Convex Optimization

(EE227A: UC Berkeley)

Lecture 11
(Duality, minimax, optimality conditions)

26 Feb, 2013

Suvrit Sra

Organizational

- Project team lists due by end of Feb
- Project suggestions out in a few days
  - Purely theoretical projects
  - Algorithms for particular problem classes
  - Application centric (engg., sig. proc., ML, etc.)
  - Systems centric (software, distributed, parallel algos)
- Initial proposal by 14th March
- Project midpoint review: 16th April
- Project **final paper**, presentations: Finals week
- **Midterm: 21st March (1.5 hours, in class)**
- Email me any concerns, doubts, questions, feedback

Recap

- $\mathcal{L}(x, \lambda, \nu) = f(x) + \sum_i \lambda_i f_i(x) + \sum_i \nu_i h_i(x)$
- $g(\lambda, \nu) := \inf_x \mathcal{L}(x, \lambda, \nu)$
- $d^* := \sup g(\lambda, \nu) \leq p^* := \inf_x f(x)$ s.t. $x \in \mathcal{X}$ (weak duality)
- Slater's constraint qualification ensures $d^* = p^*$ (strong duality)

Example: regularized optimization

$$\inf_{x\in\mathcal{X}} \quad f(x) + r(Ax) \quad \text{s.t.} \quad Ax \in \mathcal{Y}.$$

Example: regularized optimization

$$\inf_{x \in \mathcal{X}} \quad f(x) + r(Ax) \quad \text{s.t.} \quad Ax \in \mathcal{Y}.$$

Dual problem

$$\inf_{u \in \mathcal{Y}} \quad f^*(-A^T u) + r^*(u).$$

Example: regularized optimization

$$\inf_{x \in \mathcal{X}} \quad f(x) + r(Ax) \quad \text{s.t.} \quad Ax \in \mathcal{Y}.$$

Dual problem

$$\inf_{u \in \mathcal{Y}} \quad f^*(-A^T u) + r^*(u).$$

- Introduce new variable $z = Ax$

$$\inf_{x \in \mathcal{X}, z \in \mathcal{Y}} \quad f(x) + r(z), \qquad \text{s.t.} \quad z = Ax.$$

Example: regularized optimization

$$\inf_{x \in \mathcal{X}} \quad f(x) + r(Ax) \quad \text{s.t.} \quad Ax \in \mathcal{Y}.$$

Dual problem

$$\inf_{u \in \mathcal{Y}} \quad f^*(-A^T u) + r^*(u).$$

- Introduce new variable $z = Ax$

$$\inf_{x \in \mathcal{X}, z \in \mathcal{Y}} \quad f(x) + r(z), \qquad \text{s.t.} \quad z = Ax.$$

- The (partial)-Lagrangian is

$$L(x, z; u) := f(x) + r(z) + u^T(Ax - z), \quad x \in \mathcal{X}, z \in \mathcal{Y};$$

Example: regularized optimization

$$\inf_{x \in \mathcal{X}} \quad f(x) + r(Ax) \quad \text{s.t.} \quad Ax \in \mathcal{Y}.$$

Dual problem

$$\inf_{u \in \mathcal{Y}} \quad f^*(-A^T u) + r^*(u).$$

- Introduce new variable $z = Ax$

$$\inf_{x \in \mathcal{X}, z \in \mathcal{Y}} \quad f(x) + r(z), \qquad \text{s.t.} \quad z = Ax.$$

- The (partial)-Lagrangian is

$$L(x, z; u) := f(x) + r(z) + u^T(Ax - z), \quad x \in \mathcal{X}, z \in \mathcal{Y};$$

- Associated dual function

$$g(u) := \inf_{x \in \mathcal{X}, z \in \mathcal{Y}} L(x, z; u).$$

Regularized optimization

$$\inf_{x \in \mathcal{X}} \quad f(x) + r(Ax) \quad \text{s.t.} \quad Ax \in \mathcal{Y}.$$

Dual problem

$$\inf_{y \in \mathcal{Y}} \quad f^*(-A^T y) + r^*(y).$$

The infimum above can be rearranged as follows

$$g(y) \quad = \quad \inf_{x \in \mathcal{X}} \; f(x) + y^T Ax + \inf_{z \in \mathcal{Y}} \; r(z) - y^T z$$

Regularized optimization

$$\inf_{x \in \mathcal{X}} \quad f(x) + r(Ax) \quad \text{s.t.} \quad Ax \in \mathcal{Y}.$$

Dual problem

$$\inf_{y \in \mathcal{Y}} \quad f^*(-A^T y) + r^*(y).$$

The infimum above can be rearranged as follows

$$\begin{aligned} g(y) &= \inf_{x \in \mathcal{X}} f(x) + y^T Ax + \inf_{z \in \mathcal{Y}} r(z) - y^T z \\ &= -\sup_{x \in \mathcal{X}} \left\{ -x^T A^T y - f(x) \right\} - \sup_{z \in \mathcal{Y}} \left\{ z^T y - r(z) \right\} \end{aligned}$$

Regularized optimization

$$\inf_{x \in \mathcal{X}} \quad f(x) + r(Ax) \quad \text{s.t.} \quad Ax \in \mathcal{Y}.$$

Dual problem

$$\inf_{y \in \mathcal{Y}} \quad f^*(-A^T y) + r^*(y).$$

The infimum above can be rearranged as follows

$$\begin{aligned} g(y) &= \inf_{x \in \mathcal{X}} f(x) + y^T Ax + \inf_{z \in \mathcal{Y}} r(z) - y^T z \\ &= -\sup_{x \in \mathcal{X}} \left\{ -x^T A^T y - f(x) \right\} - \sup_{z \in \mathcal{Y}} \left\{ z^T y - r(z) \right\} \\ &= -f^*(-A^T y) - r^*(y) \quad \text{s.t. } y \in \mathcal{Y}. \end{aligned}$$

Regularized optimization

$$\inf_{x\in\mathcal{X}} \quad f(x) + r(Ax) \quad \text{s.t.} \quad Ax \in \mathcal{Y}.$$

Dual problem

$$\inf_{y\in\mathcal{Y}} \quad f^*(-A^Ty) + r^*(y).$$

The infimum above can be rearranged as follows

$$\begin{aligned} g(y) &= \inf_{x\in\mathcal{X}} f(x) + y^TAx + \inf_{z\in\mathcal{Y}} r(z) - y^Tz \\ &= -\sup_{x\in\mathcal{X}} \left\{-x^TA^Ty - f(x)\right\} - \sup_{z\in\mathcal{Y}} \left\{z^Ty - r(z)\right\} \\ &= -f^*(-A^Ty) - r^*(y) \quad \text{s.t. } y \in \mathcal{Y}. \end{aligned}$$

Dual problem computes $\sup_{u\in\mathcal{Y}} g(u)$; so equivalently,

$$\inf_{y\in\mathcal{Y}} \quad f^*(-A^Ty) + r^*(y).$$

Regularized optimization

Strong duality

$$\inf_x \{f(x) + r(Ax)\} = \sup_y \left\{-f^*(-A^T y) + r^*(y)\right\}$$

Regularized optimization

Strong duality

$$\inf_x \{f(x) + r(Ax)\} = \sup_y \left\{-f^*(-A^T y) + r^*(y)\right\}$$

- 'sup' attained at some y, and
- 'inf' attained at some x

Regularized optimization

Strong duality

$$\inf_x \{f(x) + r(Ax)\} = \sup_y \left\{-f^*(-A^T y) + r^*(y)\right\}$$

- 'sup' attained at some y, and
- 'inf' attained at some x

Ensured, if either of the following conditions holds:

- $\exists x \in \mathrm{ri}(\mathrm{dom}\, f)$ such that $Ax \in \mathrm{ri}(\mathrm{dom}\, r)$
- $\exists y \in \mathrm{ri}(\mathrm{dom}\, r^*)$ such that $A^T y \in \mathrm{ri}(\mathrm{dom}\, f^*)$

