对偶理论(续)

王 尧

西安交通大学智能决策与机器学习中心
(Email: yao.s.wang@gmail.com)

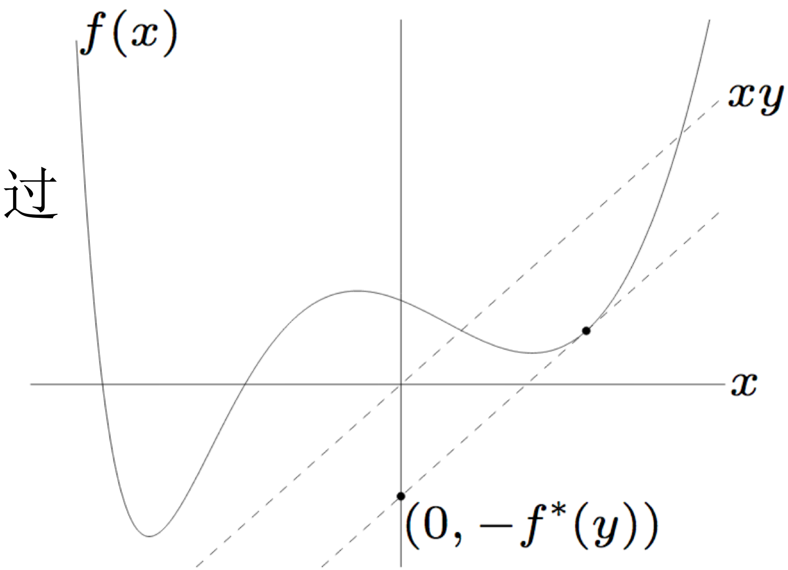
2022. 5

回顾：共轭函数

the **conjugate** of a function f is

$$f^*(y) = \sup_{x \in \mathbf{dom}\, f} (y^T x - f(x))$$

找到f上的一点x，使得通过它的切线斜率为 y



- f^* is convex (even if f is not)

回顾：Lagrange对偶与共轭函数

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & Ax \preceq b, \quad Cx = d \end{array}$$

dual function

$$\begin{aligned} g(\lambda, \nu) &= \inf_{x \in \mathbf{dom}\, f_0} \left(f_0(x) + (A^T\lambda + C^T\nu)^T x - b^T\lambda - d^T\nu \right) \\ &= -f_0^*(-A^T\lambda - C^T\nu) - b^T\lambda - d^T\nu \end{aligned}$$

- recall definition of conjugate $f^*(y) = \sup_{x \in \mathbf{dom}\, f} (y^T x - f(x))$
- simplifies derivation of dual if conjugate of f_0 is known

回顾：等式约束范数极小化

$$\begin{array}{ll} \text{minimize} & \|x\| \\ \text{subject to} & Ax = b \end{array}$$

dual function

$$g(\nu) = \inf_x (\|x\| - \nu^T A x + b^T \nu) = \begin{cases} b^T \nu & \|A^T \nu\|_* \le 1 \\ -\infty & \text{otherwise} \end{cases}$$

where $\|v\|_* = \sup_{\|u\| \le 1} u^T v$ is dual norm of $\|\cdot\|$

The conjugate of $f_0 = \|\cdot\|$ is given by

$$f_0^*(y) = \begin{cases} 0 & \|y\|_* \le 1 \\ \infty & \text{otherwise} \end{cases}$$

回顾： Lagrange对偶问题

Lagrange dual problem

$$\begin{array}{ll} \text{maximize} & g(\lambda, \nu) \\ \text{subject to} & \lambda \succeq 0 \end{array}$$

