CSE 203B: Convex Optimization
Week 6 Discuss Session

Xinyuan Wang

02/13/2019

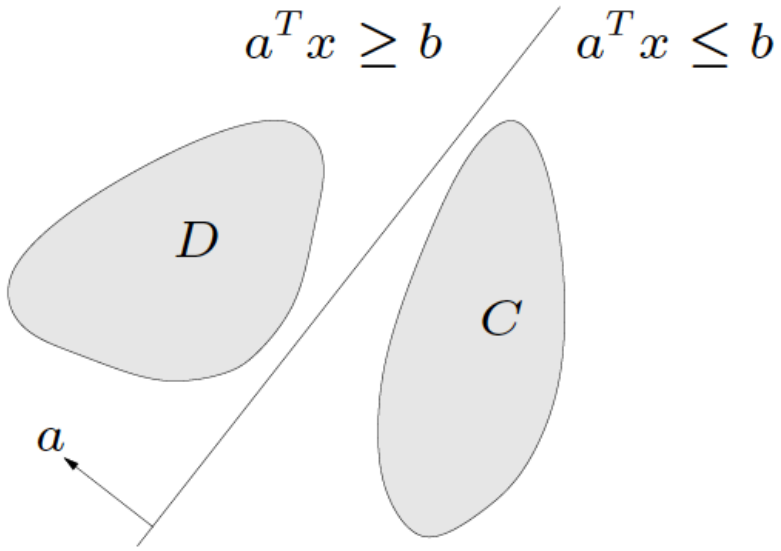
Contents

- Theorems and Proofs
  - Separating hyperplane
  - Establish the convexity
  - Pointwise supremum and Conjugate function
  - Lagrange dual function
  - KKT condition

Separating hyperplane theorem

if C and D are nonempty disjoint convex sets, there exist $a \neq 0$, b s.t.

$$a^T x \leq b \text{ for } x \in C, \qquad a^T x \geq b \text{ for } x \in D$$



the hyperplane $\{x \mid a^T x = b\}$ separates C and D

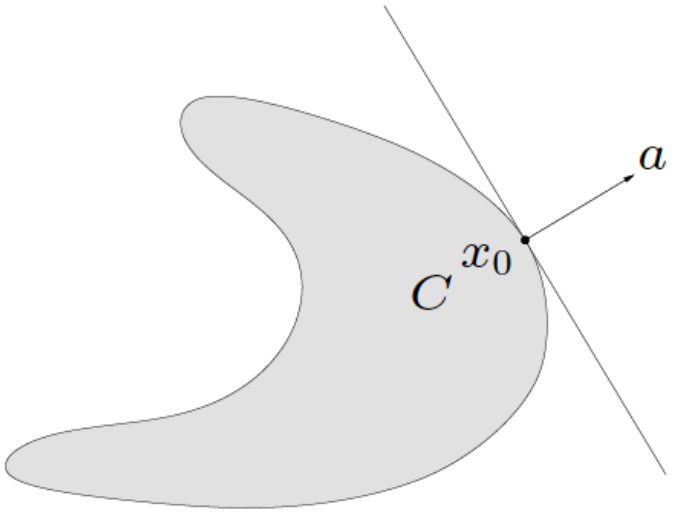
strict separation requires additional assumptions ($e.g.$, C is closed, D is a singleton)

Supporting hyperplane theorem

supporting hyperplane to set C at boundary point x_0:

$$\{x \mid a^T x = a^T x_0\}$$

where $a \neq 0$ and $a^T x \leq a^T x_0$ for all $x \in C$



supporting hyperplane theorem: if C is convex, then there exists a supporting hyperplane at every boundary point of C