Example: norm regularized problems

$$\min \quad f(x) + \|Ax\|$$

Example: norm regularized problems

$$\min \quad f(x) + \|Ax\|$$

Dual problem

$$\min_{y} \quad f^*(-A^T y) \quad \text{s.t. } \|y\|_* \leq 1.$$

Example: norm regularized problems

$$\min \quad f(x) + \|Ax\|$$

Dual problem

$$\min_y \quad f^*(-A^T y) \quad \text{s.t. } \|y\|_* \leq 1.$$

Say $\|\bar{y}\|_* < 1$, such that $A^T\bar{y} \in \mathrm{ri}(\mathrm{dom}\, f^*)$, then we have strong duality (e.g., for instance $0 \in \mathrm{ri}(\mathrm{dom}\, f^*)$)

Dual via Fenchel conjugates

$\min f(x) \quad \text{s.t. } f_i(x) \leq 0, Ax = b.$

$$\mathcal{L}(x, \lambda, \nu) \quad := \quad f_0(x) + \sum_i \lambda_i f_i(x) + \nu^T(Ax - b)$$

Dual via Fenchel conjugates

$\min f(x)$ s.t. $f_i(x) \leq 0, Ax = b$.

$$\begin{aligned} \mathcal{L}(x, \lambda, \nu) &:= f_0(x) + \sum_i \lambda_i f_i(x) + \nu^T (Ax - b) \\ g(\lambda, \nu) &= \inf_x \mathcal{L}(x, \lambda, \nu) \end{aligned}$$

Dual via Fenchel conjugates

$\min f(x)$ s.t. $f_i(x) \leq 0, Ax = b.$

$$\begin{aligned}
\mathcal{L}(x,\lambda,\nu) &:= f_0(x) + \sum_i \lambda_i f_i(x) + \nu^T(Ax - b)\\
g(\lambda,\nu) &= \inf_x \mathcal{L}(x,\lambda,\nu)\\
g(\lambda,\nu) &= -\nu^T b + \inf_x x^T A^T \nu + F(x)
\end{aligned}$$

Dual via Fenchel conjugates

$\min f(x) \quad \text{s.t. } f_i(x) \leq 0, Ax = b.$

$$\begin{aligned}
\mathcal{L}(x, \lambda, \nu) &:= f_0(x) + \sum_i \lambda_i f_i(x) + \nu^T(Ax - b) \\
g(\lambda, \nu) &= \inf_x \mathcal{L}(x, \lambda, \nu) \\
g(\lambda, \nu) &= -\nu^T b + \inf_x x^T A^T \nu + F(x) \\
F(x) &:= f_0(x) + \sum_i \lambda_i f_i(x)
\end{aligned}$$

Dual via Fenchel conjugates

$\min f(x)$ s.t. $f_i(x) \leq 0, Ax = b.$

$$
\begin{aligned}
\mathcal{L}(x, \lambda, \nu) &:= f_0(x) + \sum_i \lambda_i f_i(x) + \nu^T (Ax - b) \\
g(\lambda, \nu) &= \inf_x \mathcal{L}(x, \lambda, \nu) \\
g(\lambda, \nu) &= -\nu^T b + \inf_x x^T A^T \nu + F(x) \\
F(x) &:= f_0(x) + \sum_i \lambda_i f_i(x) \\
g(\lambda, \nu) &= -\nu^T b - \sup_x \langle x, -A^T \nu \rangle - F(x)
\end{aligned}
$$

Dual via Fenchel conjugates

$\min f(x)$ s.t. $f_i(x) \le 0, Ax = b.$

$$
\begin{aligned}
\mathcal{L}(x,\lambda,\nu) &:= f_0(x) + \sum_i \lambda_i f_i(x) + \nu^T(Ax - b) \\
g(\lambda,\nu) &= \inf_x \mathcal{L}(x,\lambda,\nu) \\
g(\lambda,\nu) &= -\nu^T b + \inf_x x^T A^T \nu + F(x) \\
F(x) &:= f_0(x) + \sum_i \lambda_i f_i(x) \\
g(\lambda,\nu) &= -\nu^T b - \sup_x \langle x, -A^T\nu \rangle - F(x) \\
g(\lambda,\nu) &= -\nu^T b - F^*(-A^T\nu).
\end{aligned}
$$

Not so useful! F^* hard to compute.

Dual via Fenchel conjugates

Introduce new variables!

Dual via Fenchel conjugates

Introduce new variables!

$$\min f(x) \quad \text{s.t.} \quad f_i(x_i) \leq 0, Ax = b$$
$$x_i = z, i = 1, \ldots, m.$$

Dual via Fenchel conjugates

Introduce new variables!

$$\min f(x) \quad \text{s.t.} \quad f_i(x_i) \leq 0, Ax = b$$
$$x_i = z, i = 1, \ldots, m.$$

$$\mathcal{L}(x, x_i z, \lambda, \nu, \pi_i)$$
$$:= f(x) + \sum\nolimits_i \lambda_i f_i(x_i) + \nu^T(Ax - b) + \sum\nolimits_i \pi_i^T(x_i - z)$$

Dual via Fenchel conjugates

Introduce new variables!

$$\min f(x) \quad \text{s.t.} \quad f_i(x_i) \leq 0, Ax = b$$
$$x_i = z, i = 1, \ldots, m.$$

$$\mathcal{L}(x, x_i z, \lambda, \nu, \pi_i)$$
$$:= f(x) + \sum\nolimits_i \lambda_i f_i(x_i) + \nu^T(Ax - b) + \sum\nolimits_i \pi_i^T(x_i - z)$$
$$g(\lambda, \nu, \pi_i) = \inf_{x, x_i, z} \mathcal{L}(x, x_i, z, \lambda, \nu, \pi_i)$$

Dual via Fenchel conjugates

Introduce new variables!

$$\min f(x) \quad \text{s.t.} \qquad f_i(x_i) \leq 0, Ax = b$$
$$x_i = z, i = 1, \ldots, m.$$

$$\mathcal{L}(x, x_i z, \lambda, \nu, \pi_i)$$
$$:= f(x) + \sum\nolimits_i \lambda_i f_i(x_i) + \nu^T(Ax - b) + \sum\nolimits_i \pi_i^T(x_i - z)$$
$$g(\lambda, \nu, \pi_i) = \inf_{x, x_i, z} \mathcal{L}(x, x_i, z, \lambda, \nu, \pi_i)$$
$$= -\nu^T b + \inf_x f(x) + \nu^T Ax + \inf_z \sum\nolimits_i -\pi_i^T z$$
$$+ \sum\nolimits_i \inf_{x_i} \pi_i^T x_i + \lambda_i f_i(x_i)$$

Dual via Fenchel conjugates

Introduce new variables!

$$\min f(x) \quad \text{s.t.} \qquad f_i(x_i) \leq 0, Ax = b$$
$$x_i = z, i = 1, \ldots, m.$$

$$\mathcal{L}(x, x_i z, \lambda, \nu, \pi_i)$$
$$:= f(x) + \sum\nolimits_i \lambda_i f_i(x_i) + \nu^T(Ax - b) + \sum\nolimits_i \pi_i^T(x_i - z)$$
$$g(\lambda, \nu, \pi_i) = \inf_{x, x_i, z} \mathcal{L}(x, x_i, z, \lambda, \nu, \pi_i)$$
$$= -\nu^T b + \inf_x f(x) + \nu^T Ax + \inf_z \sum\nolimits_i -\pi_i^T z$$
$$+ \sum\nolimits_i \inf_{x_i} \pi_i^T x_i + \lambda_i f_i(x_i)$$
$$= \begin{cases} -\nu^T b - f^*(-A^T\nu) - \sum_i (\lambda_i f_i)^*(-\pi_i) & \text{if } \sum_i \pi_i = 0 \\ -\infty & \text{otherwise.} \end{cases}$$

Example

Exercise: Derive the Lagrangian dual in terms of Fenchel conjugates for the following linearly constrained problem:

$$\min \quad f(x) \quad \text{s.t. } Ax \le b, \quad Cx = d.$$

Hint: No need to introduce extra variables.

