Convex Duality and Interior Point Method

Stanley

Outline

- Convexity
- Duality
- Optimality
- Examples: Linear Prog. and SVM
- (Interior Point Method)

Convex Set

A set $S \in R^n$ is said convex if
$\forall x, y \in S$, and $0 \leq \theta \leq 1$,

$$\theta x + (1 - \theta)y \in S$$

Examples:

- $\{x \in R^n : |x - x_0| < r,$ for some $x_0 \in R^n,$ and some $r > 0\}$
- $\{x \in R^n : a_i^T x \geq b, \quad i = 1, \ldots, m\}$

Thm 1: $\bigcap_\alpha S_\alpha$ is convex

Theorem 1
If $\{S_\alpha\}$ is convex, then $\bigcap_\alpha S_\alpha$ is convex.

proof:

$$x, y \in \bigcap_\alpha S_\alpha \Rightarrow x, y \in S_\alpha, \forall \alpha.$$

S_α is convex $\Rightarrow \forall \theta \in [0, 1], \theta x + (1 - \theta)y \in S_\alpha, \forall \alpha$

$$\Rightarrow \theta x + (1 - \theta)y \in \bigcap_\alpha S_\alpha$$

Convex Function

A function $f : R^n \rightarrow R$ is said convex if
1. dom$f$ is convex
2.$\forall x, y \in$ dom$f$, and $0 \leq \theta \leq 1$,

$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y)$$

remark: f is concave if $-f$ is convex.

Optimization Problem

General optimization is in the form: (primal problem)

$$\begin{aligned} \text{min}\, & f_0(x) \\ \text{s.t.}\, & f_i(x) \leq 0, i = 1, \ldots, m \\ & h_i(x) = 0, i = 1, \ldots, p \end{aligned}$$

We make NO assumption about the convexity.

Lagrangian $L(x, \lambda, \nu)$

Lagrangian is defined as
$L : R^n \times R^m \times R^p \to R$:

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x)$$

We call λ, ν dual variables, or Lagrange multiplier vector.

Lagrange Dual Function $g(\lambda, \nu)$

Lagrangian dual function is defined as
$g : R^m \times R^p \longrightarrow R$:

$$g(\lambda, \nu) = \inf_{x \in D} \left\{ f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \right\}$$

Thm 2: $g(\lambda, \nu)$ is concave

Theorem 2
$g(\lambda, \nu)$ is concave.

Proof:
First, claim that inf$(f + g) \geq$ inf$f +$ inf$g$.
proof: Use the fact

1. sup f = -inf (-f).
2. sup$(f + g) \leq$ sup$f +$ sup$g$

$g(\lambda, \nu)$ is concave (2)

Next, consider $\phi_i = (\lambda_i, \nu_i) \in$ dom$(-g), i = 1, 2$. We can show that

$$
\begin{aligned}
-g(\theta\phi_1 + (1-\theta)\phi_2) &= \dots \\
&\leq -\{\theta g(\phi_1) + (1-\theta) g(\phi_2)\}
\end{aligned}
$$

Thus $-g(\lambda, \nu)$ is convex, hence $g(\lambda, \nu)$ is concave.

Thm 3: $g(\lambda, \nu) \leq p^*$

Theorem 3
If p^* is the optimal value of the primal problem, then $\forall \lambda \geq 0$, and any ν, we have

$$g(\lambda, \nu) \leq p^*.$$

Proof: Since $\lambda_i \geq 0$, $f_i(x^*) \leq 0$, $h_i(x^*) = 0$, we have $\sum_{i=1}^m \lambda_i f_i(x^*) + \sum_{i=1}^p \nu_i h_i(x^*) \leq 0$. So,

$$g(\lambda, \nu) = \text{inf } L(x, \lambda, \nu) \leq L(x^*, \lambda, \nu)$$

$$= f_0(x^*) + \sum_{i=1}^m \lambda_i f_i(x^*) + \sum_{i=1}^p \nu_i h_i(x^*) \leq f_0(x^*)$$

Dual Problem

Theorem 3 says $g(\lambda, \nu)$ is a lower bound.
$\Rightarrow$ want to find the best lower bound!