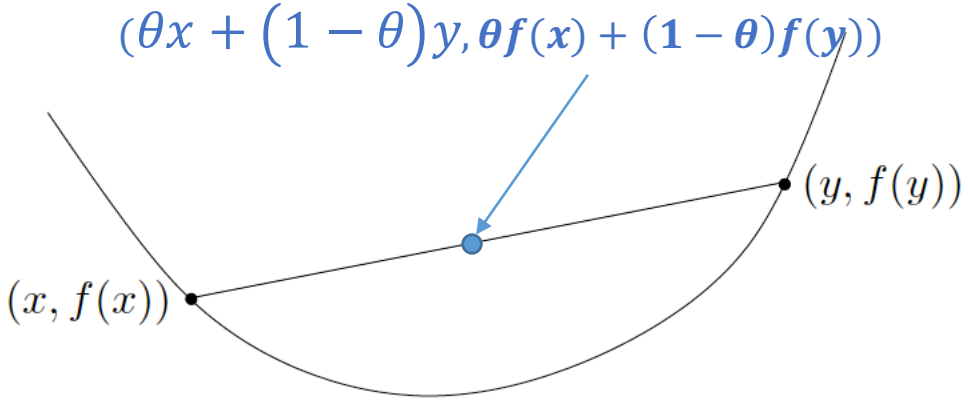
Definition of Convex Functions

- A function $f: R^n \rightarrow R$ is convex if $\mathrm{dom}\ f$ is a convex set and if for all $x, y \in \mathrm{dom}\ f$ and $0 \leq \theta \leq 1$

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

sometimes called Jensen's inequality

- Review the proof in class: necessary and sufficiency
- Strict convexity: $f(\theta x + (1-\theta)y) < \theta f(x) + (1-\theta) f(y), x \neq y, 0 < \theta < 1$
- Concave functions: $-f$ is convex

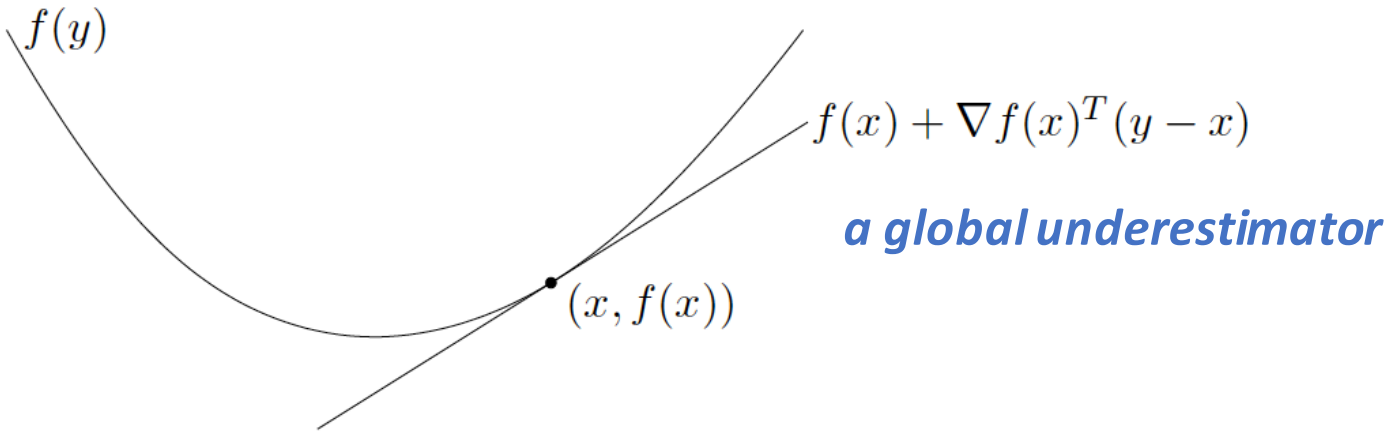


First-order Condition

- Suppose f is differentiable ($dom\ f$ is open and ∇f exists at $\forall x \in dom\ f$), then f is convex iff $\mathrm{dom}\ f$ is convex and for all $x, y \in dom\ f$

$$f(y) \geq f(x) + \nabla f(x)^T (y - x)$$

- Review the proof in class: necessary and sufficiency
- Strict convexity: $f(y) > f(x) + \nabla f(x)^T (y - x), x \neq y$
- Concave functions: $f(y) \leq f(x) + \nabla f(x)^T (y - x)$



a global underestimator

Second Order Condition

- Suppose f is twice differentiable ($dom\ f$ is open and its Hessian exists at $\forall x \in dom\ f$), then f is convex iff $\mathrm{dom}\ f$ is convex and for all $x, y \in dom\ f$

$$\nabla^2 f(x) \succcurlyeq 0 \text{ (positive semidefinite)}$$

- Review the proof in class: necessary and sufficiency
- Strict convexity: $\nabla^2 f(x) \succ 0$
- Concave functions: $\nabla^2 f(x) \preccurlyeq 0$