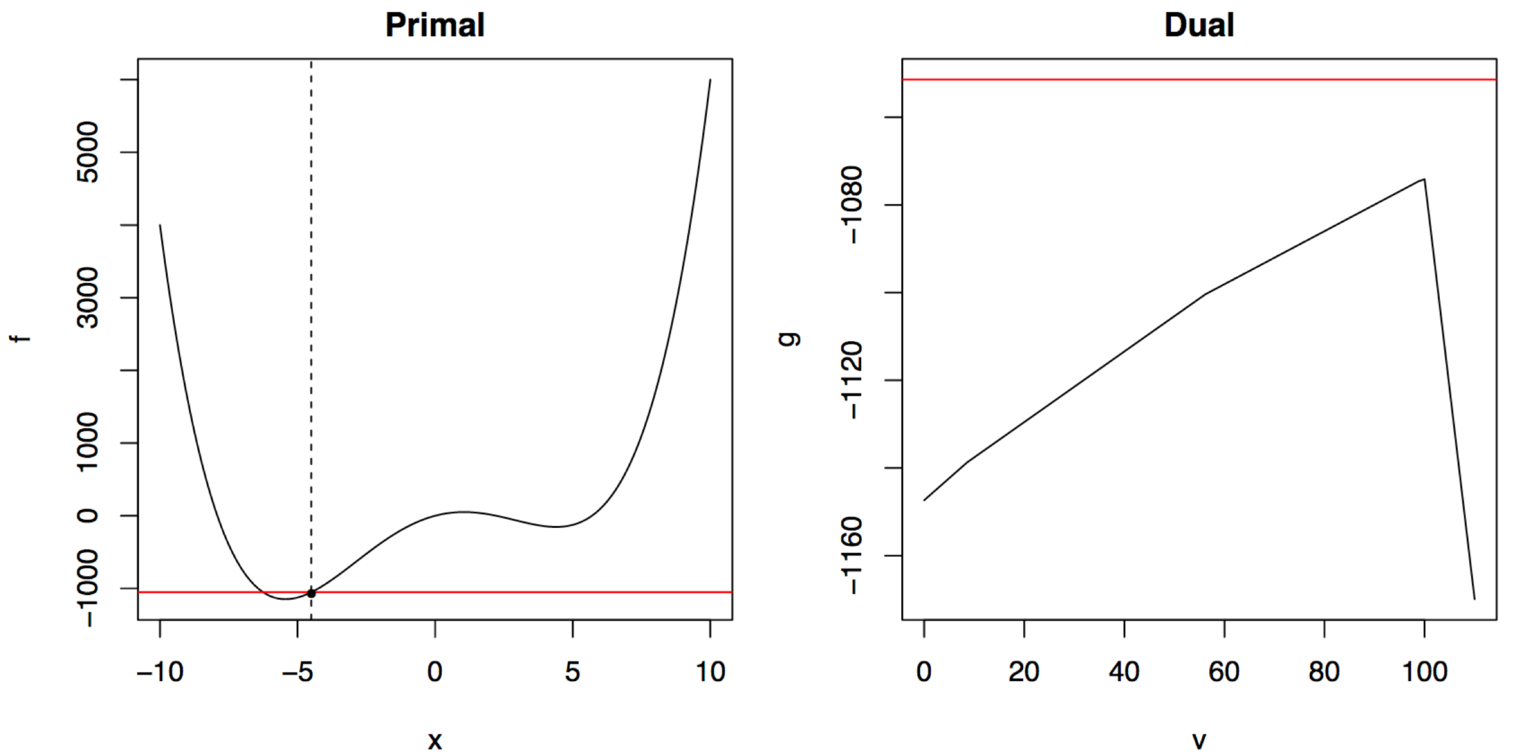
- finds best lower bound on $p^\star$, obtained from Lagrange dual function
- a convex optimization problem; optimal value denoted $d^\star$
- λ, ν are dual feasible if $\lambda \succeq 0$, $(\lambda, \nu) \in \mathbf{dom}\, g$

example: standard form LP and its dual

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & x \succeq 0 \end{array} \qquad \begin{array}{ll} \text{maximize} & -b^T \nu \\ \text{subject to} & A^T \nu + c \succeq 0 \end{array}$$

Example: nonconvex quartic minimization

Define $f(x) = x^4 - 50x^2 + 100x$ (nonconvex), minimize subject to constraint $x \geq -4.5$