Dual problem:

$$\begin{aligned} \max_{\lambda\,\nu} \quad & g(\lambda, \nu) \\ \text{s.t.} \quad & \lambda \geq 0 \end{aligned}$$

This is convex!

Strong and Weak Duality

Let p^* be the optimal value of primal, and d^* of dual, then since $g(\lambda, \nu) \leq p^*, \forall \lambda, \nu$ (Thm 3)
Weak duality:

$$d^* \leq p^*$$

always hold.

If

$$d^* = p^*$$

then we say **Strong duality** holds (e.g. when primal constraints are convex.)

Geometric Interpretation

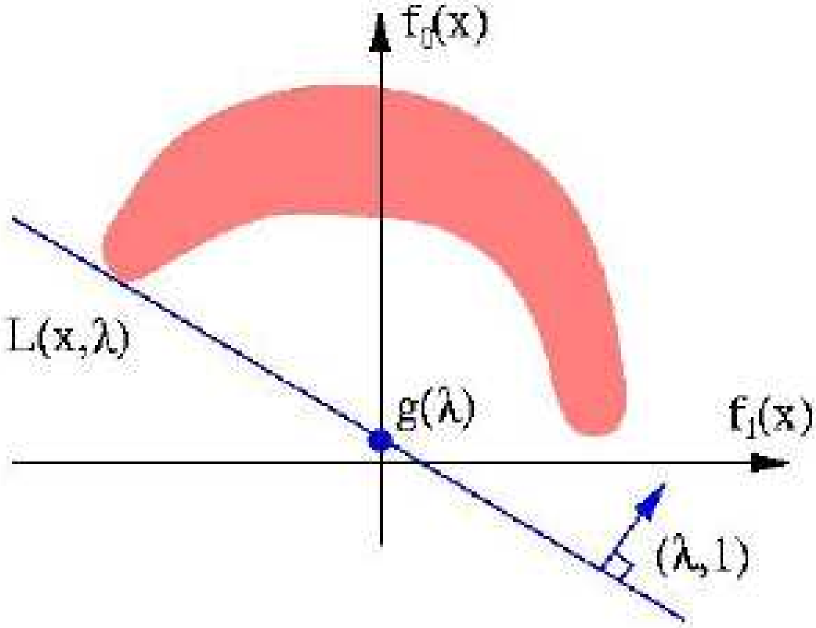

Figure 1: Blue lines is $L(x, \lambda) = f_0(x) + \lambda f_1(x)$ and λ is the slope. $g(\lambda) = \inf L$ is achieved at $f_1(x) = 0$.

Geometric Interpretation (2)