Operations that preserve convexity

Practical methods for establishing convexity of a function

- Verify definition (often simplified by restricting to a line)
- For twice differentiable functions, show $\nabla^2 f(x) \succcurlyeq 0$
- Show that f is obtained from simple convex functions by operations that preserve convexity (Ref. Chap. 3.2)
  - Nonnegative weighted sum
  - Composition with affine function
  - Pointwise maximum and supremum
  - Composition
  - Minimization
  - Perspective

Relation between convex sets and convex functions

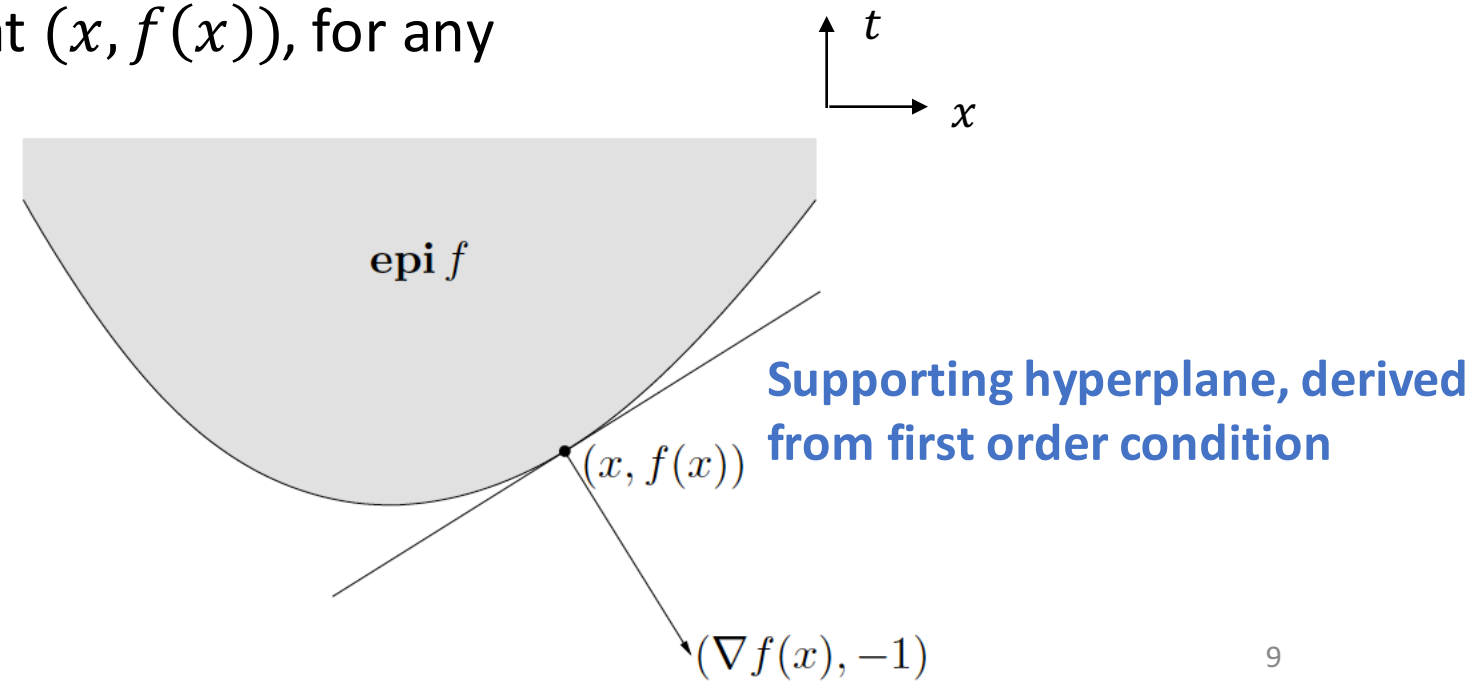
- A function is convex **iff** its epigraph is a convex set.
- Consider a convex function f and $x, y \in dom\ f$

$$t \geq f(y) \geq f(x) + \nabla f(x)^T (y - x)$$

epi f **First order condition for convexity**

- The hyperplane supports **epi f** at $(x, f(x))$, for any

$$(y, t) \in \mathbf{epi}\ f \Rightarrow$$
$$\nabla f(x)^T (y - x) + f(x) - t \leq 0$$
$$\Rightarrow \begin{bmatrix} \nabla f(x) \\ -1 \end{bmatrix}^T \left(\begin{bmatrix} y \\ t \end{bmatrix} - \begin{bmatrix} x \\ f(x) \end{bmatrix} \right) \leq 0$$