弱对偶与强对偶

weak duality: $d^\star \leq p^\star$

- always holds (for convex and nonconvex problems)
- can be used to find nontrivial lower bounds for difficult problems

strong duality: $d^\star = p^\star$

- does not hold in general
- (usually) holds for convex problems
- conditions that guarantee strong duality in convex problems are called **constraint qualifications** (约束规范性条件)

Duality gap: $p^* - d^*$

A Counterexample

$$\min_{x,y>0} \; e^{-x} \;\; : \;\; x^2/y \leq 0$$

- We have $p^* = 1.$
- The Lagrange dual function is

$$g(\lambda) = \inf_{x,y>0} (e^{-x} + \lambda x^2/y) = \begin{cases} 0 & \lambda \geq 0 \\ -\infty & \lambda < 0 \end{cases}$$

- We can write the dual problem as

$$d^* = \max_{\lambda} \; 0 \;\; : \;\; \lambda \geq 0$$

Note: Convexity alone is not enough to guarantee strong duality.

Slater条件

strong duality holds for a convex problem

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & Ax = b \end{array}$$

if it is strictly feasible, $i.e.$,

$$\exists x \in \mathbf{int}\, \mathcal{D}: \qquad f_i(x) < 0, \quad i = 1, \ldots, m, \qquad Ax = b$$

- also guarantees that the dual optimum is attained (if $p^\star > -\infty$)

Refinement: actually only need strict inequalities for non-affine f_i

通常称为weak slater条件，更多介绍参见Boyd版5.2.3节与5.3.2节

二次规划

primal problem (assume $P \in \mathbf{S}^n_{++}$)

$$\begin{array}{ll} \text{minimize} & x^T P x \\ \text{subject to} & Ax \preceq b \end{array}$$

dual function

$$g(\lambda) = \inf_x \left(x^T P x + \lambda^T (Ax - b)\right) = -\frac{1}{4}\lambda^T A P^{-1} A^T \lambda - b^T \lambda$$

dual problem

$$\begin{array}{ll} \text{maximize} & -(1/4)\lambda^T A P^{-1} A^T \lambda - b^T \lambda \\ \text{subject to} & \lambda \succeq 0 \end{array}$$

- from Slater's condition: $p^\star = d^\star$ if $A\tilde{x} \prec b$ for some $\tilde{x}$
- in fact, $p^\star = d^\star$ always (除非原问题无可行解，由weak slater条件即可知)

不等式约束的线性规划

primal problem

$$\begin{array}{ll}\text{minimize} & c^T x \\ \text{subject to} & Ax \preceq b\end{array}$$

dual function

$$g(\lambda) = \inf_x \left((c + A^T\lambda)^T x - b^T\lambda\right) = \begin{cases} -b^T\lambda & A^T\lambda + c = 0 \\ -\infty & \text{otherwise}\end{cases}$$

dual problem

$$\begin{array}{ll}\text{maximize} & -b^T\lambda \\ \text{subject to} & A^T\lambda + c = 0, \quad \lambda \succeq 0\end{array}$$

- from Slater's condition: $p^\star = d^\star$ if $A\tilde{x} \prec b$ for some $\tilde{x}$
- in fact, $p^\star = d^\star$ except when primal and dual are infeasible (由weak slater条件可知)

Karush-Kuhn-Tucker (KKT)条件

Given general problem

$$\begin{aligned} \min_x \quad & f(x) \\ \text{subject to} \quad & h_i(x) \le 0, \ i = 1, \ldots, m \\ & \ell_j(x) = 0, \ j = 1, \ldots, r \end{aligned}$$

The Karush-Kuhn-Tucker conditions or KKT conditions are:

- $0 \in \partial_x \left(f(x) + \sum_{i=1}^{m} u_i h_i(x) + \sum_{j=1}^{r} v_j \ell_j(x) \right)$ (stationarity)
- $u_i \cdot h_i(x) = 0$ for all i (complementary slackness)
- $h_i(x) \le 0, \ \ell_j(x) = 0$ for all i, j (primal feasibility)
- $u_i \ge 0$ for all i (dual feasibility)

What's in a name?