Example: variable splitting

$$\min \quad f(x) + g(x)$$

Example: variable splitting

$$\min \quad f(x) + g(x)$$

Exercise: Fill in the details for the following steps

$$\min_{x,z} \quad f(x) + g(z) \quad \text{s.t.} \quad x = z$$

Example: variable splitting

$$\min \quad f(x) + g(x)$$

Exercise: Fill in the details for the following steps

$$\min_{x,z} \quad f(x) + g(z) \quad \text{s.t.} \quad x = z$$

$$L(x, z, \nu) = f(x) + g(z) + \nu^T(x - z)$$

Example: variable splitting

$$\min \quad f(x) + g(x)$$

Exercise: Fill in the details for the following steps

$$\min_{x,z} \quad f(x) + g(z) \quad \text{s.t.} \quad x = z$$

$$L(x, z, \nu) = f(x) + g(z) + \nu^T (x - z)$$

$$g(\nu) = \inf_{x,z} L(x, z, \nu)$$

Conic duality

LP Duality

- Consider linear program

$$\min \quad c^T x \qquad Ax \leq b.$$

LP Duality

- Consider linear program

$$\min \quad c^T x \qquad Ax \leq b.$$

- Corresponding dual is

$$\max \quad b^T \lambda \qquad A^T \lambda + c = 0, \quad \lambda \geq 0.$$

LP Duality

- Consider linear program

$$\min \quad c^T x \qquad Ax \le b.$$

- Corresponding dual is

$$\max \quad b^T \lambda \qquad A^T \lambda + c = 0, \quad \lambda \ge 0.$$

- LP duality facts:
  - If either p^* or d^* finite, then $p^* = d^*$, and both primal, dual problem have optimal solutions
  - If $p^* = -\infty$, then $d^* = -\infty$ (follows from weak-duality)
  - If $d^* = \infty$, then $p^* = \infty$ (again, weak-duality)

LP Duality

- Consider linear program

$$\min \quad c^T x \qquad Ax \le b.$$

- Corresponding dual is

$$\max \quad b^T \lambda \qquad A^T\lambda + c = 0, \quad \lambda \ge 0.$$

- LP duality facts:
  - If either p^* or d^* finite, then $p^* = d^*$, and both primal, dual problem have optimal solutions
  - If $p^* = -\infty$, then $d^* = -\infty$ (follows from weak-duality)
  - If $d^* = \infty$, then $p^* = \infty$ (again, weak-duality)

Proof: See lecture notes.

LP Duality

- Consider linear program

$$\min \quad c^T x \qquad Ax \leq b.$$

- Corresponding dual is

$$\max \quad b^T \lambda \qquad A^T \lambda + c = 0, \quad \lambda \geq 0.$$

- LP duality facts:
  - If either p^* or d^* finite, then $p^* = d^*$, and both primal, dual problem have optimal solutions
  - If $p^* = -\infty$, then $d^* = -\infty$ (follows from weak-duality)
  - If $d^* = \infty$, then $p^* = \infty$ (again, weak-duality)

Proof: See lecture notes.

If LP is feasible, strong duality holds.

SOCP Duality

- Consider SOCP

$$\min \quad f^T x \qquad \|A_i x + b_i\|_2 \leq c_i^T x + d_i, \quad i = 1, \ldots, m.$$

SOCP Duality

- Consider SOCP

$$\min \quad f^T x \qquad \|A_i x + b_i\|_2 \leq c_i^T x + d_i, \quad i = 1, \ldots, m.$$

- Lagrangian (ordinary)

$$\mathcal{L}(x, \lambda) := f^T x + \sum\nolimits_i \lambda_i (\|A_i x + b_i\|_2 - c_i^T x + d_i)$$

SOCP Duality

- Consider SOCP

$$\min \quad f^T x \qquad \|A_i x + b_i\|_2 \leq c_i^T x + d_i, \quad i = 1, \ldots, m.$$

- Lagrangian (ordinary)

$$\mathcal{L}(x, \lambda) := f^T x + \sum_i \lambda_i (\|A_i x + b_i\|_2 - c_i^T x + d_i)$$

- Recall that $\|x\|_2 = \sup \left\{ u^T x \mid \|u\|_2 \leq 1 \right\}$.

$$\lambda_i \|A_i x + b_i\|_2 \quad = \quad \max_{u_i} \; (\lambda_i u_i)^T (A_i x + b_i) \qquad \|u_i\|_2 \leq 1$$

SOCP Duality

- Consider SOCP

$$\min \quad f^T x \qquad \|A_i x + b_i\|_2 \leq c_i^T x + d_i, \quad i = 1, \ldots, m.$$

- Lagrangian (ordinary)

$$\mathcal{L}(x, \lambda) := f^T x + \sum_i \lambda_i (\|A_i x + b_i\|_2 - c_i^T x + d_i)$$

- Recall that $\|x\|_2 = \sup \left\{ u^T x \mid \|u\|_2 \leq 1 \right\}$.

$$\begin{aligned} \lambda_i \|A_i x + b_i\|_2 &= \max_{u_i} \ (\lambda_i u_i)^T (A_i x + b_i) && \|u_i\|_2 \leq 1 \\ &= \max_{v_i} \ v_i^T (A_i x + b_i) && \|v_i\|_2 \leq \lambda_i \end{aligned}$$

SOCP Duality

- Consider SOCP
$$\min \quad f^T x \qquad \|A_i x + b_i\|_2 \leq c_i^T x + d_i, \quad i = 1, \ldots, m.$$
- Lagrangian (ordinary)
$$\mathcal{L}(x, \lambda) := f^T x + \sum_i \lambda_i (\|A_i x + b_i\|_2 - c_i^T x + d_i)$$
- Recall that $\|x\|_2 = \sup \left\{ u^T x \mid \|u\|_2 \leq 1 \right\}$.
$$\begin{aligned} \lambda_i \|A_i x + b_i\|_2 &= \max_{u_i} \ (\lambda_i u_i)^T (A_i x + b_i) \qquad \|u_i\|_2 \leq 1 \\ &= \max_{v_i} \ v_i^T (A_i x + b_i) \qquad \|v_i\|_2 \leq \lambda_i \end{aligned}$$
- Thus, with $v_1, \ldots, v_m$ also as dual variables we have

SOCP Duality

- Consider SOCP

$$\min \quad f^T x \qquad \|A_i x + b_i\|_2 \le c_i^T x + d_i, \quad i = 1, \ldots, m.$$

- Lagrangian (ordinary)

$$\mathcal{L}(x, \lambda) := f^T x + \sum_i \lambda_i (\|A_i x + b_i\|_2 - c_i^T x + d_i)$$

- Recall that $\|x\|_2 = \sup \left\{ u^T x \mid \|u\|_2 \le 1 \right\}$.

$$\begin{aligned} \lambda_i \|A_i x + b_i\|_2 &= \max_{u_i} \; (\lambda_i u_i)^T (A_i x + b_i) \qquad \|u_i\|_2 \le 1 \\ &= \max_{v_i} \; v_i^T (A_i x + b_i) \qquad \|v_i\|_2 \le \lambda_i \end{aligned}$$

- Thus, with $v_1, \ldots, v_m$ also as dual variables we have

$$p^* = \inf_x \sup_{\lambda, v_1, \ldots, v_m} \; f^T x + \sum_i v_i^T (A_i x + b_i) - \sum_i \lambda_i (c_i^T x + d_i)$$

$$\text{s.t.} \quad \|v_i\|_2 \le \lambda_i, \quad i = 1, \ldots, m.$$

SOCP Duality

- The dual problem is

$$d^* = \sup_{\lambda, v_1, \ldots, v_m} \inf_x \; f^T x + \sum_i v_i^T (A_i x + b_i) - \sum_i \lambda_i (c_i^T x + d_i)$$

$$\text{s.t.} \quad \|v_i\|_2 \leq \lambda_i, \quad i = 1, \ldots, m.$$

SOCP Duality

- The dual problem is

$$d^* = \sup_{\lambda, v_1, \ldots, v_m} \inf_x f^T x + \sum_i v_i^T (A_i x + b_i) - \sum_i \lambda_i (c_i^T x + d_i)$$
$$\text{s.t.} \quad \|v_i\|_2 \le \lambda_i, \quad i = 1, \ldots, m.$$

- Inner minimization over x very easy (unconstrained)

SOCP Duality

- The dual problem is

$$d^* = \sup_{\lambda, v_1, \ldots, v_m} \inf_x f^T x + \sum\nolimits_i v_i^T (A_i x + b_i) - \sum\nolimits_i \lambda_i (c_i^T x + d_i)$$

$$\text{s.t.} \quad \|v_i\|_2 \le \lambda_i, \quad i = 1, \ldots, m.$$

- Inner minimization over x very easy (unconstrained)
- $f + \sum_i A_i^T v_i - \lambda_i c_i = 0$