Supporting hyperplane, derived from first order condition

Pointwise Supremum

- If for each $y \in U, f(x, y): R^n \to R$ is convex in x, then function

$$g(x) = \sup_{y \in U} f(x, y)$$

  is convex in x.

- **Proof with definition of convex functions.**

- Epigraph of $g(x)$ is the intersection of epigraphs with f and set U

$$\mathbf{epi}\ g = \cap_{y \in U}\ \mathbf{epi}\ f(\cdot, y)$$

  knowing $\mathbf{epi}\ f(\cdot, y)$ is a convex set (f is a convex function in x and regard y as a const), so $\mathbf{epi}\ g$ is convex.

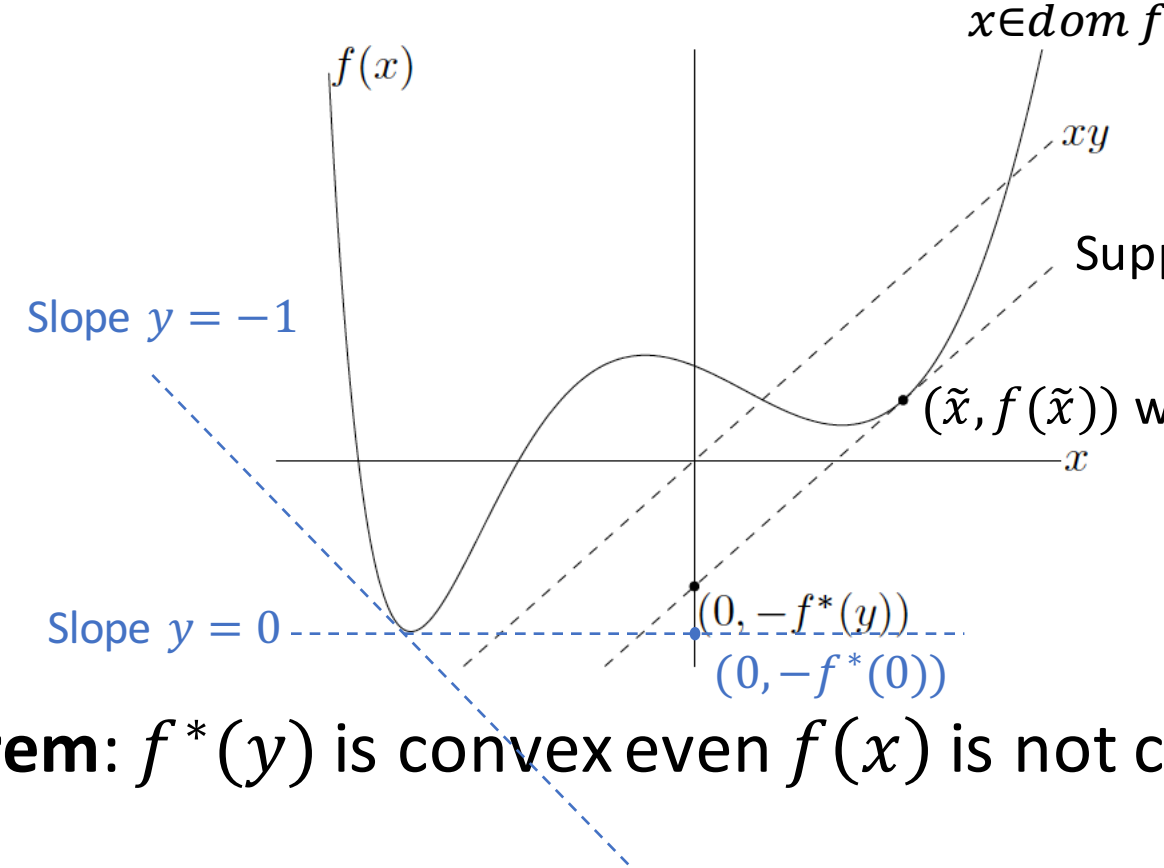
- An interesting method to establish convexity of a function: the pointwise supremum of a family of affine functions. (ref. Chap 3.2.4)