Older folks will know these as the KT (Kuhn-Tucker) conditions:

- First appeared in publication by Kuhn and Tucker in 1951
- Later people found out that Karush had the conditions in his unpublished master's thesis of 1939

For unconstrained problems, the KKT conditions are nothing more than the subgradient optimality condition

充分性

If there exists $x^\star, u^\star, v^\star$ that satisfy the KKT conditions, then

$$\begin{aligned} g(u^\star, v^\star) &= f(x^\star) + \sum_{i=1}^{m} u_i^\star h_i(x^\star) + \sum_{j=1}^{r} v_j^\star \ell_j(x^\star) \\ &= f(x^\star) \end{aligned}$$

where the first equality holds from stationarity, and the second holds from complementary slackness

Therefore the duality gap is zero (and $x^\star$ and $u^\star, v^\star$ are primal and dual feasible) so $x^\star$ and $u^\star, v^\star$ are primal and dual optimal. Hence, we've shown:

If $x^\star$ and $u^\star, v^\star$ satisfy the KKT conditions, then $x^\star$ and $u^\star, v^\star$ are primal and dual solutions

必要性

Let $x^\star$ and $u^\star, v^\star$ be primal and dual solutions with zero duality gap (strong duality holds, e.g., under Slater's condition). Then

$$
\begin{aligned}
f(x^\star) &= g(u^\star, v^\star) \\
&= \min_x \; f(x) + \sum_{i=1}^{m} u_i^\star h_i(x) + \sum_{j=1}^{r} v_j^\star \ell_j(x) \\
&\le f(x^\star) + \sum_{i=1}^{m} u_i^\star h_i(x^\star) + \sum_{j=1}^{r} v_j^\star \ell_j(x^\star) \\
&\le f(x^\star)
\end{aligned}
$$

In other words, all these inequalities are actually equalities

Two things to learn from this:

- The point $x^\star$ minimizes $L(x, u^\star, v^\star)$ over $x \in \mathbb{R}^n$. Hence the subdifferential of $L(x, u^\star, v^\star)$ must contain 0 at $x = x^\star$—this is exactly the stationarity condition
- We must have $\sum_{i=1}^m u_i^\star h_i(x^\star) = 0$, and since each term here is ≤ 0, this implies $u_i^\star h_i(x^\star) = 0$ for every i—this is exactly complementary slackness

If $x^\star$ and $u^\star, v^\star$ are primal and dual solutions, with zero duality gap, then $x^\star, u^\star, v^\star$ satisfy the KKT conditions

总结

In summary, KKT conditions are equivalent to zero duality gap:

- always sufficient
- necessary under strong duality

Putting it together:

For a problem with strong duality (e.g., assume Slater's condition: convex problem and there exists x strictly satisfying non-affine inequality contraints),

$$x^\star \text{ and } u^\star, v^\star \text{ are primal and dual solutions}$$
$$\iff x^\star \text{ and } u^\star, v^\star \text{ satisfy the KKT conditions}$$

(Warning, concerning the stationarity condition: for a differentiable function f, we cannot use $\partial f(x) = \{\nabla f(x)\}$ unless f is convex!)

课堂作业：SVM的KKT条件

Given $y \in \{-1, 1\}^n$, and $X \in \mathbb{R}^{n \times p}$, the support vector machine problem is:

$$\begin{aligned} \min_{\beta, \beta_0, \xi} \quad & \frac{1}{2}\|\beta\|_2^2 + C \sum_{i=1}^{n} \xi_i \\ \text{subject to} \quad & \xi_i \geq 0, \;\; i = 1, \ldots, n \\ & y_i(x_i^T \beta + \beta_0) \geq 1 - \xi_i, \;\; i = 1, \ldots, n \end{aligned}$$

Thank you for your attentions !