SOCP Duality

- The dual problem is

$$d^* = \sup_{\lambda, v_1, \ldots, v_m} \inf_x f^T x + \sum_i v_i^T (A_i x + b_i) - \sum_i \lambda_i (c_i^T x + d_i)$$
$$\text{s.t.} \quad \|v_i\|_2 \le \lambda_i, \quad i = 1, \ldots, m.$$

- Inner minimization over x very easy (unconstrained)
- $f + \sum_i A_i^T v_i - \lambda_i c_i = 0$
- Dual problem becomes

$$d^* = \sup_{\lambda, v_1, \ldots, v_m} -\lambda^T d + \sum_i v_i^T b_i$$
$$\text{s.t.} \quad f + \sum_i A_i^T v_i - \lambda_i c_i = 0, \quad \|v_i\|_2 \le \lambda_i, \quad i = 1, \ldots, m.$$

- Also an SOCP, like the primal

SOCP Duality

- The dual problem is

$$d^* = \sup_{\lambda, v_1, \ldots, v_m} \inf_x f^T x + \sum_i v_i^T (A_i x + b_i) - \sum_i \lambda_i (c_i^T x + d_i)$$
$$\text{s.t.} \quad \|v_i\|_2 \le \lambda_i, \quad i = 1, \ldots, m.$$

- Inner minimization over x very easy (unconstrained)
- $f + \sum_i A_i^T v_i - \lambda_i c_i = 0$
- Dual problem becomes

$$d^* = \sup_{\lambda, v_1, \ldots, v_m} -\lambda^T d + \sum_i v_i^T b_i$$
$$\text{s.t.} \quad f + \sum_i A_i^T v_i - \lambda_i c_i = 0, \quad \|v_i\|_2 \le \lambda_i, \quad i = 1, \ldots, m.$$

- Also an SOCP, like the primal
- Apply Slater to obtain a condition for strong duality.

SDP duality

- SDP primal form

$$p^* := \min \operatorname{Tr}(CX), \quad \text{s.t. } \operatorname{Tr}(A_i X) = b_i, \quad i = 1, \ldots, m, \quad X \succeq 0.$$

SDP duality

- SDP primal form

$$p^* := \min \mathrm{Tr}(CX), \quad \text{s.t. } \mathrm{Tr}(A_i X) = b_i, \quad i = 1, \ldots, m, \quad X \succeq 0.$$

- How to handle the matrix constraint $X \succeq 0$?

SDP duality

- SDP primal form

$$p^* := \min \operatorname{Tr}(CX), \quad \text{s.t. } \operatorname{Tr}(A_i X) = b_i, \quad i = 1, \ldots, m, \quad X \succeq 0.$$

- How to handle the matrix constraint $X \succeq 0$?
- Introduce **conic Lagrangian**

$$\mathcal{L}(X, \nu, Y) := \operatorname{Tr}(CX) + \sum_i \nu_i (\operatorname{Tr}(A_i X) - b_i) - \operatorname{Tr}(YX)$$

where we have a matrix dual variable $Y \succeq 0$.

SDP duality

- SDP primal form

$$p^* := \min \mathrm{Tr}(CX), \quad \text{s.t. } \mathrm{Tr}(A_i X) = b_i, \quad i = 1, \ldots, m, \quad X \succeq 0.$$

- How to handle the matrix constraint $X \succeq 0$?
- Introduce **conic Lagrangian**

$$\mathcal{L}(X, \nu, Y) := \mathrm{Tr}(CX) + \sum\nolimits_i \nu_i(\mathrm{Tr}(A_i X) - b_i) - \mathrm{Tr}(YX)$$

where we have a matrix dual variable $Y \succeq 0$.

- Note: $\mathrm{Tr}(YX) \geq 0$; so $p^* \geq \sup_{\nu,Y} \mathcal{L}(X, \nu, Y)$ for any feasible X

SDP duality

- SDP primal form

$$p^* := \min \mathrm{Tr}(CX), \quad \text{s.t. } \mathrm{Tr}(A_i X) = b_i, \quad i = 1, \ldots, m, \quad X \succeq 0.$$

- How to handle the matrix constraint $X \succeq 0$?
- Introduce **conic Lagrangian**

$$\mathcal{L}(X, \nu, Y) := \mathrm{Tr}(CX) + \sum_i \nu_i(\mathrm{Tr}(A_i X) - b_i) - \mathrm{Tr}(YX)$$

where we have a matrix dual variable $Y \succeq 0$.

- Note: $\mathrm{Tr}(YX) \geq 0$; so $p^* \geq \sup_{\nu,Y} \mathcal{L}(X, \nu, Y)$ for any feasible X
- As before, $p^* \geq d^* := \sup_{\nu, Y \succeq 0} \inf_X \mathcal{L}(X, \nu, Y)$
- Simplifying $\inf_X \mathcal{L}$, we obtain dual function

$$g(\nu, Y) = \begin{cases} b^T \nu & \text{if } C - \sum_i \nu_i A_i - Y = 0, \\ -\infty & \text{otherwise.} \end{cases}$$

SDP Duality

Dual problem

$$\max_{\nu, Y \succeq 0} \quad b^T \nu \quad \text{s.t. } C - \sum_i \nu_i A_i = Y \succeq 0$$

SDP Duality

Dual problem

$$\max_{\nu, Y \succeq 0} \quad b^T \nu \quad \text{s.t. } C - \sum_i \nu_i A_i = Y \succeq 0$$

$$\max_{\nu} \quad b^T \nu \quad \text{s.t. } \sum_i \nu_i A_i \preceq C.$$

SDP Duality

Dual problem

$$\max_{\nu, Y \succeq 0} \quad b^T\nu \quad \text{s.t. } C - \sum_i \nu_i A_i = Y \succeq 0$$

$$\max_{\nu} \quad b^T\nu \quad \text{s.t. } \sum_i \nu_i A_i \preceq C.$$

This is the conic form we saw in Lecture 5!

SDP Duality

Dual problem

$$\max_{\nu, Y \succeq 0} \quad b^T \nu \quad \text{s.t. } C - \sum_i \nu_i A_i = Y \succeq 0$$

$$\max_{\nu} \quad b^T \nu \quad \text{s.t. } \sum_i \nu_i A_i \preceq C.$$

This is the conic form we saw in Lecture 5!

- Weak-duality: $\mathrm{Tr}(CX) \geq \nu^T b$ for any feasible pair (X, ν)

SDP Duality

Dual problem

$$\max_{\nu, Y \succeq 0} \quad b^T\nu \quad \text{s.t. } C - \sum\nolimits_i \nu_i A_i = Y \succeq 0$$

$$\max_{\nu} \quad b^T\nu \quad \text{s.t. } \sum\nolimits_i \nu_i A_i \preceq C.$$

This is the conic form we saw in Lecture 5!

- Weak-duality: $\mathrm{Tr}(CX) \geq \nu^T b$ for any feasible pair (X, ν)
- Strong-duality: If primal strictly feasible,

SDP Duality

Dual problem

$$\max_{\nu, Y \succeq 0} \quad b^T\nu \quad \text{s.t. } C - \sum_i \nu_i A_i = Y \succeq 0$$

$$\max_{\nu} \quad b^T\nu \quad \text{s.t. } \sum_i \nu_i A_i \preceq C.$$

This is the conic form we saw in Lecture 5!

- Weak-duality: $\mathrm{Tr}(CX) \geq \nu^T b$ for any feasible pair (X, ν)
- Strong-duality: If primal strictly feasible, $X \succ 0$ such that $\mathrm{Tr}(A_i X) = b_i$, for $i = 1, \ldots, m$, we have strong duality.

SDP Duality

Dual problem

$$\max_{\nu, Y \succeq 0} \quad b^T\nu \quad \text{s.t. } C - \sum_i \nu_i A_i = Y \succeq 0$$

$$\max_{\nu} \quad b^T\nu \quad \text{s.t. } \sum_i \nu_i A_i \preceq C.$$

This is the conic form we saw in Lecture 5!

- Weak-duality: $\mathrm{Tr}(CX) \geq \nu^T b$ for any feasible pair (X, ν)
- Strong-duality: If primal strictly feasible, $X \succ 0$ such that $\mathrm{Tr}(A_i X) = b_i$, for $i = 1, \ldots, m$, we have strong duality.
- Alternatively, if dual strictly feasible, we have strong duality.