Conjugate Function

- Given function $f: R^n \rightarrow R$, the conjugate function

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

  - The dom f^* consists $y \in R^n$ for which $\sup_{x \in dom\, f} \; y^T x \; - f(x)$ is bounded



Supporting hyperplane: if f is convex in that domain

$(\tilde{x}, f(\tilde{x}))$ where $\nabla f^T(\tilde{x}) = y$ if f is differentiable

Theorem: $f^*(y)$ is convex even $f(x)$ is not convex.

(Proof with pointwise supremum)

$y^T x \; - f(x)$ is affine function in y

Duality

standard form problem (not necessarily convex)

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

variable $x \in \mathbf{R}^n$, domain $\mathcal{D}$, optimal value $p^\star$

Lagrangian: $L : \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R}^p \to \mathbf{R}$, with $\mathbf{dom}\, L = \mathcal{D} \times \mathbf{R}^m \times \mathbf{R}^p$,

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

Lagrange dual function: $g : \mathbf{R}^m \times \mathbf{R}^p \to \mathbf{R}$,

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu)$$

Duality

- The Lagrange dual function is concave
- The dual function yields lower bounds on optimal value

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

Prove the properties.

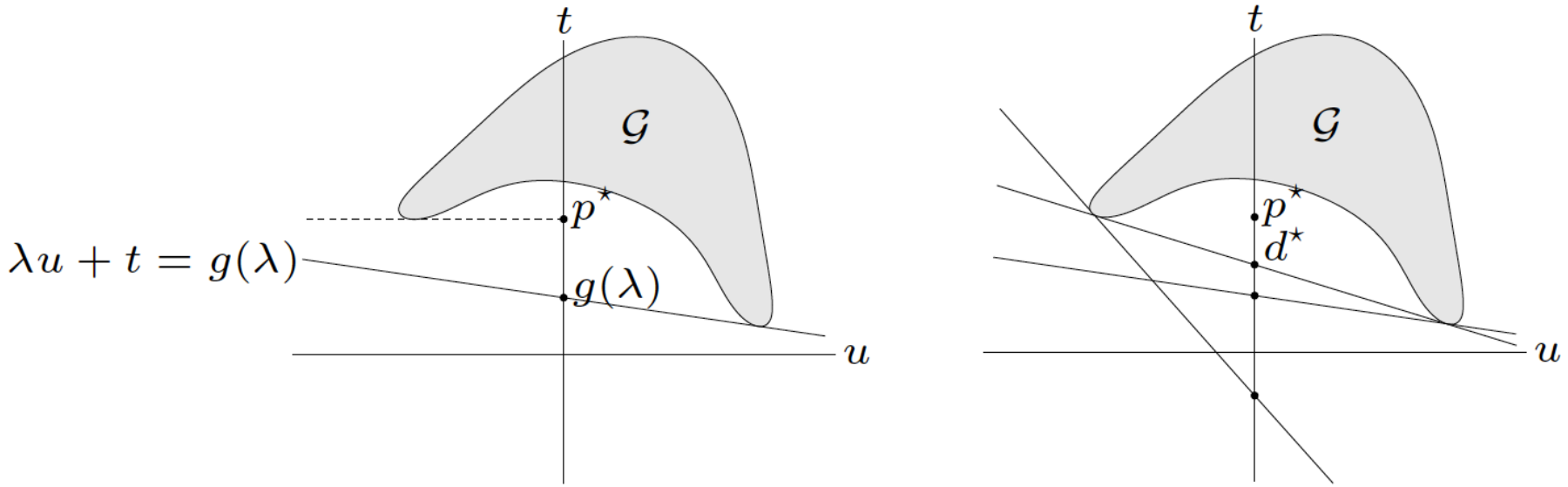
Lagrange dual problem

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

- 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 $\lambda \succeq 0$, $(\lambda, \nu) \in \mathbf{dom}\, g$
- often simplified by making implicit constraint $(\lambda, \nu) \in \mathbf{dom}\, g$ explicit