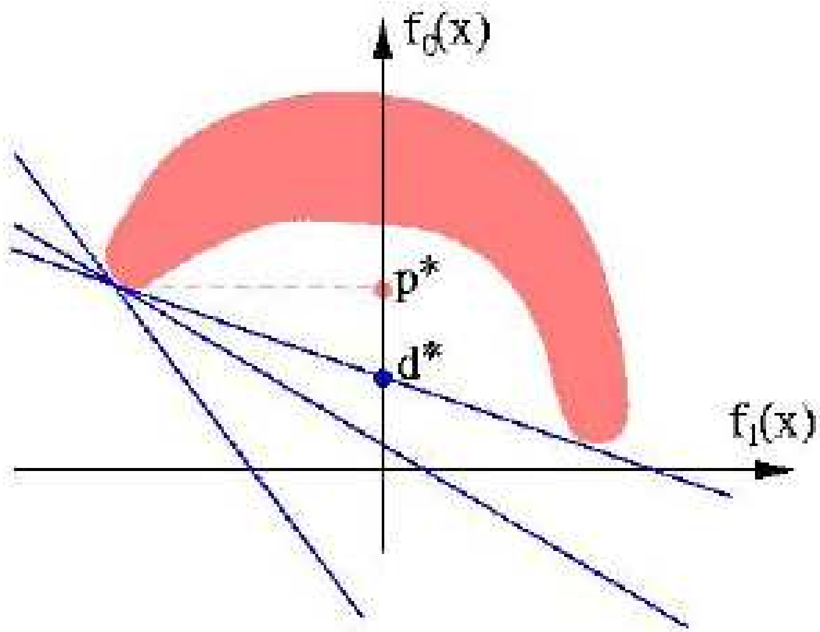

Figure 2: If primal is not convex, then $p^* \neq d^*$

KKT Condition

KKT: If x^* is a minimizer, then

1. (feasibility) $f_i(x^*) \leq 0$, $h_i(x^*) \leq 0$
2. (dual feasible) $\lambda_i^* \geq 0$
3. (complementary slackness) $\lambda_i^* f_i(x^*) = 0$
4. (first order stationarity)
   $\nabla f_0(x^*) + \sum \lambda_i \nabla f_i(x^*) + \sum \nu_i \nabla h_i(x^*) = 0$.

The converse is true only if the primal is convex.

Why dual?

1. Dual optimal = primal optimal if primal is convex
2. Dual problem is always convex
3. Convex problems can be solved efficiently via interior method

Example 1: Linear Program

Linear Programming in standard form:

$$\begin{aligned} \text{min} \quad & c^T x \\ \text{s.t.} \quad & Ax = b \\ & x \geq 0 \end{aligned}$$

Let's start from the Lagrangian:

$$\begin{aligned} L(x, \lambda, \nu) &= c^T x - \lambda^T x + \nu^T (Ax - b) \\ &= -b^T \nu + (c + A^T \nu - \lambda)^T x \end{aligned}$$

Example 1: Linear Prog. (2)

Now, having the Lagrangian,

$$L(x, \lambda, \nu) = -b^T \nu + (c + A^T \nu - \lambda)^T x,$$

we can compute Lagrange dual function

$$\begin{aligned} g(\lambda, \nu) &= \inf_{x \in D} L(x, \lambda, \nu) \\ &= -b^T \nu + \inf_{x \in D} \{(c + A^T \nu - \lambda)^T x\} \\ &= \begin{cases} -b^T \nu & : \quad c + A^T \nu - \lambda = 0, \\ -\infty & : \quad \text{otherwise.} \end{cases} \end{aligned}$$

Example 1: Linear Prog. (3)

Therefore, the dual problem is:

$$\begin{aligned} \max_{\nu,\lambda} \quad & -b^T\nu \\ \text{s.t.} \quad & c + A^T\nu - \lambda = 0, \quad \lambda \geq 0 \end{aligned}$$

Equivalent to:

$$\begin{aligned} \max_{\nu} \quad & -b^T\nu \\ \text{s.t.} \quad & c + A^T\nu \geq 0 \end{aligned}$$

Example 2: SV classification

Optimal Hyperplane problem:

$$\min_{\mathbf{w},b} \quad \frac{1}{2}\|\mathbf{w}\|^2$$
$$\text{s.t.} \quad y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1, \quad i = 1, \ldots, m.$$

The Lagrangian is

$$L((\mathbf{w}, b), \lambda) = \frac{1}{2}\|\mathbf{w}\|^2 + \sum_{i=1}^{m} \lambda_i \left(1 - y_i(\mathbf{w}^T\mathbf{x}_i + b)\right)$$

Example 2: SV classification (2)

Lagrange dual function is

$$g(\lambda) = \inf_{\mathbf{w},b} L((\mathbf{w}, b), \lambda)$$

The infimum can be achieved when

$$\frac{\partial}{\partial \mathbf{w}} L = 0 \quad \text{and} \quad \frac{\partial}{\partial b} L = 0.$$

This leads to

$$\mathbf{w} = \sum_{i=1}^{m} \lambda_i y_i \mathbf{x}_i \quad \text{and} \quad \sum_{i=1}^{m} \lambda_i y_i = 0$$

Example 2: SV classification (3)

So we have

$$g(\lambda) = \sum_{i=1}^{m} \lambda_i - \frac{1}{2} \sum_{i,j=1}^{m} \lambda_i \lambda_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j$$

Hence the dual problem is:

$$\max_{\lambda} \quad \sum_{i=1}^{m} \lambda_i - \frac{1}{2} \sum_{i,j=1}^{m} \lambda_i \lambda_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j$$

$$\text{s.t.} \quad \sum_{i=1}^{m} \lambda_i y_i = 0, \quad \lambda \geq 0.$$

Questions?

Thank you.