SDP Duality

Dual problem

$$\max_{\nu, Y \succeq 0} \quad b^T\nu \quad \text{s.t. } C - \sum_i \nu_i A_i = Y \succeq 0$$

$$\max_{\nu} \quad b^T\nu \quad \text{s.t. } \sum_i \nu_i A_i \preceq C.$$

This is the conic form we saw in Lecture 5!

- Weak-duality: $\mathrm{Tr}(CX) \geq \nu^T b$ for any feasible pair (X, ν)
- Strong-duality: If primal strictly feasible, $X \succ 0$ such that $\mathrm{Tr}(A_i X) = b_i$, for $i = 1, \ldots, m$, we have strong duality.
- Alternatively, if dual strictly feasible, we have strong duality.
- But, contrary to LPs, **feasibility alone does not suffice!**

Example: failure of strong duality

Primal problem

$$p^* = \min_X \quad x_2 \quad \begin{bmatrix} x_2 + 1 & 0 & 0 \\ 0 & x_1 & x_2 \\ 0 & x_2 & 0 \end{bmatrix} \succeq 0.$$

Example: failure of strong duality

Primal problem

$$p^* = \min_X \quad x_2 \quad \begin{bmatrix} x_2+1 & 0 & 0 \\ 0 & x_1 & x_2 \\ 0 & x_2 & 0 \end{bmatrix} \succeq 0.$$

Any primal feasible requires $\begin{bmatrix} x_1 & x_2 \\ x_2 & 0 \end{bmatrix} \succeq 0$;

Example: failure of strong duality

Primal problem

$$p^* = \min_{X} \quad x_2 \quad \begin{bmatrix} x_2 + 1 & 0 & 0 \\ 0 & x_1 & x_2 \\ 0 & x_2 & 0 \end{bmatrix} \succeq 0.$$

Any primal feasible requires $\begin{bmatrix} x_1 & x_2 \\ x_2 & 0 \end{bmatrix} \succeq 0$; $x_1 \geq 0$ and $x_2^2 \leq 0$.

Example: failure of strong duality

Primal problem

$$p^* = \min_X \quad x_2 \quad \begin{bmatrix} x_2 + 1 & 0 & 0 \\ 0 & x_1 & x_2 \\ 0 & x_2 & 0 \end{bmatrix} \succeq 0.$$

Any primal feasible requires $\begin{bmatrix} x_1 & x_2 \\ x_2 & 0 \end{bmatrix} \succeq 0$; $x_1 \geq 0$ and $x_2^2 \leq 0$.

Thus, we have $x_2 = 0$, whereby $p^* = 0$.

Primal obj: $\mathrm{Tr}(CX)$ with $c_{23} = c_{32} = 1/2$ (rest zeros).

Example: failure of strong duality

Primal problem

$$p^* = \min_X \quad x_2 \quad \begin{bmatrix} x_2+1 & 0 & 0 \\ 0 & x_1 & x_2 \\ 0 & x_2 & 0 \end{bmatrix} \succeq 0.$$

Any primal feasible requires $\begin{bmatrix} x_1 & x_2 \\ x_2 & 0 \end{bmatrix} \succeq 0$; $x_1 \geq 0$ and $x_2^2 \leq 0$.
Thus, we have $x_2 = 0$, whereby $p^* = 0$.

Primal obj: $\mathrm{Tr}(CX)$ with $c_{23} = c_{32} = 1/2$ (rest zeros).

Lagrangian: $\mathrm{Tr}([C - X]^T Y)$

Example: failure of strong duality

$$\begin{aligned}\operatorname{Tr}([C - X]^T Y) &= -(x_2+1)y_{11} - x_1 y_{22} + x_2 - 2x_2 y_{23} \\ &= -y_{11} - x_1 y_{22} + x_2 - x_2 y_{11} - 2x_2 y_{23}.\end{aligned}$$

Dual function

$$g(Y) = \inf_{X \succeq 0} \operatorname{Tr}([C-X]^T Y) = \begin{cases} -y_{11} & y_{22} = 0, 1 - y_{11} - 2y_{23} = 0 \\ -\infty & \text{otherwise.} \end{cases}$$

Example: failure of strong duality

$$\begin{aligned}\mathrm{Tr}([C-X]^T Y) &= -(x_2+1)y_{11} - x_1 y_{22} + x_2 - 2x_2 y_{23} \\ &= -y_{11} - x_1 y_{22} + x_2 - x_2 y_{11} - 2x_2 y_{23}.\end{aligned}$$

Dual function

$$g(Y) = \inf_{X\succeq 0} \mathrm{Tr}([C-X]^T Y) = \begin{cases} -y_{11} & y_{22}=0, 1-y_{11}-2y_{23}=0 \\ -\infty & \text{otherwise.}\end{cases}$$

Dual SDP

$$d^* = \max_{Y\succeq 0} \; -y_{11}, \quad y_{22}=0,\; 1-y_{11}-2y_{23}=0.$$

Example: failure of strong duality

$$\begin{aligned}\mathrm{Tr}([C-X]^TY) &= -(x_2+1)y_{11} - x_1y_{22} + x_2 - 2x_2y_{23}\\ &= -y_{11} - x_1y_{22} + x_2 - x_2y_{11} - 2x_2y_{23}.\end{aligned}$$

Dual function

$$g(Y) = \inf_{X\succeq 0} \mathrm{Tr}([C-X]^TY) = \begin{cases} -y_{11} & y_{22}=0, 1-y_{11}-2y_{23}=0 \\ -\infty & \text{otherwise.}\end{cases}$$

Dual SDP

$$d^* = \max_{Y\succeq 0}\ -y_{11}, \quad y_{22}=0,\ 1-y_{11}-2y_{23}=0.$$

- Any feasible Y satisfies, $y_{23}=0$ (since $y_{22}=0$)

Example: failure of strong duality

$$\begin{aligned}\mathrm{Tr}([C-X]^T Y) &= -(x_2+1)y_{11} - x_1 y_{22} + x_2 - 2x_2 y_{23} \\ &= -y_{11} - x_1 y_{22} + x_2 - x_2 y_{11} - 2x_2 y_{23}.\end{aligned}$$

Dual function

$$g(Y) = \inf_{X \succeq 0} \mathrm{Tr}([C-X]^T Y) = \begin{cases} -y_{11} & y_{22} = 0, 1 - y_{11} - 2y_{23} = 0 \\ -\infty & \text{otherwise.} \end{cases}$$

Dual SDP

$$d^* = \max_{Y \succeq 0} \; -y_{11}, \quad y_{22} = 0,\; 1 - y_{11} - 2y_{23} = 0.$$

- Any feasible Y satisfies, $y_{23} = 0$ (since $y_{22} = 0$)
- Thus $y_{11} = 1$, so $d^* = -1.$
- **duality gap**: $p^* - d^* = 1$

Optimality conditions

Optimality conditions

$$\min \quad f_0(x) \qquad f_i(x) \leq 0, \quad i = 1, \ldots, m.$$

Optimality conditions

$$\min \quad f_0(x) \qquad f_i(x) \leq 0, \quad i = 1, \ldots, m.$$

- Recall: $\langle \nabla f_0(x^*), x - x^* \rangle \geq 0$ for all feasible $x \in \mathcal{X}$

Optimality conditions

$$\min \quad f_0(x) \qquad f_i(x) \leq 0, \quad i = 1, \ldots, m.$$

- Recall: $\langle \nabla f_0(x^*), x - x^* \rangle \geq 0$ for all feasible $x \in \mathcal{X}$
- Can we simplify this using Lagrangian?

Optimality conditions

$$\min \quad f_0(x) \qquad f_i(x) \le 0, \quad i = 1, \ldots, m.$$

- Recall: $\langle \nabla f_0(x^*), x - x^* \rangle \geq 0$ for all feasible $x \in \mathcal{X}$
- Can we simplify this using Lagrangian?
- $g(\lambda) = \inf_x \mathcal{L}(x, \lambda) := f_0(x) + \sum_i \lambda_i f_i(x)$

Optimality conditions

$$\min \quad f_0(x) \qquad f_i(x) \leq 0, \quad i = 1, \ldots, m.$$

- Recall: $\langle \nabla f_0(x^*), x - x^* \rangle \geq 0$ for all feasible $x \in \mathcal{X}$
- Can we simplify this using Lagrangian?
- $g(\lambda) = \inf_x \mathcal{L}(x, \lambda) := f_0(x) + \sum_i \lambda_i f_i(x)$

Assume strong duality and that both p^* and d^* attained!