Geometric interpretation

- Let :$\mathcal{G} = \{(f_1(x), \ldots, f_m(x), h_1(x), \ldots, h_p(x), f_0(x)) \in R^m \times R^p \times R | \; x \in \mathcal{D}\}$
  - Optimal value $p^* = \inf\{t | (u, v, t) \in \mathcal{G}, u \leq 0, v = 0\}$
  - The dual function $g(\lambda, \nu) = \inf\{(\lambda, \nu, 1)^T (u, v, t) \; | (u, v, t) \in \mathcal{G}\}$
  - Lower bound property holds: $p^* \geq g(\lambda, \nu)$



- $\lambda u + t = g(\lambda)$ is (non-vertical) supporting hyperplane to $\mathcal{G}$
- hyperplane intersects t-axis at $t = g(\lambda)$

Saddle point interpretation

- *Max-min inequality*: for any $f: R^n \times R^m \to R$

$$\sup_{z \in Z} \inf_{w \in W} f(w,z) \leq \inf_{w \in W} \sup_{z \in Z} f(w,z)$$

Consider the optimal values of primal problem and dual problem (no equality constraint),

$$p^* = \inf_x \sup_{\lambda \succcurlyeq 0} L(x,\lambda), \qquad d^* = \sup_{\lambda \succcurlyeq 0} \inf_x L(x,\lambda)$$

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

- *Strong max-min property* (saddle-point property)

$$\sup_{z \in Z} \inf_{w \in W} f(w,z) = \inf_{w \in W} \sup_{z \in Z} f(w,z)$$

- Saddle-point $\tilde{w} \in W, \tilde{z} \in Z$ for f if

$$f(\tilde{w}, z) \leq f(\tilde{w}, \tilde{z}) \leq f(w, \tilde{z})$$

for all $w \in W, z \in Z$. $\Leftrightarrow$ strong max-min holds.

KKT condition

the following four conditions are called KKT conditions (for a problem with differentiable f_i, h_i):

1. primal constraints: $f_i(x) \leq 0,\ i = 1, \ldots, m,\ h_i(x) = 0,\ i = 1, \ldots, p$
2. dual constraints: $\lambda \succeq 0$
3. complementary slackness: $\lambda_i f_i(x) = 0,\ i = 1, \ldots, m$
4. gradient of Lagrangian with respect to x vanishes:

$$\nabla f_0(x) + \sum_{i=1}^{m} \lambda_i \nabla f_i(x) + \sum_{i=1}^{p} \nu_i \nabla h_i(x) = 0$$

- **Any** optimization problem with differential $f_0, \ldots, f_m$ for which strong duality obtains $p^* = d^*$, any pair of primal (x^*) and dual optimal points (λ^*, ν^*) must satisfy the KKT conditions.

Contents

- Case Studies (see the exercises and midterm sample)
  - Supporting hyperplane
  - Dual cone
  - Conjugate function
  - Primal dual formulation
- Review the homework exercises
  - HW1, 2 and 3

Primal and Dual formulation of a LP problem

- Consider a linear programming problem

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to} \quad & Ax \preccurlyeq 0 \end{aligned}$$

  - The feasible x is in the cone $K = \{x | Ax \preccurlyeq 0\}$. The optimal condition $\nabla f_0(x)^T (y - x) \geq 0$ should holds for all $y \in K$. So the gradient $\nabla f_0(x) = c$ falls in the dual cone $K^* = \{-A^T y | y \geq 0\}$,

$$\Rightarrow A^T y + c = 0, \qquad \exists y \succcurlyeq 0$$

  - The Lagrange dual function of the primal problem is

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

  The dual problem is

$$\begin{aligned} \max \quad & 0 \\ \text{subject to} \quad & A^T \lambda + c = 0 \\ & \lambda \geq 0 \end{aligned}$$

- ***Farkas' lemma***: the system of inequalities

$$Ax \preccurlyeq 0, \qquad c^T x < 0$$

where $A \in R^{m \times n}, c \in R^n$, and the system of equalities ad inequalities

$$A^T y + c = 0, \qquad y \succcurlyeq 0$$

are strong alternatives.