qquad \sum_{i=1}^{m} \lambda_i^* f_i(x^*) &\geq 0 \\
\text{But,} \qquad f_i(x^*) &\leq 0 \\
\text{and} \qquad \lambda_i^* &\geq 0 \\
\text{So,} \qquad \sum_{i=1}^{m} \lambda_i^* f_i(x^*) &= 0 \\
\text{and hence} \qquad \lambda_i^* f_i(x^*) &= 0 \qquad \forall i
\end{aligned}
$$

If constraint i is inactive at x^* (i.e. $f_i(x^*) < 0$) then $\lambda_i^* = 0$.

8 KKT Optimality Conditions

If f_0 and f_i are differentiable, $\exists\ x^*,\ \lambda^*$ which are optimal, and the duality gap is zero

$$\Rightarrow\ KKT(x^*,\lambda^*)\ =\ \begin{cases} f_i(x^*) \ \le\ 0 & i=1\dots m \\ \lambda_i^* \ \ge\ 0 & i=1\dots m \\ \lambda_i^* f_i(x^*) \ =\ 0 & i=1\dots m \\ \nabla f_0(x^*)\ +\ \sum_{i=1}^m \lambda_i^* \nabla f_i(x^*) \ =\ 0 & \end{cases}$$

Also, $KKT(x,\lambda)$ and f_0, f_i convex $\Rightarrow\ x,\ \lambda$ are optimal and the duality gap is zero.

If f_0, f_i are convex, differentiable, and the duality gap is zero then $KKT(x,\lambda)\ \Leftrightarrow\ (x,\lambda)$ optimal.

9 SVM

$$\text{Primal}\quad \boxed{\begin{array}{lll} \min & \frac{1}{2}\|w\|^2 & \\ \text{s.t.} & y_i w' x_i\ \ge\ 1 & i=\dots m \end{array}}$$

$$L(w,\alpha)\ =\frac{1}{2}\|w\|^2\ +\ \sum_{i=1}^n \alpha_i(1\ -\ y_i w' x_i)$$

$$\text{substituting}\qquad w^*\ =\sum_{i=1}^n \alpha_i y_i x_i$$

$$g(\alpha)\ =\ \sum_{i=1}^n \alpha_i\ -\ \frac{1}{2}\sum_{i,j}\alpha_i\alpha_j y_i y_j x_i' x_j$$

Strong duality (if feasible).

Complementary Slackness:

$$y_i {w^*}' x_i\ > 1\ \Rightarrow\ \alpha_i^*\ =\ 0$$

$$\text{for } i \text{ s.t.}\qquad \alpha_i^*\ > 0\qquad x_i \text{ is a support vector}$$