Optimality conditions

$$\min \quad f_0(x) \qquad f_i(x) \leq 0, \quad i = 1, \ldots, m.$$

- Recall: $\langle \nabla f_0(x^*), x - x^* \rangle \geq 0$ for all feasible $x \in \mathcal{X}$
- Can we simplify this using Lagrangian?
- $g(\lambda) = \inf_x \mathcal{L}(x, \lambda) := f_0(x) + \sum_i \lambda_i f_i(x)$

Assume strong duality and that both p^* and d^* attained!

Thus, there exists a pair (x^*, λ^*) such that

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

Optimality conditions

$$\min \quad f_0(x) \qquad f_i(x) \leq 0, \quad i = 1, \ldots, m.$$

- Recall: $\langle \nabla f_0(x^*), x - x^* \rangle \geq 0$ for all feasible $x \in \mathcal{X}$
- Can we simplify this using Lagrangian?
- $g(\lambda) = \inf_x \mathcal{L}(x, \lambda) := f_0(x) + \sum_i \lambda_i f_i(x)$

Assume strong duality and that both p^* and d^* attained!

Thus, there exists a pair (x^*, λ^*) such that

$$p^* = f_0(x^*) = d^* = g(\lambda^*)$$

Optimality conditions

$$\min \quad f_0(x) \qquad f_i(x) \le 0, \quad i = 1, \dots, m.$$

- Recall: $\langle \nabla f_0(x^*),\, x - x^* \rangle \ge 0$ for all feasible $x \in \mathcal{X}$
- Can we simplify this using Lagrangian?
- $g(\lambda) = \inf_x \mathcal{L}(x, \lambda) := f_0(x) + \sum_i \lambda_i f_i(x)$

Assume strong duality and that both p^* and d^* attained!

Thus, there exists a pair (x^*, λ^*) such that

$$p^* = f_0(x^*) = d^* = g(\lambda^*) = \min_x \mathcal{L}(x, \lambda^*)$$

Optimality conditions

$$\min \quad f_0(x) \qquad f_i(x) \leq 0, \quad i = 1, \ldots, m.$$

- Recall: $\langle \nabla f_0(x^*), x - x^* \rangle \geq 0$ for all feasible $x \in \mathcal{X}$
- Can we simplify this using Lagrangian?
- $g(\lambda) = \inf_x \mathcal{L}(x, \lambda) := f_0(x) + \sum_i \lambda_i f_i(x)$

Assume strong duality and that both p^* and d^* attained!

Thus, there exists a pair (x^*, λ^*) such that

$$p^* = f_0(x^*) = d^* = g(\lambda^*) = \min_x \mathcal{L}(x, \lambda^*) \leq \mathcal{L}(x^*, \lambda^*)$$

Optimality conditions

$$\min \quad f_0(x) \qquad f_i(x) \leq 0, \quad i = 1, \ldots, m.$$

- Recall: $\langle \nabla f_0(x^*), x - x^* \rangle \geq 0$ for all feasible $x \in \mathcal{X}$
- Can we simplify this using Lagrangian?
- $g(\lambda) = \inf_x \mathcal{L}(x, \lambda) := f_0(x) + \sum_i \lambda_i f_i(x)$

Assume strong duality and that both p^* and d^* attained!

Thus, there exists a pair (x^*, λ^*) such that

$$p^* = f_0(x^*) = d^* = g(\lambda^*) = \min_x \mathcal{L}(x, \lambda^*) \leq \mathcal{L}(x^*, \lambda^*) \leq f_0(x^*) = p^*$$

Optimality conditions

$$\min \quad f_0(x) \qquad f_i(x) \leq 0, \quad i = 1, \ldots, m.$$

- Recall: $\langle \nabla f_0(x^*), x - x^* \rangle \geq 0$ for all feasible $x \in \mathcal{X}$
- Can we simplify this using Lagrangian?
- $g(\lambda) = \inf_x \mathcal{L}(x, \lambda) := f_0(x) + \sum_i \lambda_i f_i(x)$

Assume strong duality and that both p^* and d^* attained!

Thus, there exists a pair (x^*, λ^*) such that

$$p^* = f_0(x^*) = d^* = g(\lambda^*) = \min_x \mathcal{L}(x, \lambda^*) \leq \mathcal{L}(x^*, \lambda^*) \leq f_0(x^*) = p^*$$

- Thus, equalities hold in above chain.

Optimality conditions

$$\min \quad f_0(x) \qquad f_i(x) \leq 0, \quad i = 1, \ldots, m.$$

- Recall: $\langle \nabla f_0(x^*), x - x^* \rangle \geq 0$ for all feasible $x \in \mathcal{X}$
- Can we simplify this using Lagrangian?
- $g(\lambda) = \inf_x \mathcal{L}(x, \lambda) := f_0(x) + \sum_i \lambda_i f_i(x)$

Assume strong duality and that both p^* and d^* attained!

Thus, there exists a pair (x^*, λ^*) such that

$$p^* = f_0(x^*) = d^* = g(\lambda^*) = \min_x \mathcal{L}(x, \lambda^*) \leq \mathcal{L}(x^*, \lambda^*) \leq f_0(x^*) = p^*$$

- Thus, equalities hold in above chain.

$$x^* \in \operatorname{argmin}_x \mathcal{L}(x, \lambda^*).$$

Optimality conditions

$$\boxed{x^* \in \operatorname{argmin}_x \mathcal{L}(x, \lambda^*).}$$

If $f_0, f_1, \ldots, f_m$ are differentiable, this implies

Optimality conditions

$$\boxed{x^* \in \operatorname{argmin}_x \mathcal{L}(x, \lambda^*).}$$

If $f_0, f_1, \dots, f_m$ are differentiable, this implies

$$\nabla_x \mathcal{L}(x, \lambda^*)|_{x=x^*} = \nabla f_0(x^*) + \sum_i \lambda_i^* \nabla f_i(x^*) = 0.$$

Optimality conditions

$$\boxed{x^* \in \operatorname{argmin}_x \mathcal{L}(x, \lambda^*).}$$

If $f_0, f_1, \ldots, f_m$ are differentiable, this implies

$$\nabla_x \mathcal{L}(x, \lambda^*)|_{x=x^*} = \nabla f_0(x^*) + \sum_i \lambda_i^* \nabla f_i(x^*) = 0.$$

Moreover, since $\mathcal{L}(x^*, \lambda^*) = f_0(x^*)$, we also have

Optimality conditions

$$x^* \in \operatorname{argmin}_x \mathcal{L}(x, \lambda^*).$$

If $f_0, f_1, \ldots, f_m$ are differentiable, this implies

$$\nabla_x \mathcal{L}(x, \lambda^*)|_{x=x^*} = \nabla f_0(x^*) + \sum_i \lambda_i^* \nabla f_i(x^*) = 0.$$

Moreover, since $\mathcal{L}(x^*, \lambda^*) = f_0(x^*)$, we also have

$$\sum_i \lambda_i^* f_i(x^*) = 0.$$

Optimality conditions

$$x^* \in \operatorname{argmin}_x \mathcal{L}(x, \lambda^*).$$

If $f_0, f_1, \ldots, f_m$ are differentiable, this implies

$$\nabla_x \mathcal{L}(x, \lambda^*)|_{x=x^*} = \nabla f_0(x^*) + \sum_i \lambda_i^* \nabla f_i(x^*) = 0.$$

Moreover, since $\mathcal{L}(x^*, \lambda^*) = f_0(x^*)$, we also have

$$\sum_i \lambda_i^* f_i(x^*) = 0.$$

But $\lambda_i^* \geq 0$ and $f_i(x^*) \leq 0$,

Optimality conditions

$$x^* \in \operatorname{argmin}_x \mathcal{L}(x, \lambda^*).$$

If $f_0, f_1, \ldots, f_m$ are differentiable, this implies

$$\nabla_x \mathcal{L}(x, \lambda^*)|_{x=x^*} = \nabla f_0(x^*) + \sum_i \lambda_i^* \nabla f_i(x^*) = 0.$$

Moreover, since $\mathcal{L}(x^*, \lambda^*) = f_0(x^*)$, we also have

$$\sum_i \lambda_i^* f_i(x^*) = 0.$$

But $\lambda_i^* \geq 0$ and $f_i(x^*) \leq 0$, so **complementary slackness**

$$\lambda_i^* f_i(x^*) = 0, \quad i = 1, \ldots, m.$$

KKT Optimality conditions

Karush-Kuhn-Tucker Conditions (KKT)

$$
\begin{aligned}
f_i(x^*) &\leq 0, \quad i = 1, \ldots, m & \text{(primal feasibility)} \\
\lambda_i^* &\geq 0, \quad i = 1, \ldots, m & \text{(dual feasibility)} \\
\lambda_i^* f_i(x^*) &= 0, \quad i = 1, \ldots, m & \text{(compl. slackness)} \\
\nabla_x \mathcal{L}(x, \lambda^*)|_{x=x^*} &= 0 & \text{(Lagrangian stationarity)}
\end{aligned}
$$

KKT Optimality conditions

Karush-Kuhn-Tucker Conditions (KKT)

$$
\begin{aligned}
f_i(x^*) &\leq 0, \quad i = 1, \dots, m && \text{(primal feasibility)} \\
\lambda_i^* &\geq 0, \quad i = 1, \dots, m && \text{(dual feasibility)} \\
\lambda_i^* f_i(x^*) &= 0, \quad i = 1, \dots, m && \text{(compl. slackness)} \\
\nabla_x \mathcal{L}(x, \lambda^*)|_{x=x^*} &= 0 && \text{(Lagrangian stationarity)}
\end{aligned}
$$

- We showed: if strong duality holds, and (x^*, λ^*) exist, then KKT conditions are **necessary** for pair (x^*, λ^*) to be optimal

KKT Optimality conditions

Karush-Kuhn-Tucker Conditions (KKT)

$$\begin{aligned} f_i(x^*) &\leq 0, \quad i = 1, \ldots, m && \text{(primal feasibility)} \\ \lambda_i^* &\geq 0, \quad i = 1, \ldots, m && \text{(dual feasibility)} \\ \lambda_i^* f_i(x^*) &= 0, \quad i = 1, \ldots, m && \text{(compl. slackness)} \\ \nabla_x \mathcal{L}(x, \lambda^*)|_{x=x^*} &= 0 && \text{(Lagrangian stationarity)} \end{aligned}$$

- We showed: if strong duality holds, and (x^*, λ^*) exist, then KKT conditions are **necessary** for pair (x^*, λ^*) to be optimal
- If problem is convex, then KKT also **sufficient**

KKT Optimality conditions

Karush-Kuhn-Tucker Conditions (KKT)

$$\begin{aligned} f_i(x^*) &\leq 0, \quad i = 1, \ldots, m && \text{(primal feasibility)} \\ \lambda_i^* &\geq 0, \quad i = 1, \ldots, m && \text{(dual feasibility)} \\ \lambda_i^* f_i(x^*) &= 0, \quad i = 1, \ldots, m && \text{(compl. slackness)} \\ \nabla_x \mathcal{L}(x, \lambda^*)|_{x=x^*} &= 0 && \text{(Lagrangian stationarity)} \end{aligned}$$

- We showed: if strong duality holds, and (x^*, λ^*) exist, then KKT conditions are **necessary** for pair (x^*, λ^*) to be optimal
- If problem is convex, then KKT also **sufficient**

Exercise: Prove the above sufficiency of KKT. *Hint:* Use that $\mathcal{L}(x, \lambda^*)$ is convex, and conclude from KKT conditions that $g(\lambda^*) = f_0(x^*)$, so that (x^*, λ^*) optimal primal-dual pair.

Read Ch. 5 of BV

Minimax

Example: Lasso-like problem

$$p^* := \min_x \quad \|Ax - b\|_2 + \lambda \|x\|_1.$$

Example: Lasso-like problem

$$p^* := \min_x \quad \|Ax - b\|_2 + \lambda \|x\|_1.$$

$$\|x\|_1 = \max \left\{ x^T v \mid \|v\|_\infty \leq 1 \right\}$$
$$\|x\|_2 = \max \left\{ x^T u \mid \|u\|_2 \leq 1 \right\}.$$

Example: Lasso-like problem

$$p^* := \min_x \quad \|Ax - b\|_2 + \lambda \|x\|_1.$$

$$\|x\|_1 = \max\left\{x^T v \mid \|v\|_\infty \le 1\right\}$$
$$\|x\|_2 = \max\left\{x^T u \mid \|u\|_2 \le 1\right\}.$$

Saddle-point formulation

$$p^* \quad = \quad \min_x \max_{u,v} \ \left\{u^T(b - Ax) + v^T x \mid \|u\|_2 \le 1, \quad \|v\|_\infty \le \lambda\right\}$$

Example: Lasso-like problem

$$p^* := \min_x \quad \|Ax - b\|_2 + \lambda \|x\|_1.$$

$$\|x\|_1 = \max \left\{ x^T v \mid \|v\|_\infty \leq 1 \right\}$$
$$\|x\|_2 = \max \left\{ x^T u \mid \|u\|_2 \leq 1 \right\}.$$

Saddle-point formulation

$$\begin{aligned} p^* &= \min_x \max_{u,v} \left\{ u^T(b - Ax) + v^T x \mid \|u\|_2 \leq 1, \quad \|v\|_\infty \leq \lambda \right\} \\ &= \max_{u,v} \min_x \left\{ u^T(b - Ax) + x^T v \mid \|u\|_2 \leq 1, \quad \|v\|_\infty \leq \lambda \right\} \end{aligned}$$

Example: Lasso-like problem

$$p^* := \min_x \quad \|Ax - b\|_2 + \lambda\|x\|_1.$$

$$\|x\|_1 = \max\left\{x^T v \mid \|v\|_\infty \le 1\right\}$$
$$\|x\|_2 = \max\left\{x^T u \mid \|u\|_2 \le 1\right\}.$$

Saddle-point formulation

$$\begin{aligned}
p^* &= \min_x \max_{u,v} \left\{u^T(b - Ax) + v^T x \mid \|u\|_2 \le 1, \quad \|v\|_\infty \le \lambda\right\} \\
&= \max_{u,v} \min_x \left\{u^T(b - Ax) + x^T v \mid \|u\|_2 \le 1, \quad \|v\|_\infty \le \lambda\right\} \\
&= \max_{u,v} \ u^T b, \qquad A^T u = v, \ \|u\|_2 \le 1, \quad \|v\|_\infty \le \lambda
\end{aligned}$$

Example: Lasso-like problem

$$p^* := \min_x \quad \|Ax - b\|_2 + \lambda \|x\|_1.$$

$$\|x\|_1 = \max\left\{x^T v \mid \|v\|_\infty \le 1\right\}$$
$$\|x\|_2 = \max\left\{x^T u \mid \|u\|_2 \le 1\right\}.$$

Saddle-point formulation

$$\begin{aligned}
p^* &= \min_x \max_{u,v} \left\{u^T(b - Ax) + v^T x \mid \|u\|_2 \le 1, \quad \|v\|_\infty \le \lambda\right\} \\
&= \max_{u,v} \min_x \left\{u^T(b - Ax) + x^T v \mid \|u\|_2 \le 1, \quad \|v\|_\infty \le \lambda\right\} \\
&= \max_{u,v} \; u^T b, \qquad A^T u = v, \; \|u\|_2 \le 1, \quad \|v\|_\infty \le \lambda \\
&= \max_u \; u^T b, \qquad \|u\|_2 \le 1, \quad \|A^T u\|_\infty \le \lambda.
\end{aligned}$$

Minimax problems

- Minimax theory treats problems involving a combination of minimization and maximization

Minimax problems

- Minimax theory treats problems involving a combination of minimization and maximization
- Let $\mathcal{X}$ and $\mathcal{Y}$ be arbitrary nonempty sets

Minimax problems

- Minimax theory treats problems involving a combination of minimization and maximization
- Let $\mathcal{X}$ and $\mathcal{Y}$ be arbitrary nonempty sets
- Let $\phi : \mathcal{X} \times \mathcal{Y} \to \mathbb{R} \cup \{\pm\infty\}$

Minimax problems

- Minimax theory treats problems involving a combination of minimization and maximization
- Let $\mathcal{X}$ and $\mathcal{Y}$ be arbitrary nonempty sets
- Let $\phi : \mathcal{X} \times \mathcal{Y} \to \mathbb{R} \cup \{\pm\infty\}$
- **inf** over $y \in \mathcal{Y}$, followed by **sup** over $x \in \mathcal{X}$

$$\sup_{x \in \mathcal{X}} \inf_{y \in \mathcal{Y}} \phi(x, y) = \sup_{x \in \mathcal{X}} \psi(y(x))$$

Minimax problems

- Minimax theory treats problems involving a combination of minimization and maximization
- Let $\mathcal{X}$ and $\mathcal{Y}$ be arbitrary nonempty sets
- Let $\phi : \mathcal{X} \times \mathcal{Y} \to \mathbb{R} \cup \{\pm\infty\}$
- **inf** over $y \in \mathcal{Y}$, followed by **sup** over $x \in \mathcal{X}$

$$\sup_{x\in\mathcal{X}} \inf_{y\in\mathcal{Y}} \ \phi(x,y) = \sup_{x\in\mathcal{X}} \psi(y(x))$$

- **sup** over $x \in \mathcal{X}$, followed by **inf** over $y \in \mathcal{Y}$

$$\inf_{y\in\mathcal{Y}} \sup_{x\in\mathcal{X}} \ \phi(x,y) = \inf_{y\in\mathcal{Y}} \xi(x(y))$$

When are "inf sup" and "sup inf" equal?

Weak minimax

Theorem Let $\phi : \mathcal{X} \times \mathcal{Y} \to \mathbb{R} \cup \{\pm\infty\}$ be any function. Then,

$$\sup_{y \in \mathcal{Y}} \inf_{x \in \mathcal{X}} \phi(x, y) \quad \leq \quad \inf_{x \in \mathcal{X}} \sup_{y \in \mathcal{Y}} \phi(x, y)$$

Weak minimax

Theorem Let $\phi : \mathcal{X} \times \mathcal{Y} \to \mathbb{R} \cup \{\pm\infty\}$ be any function. Then,

$$\sup_{y \in \mathcal{Y}} \inf_{x \in \mathcal{X}} \phi(x, y) \quad \leq \quad \inf_{x \in \mathcal{X}} \sup_{y \in \mathcal{Y}} \phi(x, y)$$

Proof:

$$\forall x, y, \quad \inf_{x' \in \mathcal{X}} \phi(x', y) \;\; \leq \;\; \phi(x, y)$$

Weak minimax

Theorem Let $\phi : \mathcal{X} \times \mathcal{Y} \to \mathbb{R} \cup \{\pm\infty\}$ be any function. Then,

$$\sup_{y \in \mathcal{Y}} \inf_{x \in \mathcal{X}} \phi(x, y) \quad \leq \quad \inf_{x \in \mathcal{X}} \sup_{y \in \mathcal{Y}} \phi(x, y)$$

Proof:

$$\forall x, y, \quad \inf_{x' \in \mathcal{X}} \phi(x', y) \;\leq\; \phi(x, y)$$

$$\forall x, y, \quad \inf_{x' \in \mathcal{X}} \phi(x', y) \;\leq\; \sup_{y' \in \mathcal{Y}} \phi(x, y')$$

Weak minimax

Theorem Let $\phi : \mathcal{X} \times \mathcal{Y} \to \mathbb{R} \cup \{\pm\infty\}$ be any function. Then,

$$\sup_{y \in \mathcal{Y}} \inf_{x \in \mathcal{X}} \phi(x, y) \quad \leq \quad \inf_{x \in \mathcal{X}} \sup_{y \in \mathcal{Y}} \phi(x, y)$$

Proof:

$$\begin{aligned}
\forall x, y, \quad \inf_{x' \in \mathcal{X}} \phi(x', y) \quad &\leq \quad \phi(x, y) \\
\forall x, y, \quad \inf_{x' \in \mathcal{X}} \phi(x', y) \quad &\leq \quad \sup_{y' \in \mathcal{Y}} \phi(x, y') \\
\forall x, \quad \sup_{y \in \mathcal{Y}} \inf_{x' \in \mathcal{X}} \phi(x', y) \quad &\leq \quad \sup_{y' \in \mathcal{Y}} \phi(x, y')
\end{aligned}$$

Weak minimax

Theorem Let $\phi : \mathcal{X} \times \mathcal{Y} \to \mathbb{R} \cup \{\pm\infty\}$ be any function. Then,

$$\sup_{y \in \mathcal{Y}} \inf_{x \in \mathcal{X}} \phi(x, y) \quad \leq \quad \inf_{x \in \mathcal{X}} \sup_{y \in \mathcal{Y}} \phi(x, y)$$

Proof:

$$\begin{aligned}
\forall x, y, \quad & \inf_{x' \in \mathcal{X}} \phi(x', y) & \leq \quad & \phi(x, y) \\
\forall x, y, \quad & \inf_{x' \in \mathcal{X}} \phi(x', y) & \leq \quad & \sup_{y' \in \mathcal{Y}} \phi(x, y') \\
\forall x, \quad & \sup_{y \in \mathcal{Y}} \inf_{x' \in \mathcal{X}} \phi(x', y) & \leq \quad & \sup_{y' \in \mathcal{Y}} \phi(x, y') \\
\Longrightarrow \quad & \sup_{y \in \mathcal{Y}} \inf_{x' \in \mathcal{X}} \phi(x', y) & \leq \quad & \inf_{x \in \mathcal{X}} \sup_{y' \in \mathcal{Y}} \phi(x, y').
\end{aligned}$$

Weak minimax

Theorem Let $\phi : \mathcal{X} \times \mathcal{Y} \to \mathbb{R} \cup \{\pm\infty\}$ be any function. Then,

$$\sup_{y \in \mathcal{Y}} \inf_{x \in \mathcal{X}} \phi(x, y) \quad \leq \quad \inf_{x \in \mathcal{X}} \sup_{y \in \mathcal{Y}} \phi(x, y)$$

Proof:

$$\begin{aligned}
\forall x, y, \quad & \inf_{x' \in \mathcal{X}} \phi(x', y) && \leq && \phi(x, y) \\
\forall x, y, \quad & \inf_{x' \in \mathcal{X}} \phi(x', y) && \leq && \sup_{y' \in \mathcal{Y}} \phi(x, y') \\
\forall x, \quad & \sup_{y \in \mathcal{Y}} \inf_{x' \in \mathcal{X}} \phi(x', y) && \leq && \sup_{y' \in \mathcal{Y}} \phi(x, y') \\
\Longrightarrow \quad & \sup_{y \in \mathcal{Y}} \inf_{x' \in \mathcal{X}} \phi(x', y) && \leq && \inf_{x \in \mathcal{X}} \sup_{y' \in \mathcal{Y}} \phi(x, y').
\end{aligned}$$

Exercise: Show that weak duality is follows from above minimax inequality. *Hint:* Use $\phi = \mathcal{L}$ (Lagrangian), and suitably choose y.

Strong minimax

- If "inf sup" equals "sup inf", common value called **saddle-value**
- Value exists if there is a **saddle-point**, i.e., pair (x^*, y^*)

$$\phi(x, y^*) \geq \phi(x^*, y^*) \geq \phi(x^*, y) \quad \text{for all } x \in \mathcal{X}, y \in \mathcal{Y}.$$

Exercise: Verify above inequality!

Strong minimax

♠ Classes of problems "dual" to each other can be generated by studying classes of functions ϕ,

Strong minimax

- Classes of problems "dual" to each other can be generated by studying classes of functions ϕ,
- More interesting question: Starting from the primal problem over $\mathcal{X}$, how to introduce a space $\mathcal{Y}$ and a "useful" function ϕ on $\mathcal{X} \times \mathcal{Y}$ so that we have a saddle-point?

Strong minimax

- ♠ Classes of problems "dual" to each other can be generated by studying classes of functions ϕ,
- ♠ More interesting question: Starting from the primal problem over $\mathcal{X}$, how to introduce a space $\mathcal{Y}$ and a "useful" function ϕ on $\mathcal{X} \times \mathcal{Y}$ so that we have a saddle-point?

Sufficient conditions for saddle-point

- Function ϕ is continuous, and
- It is convex-concave ($\phi(\cdot, y)$ convex for every $y \in \mathcal{Y}$, and $\phi(x, \cdot)$ concave for every $x \in \mathcal{X}$), and
- Both $\mathcal{X}$ and $\mathcal{Y}$ are convex; one of them is compact.

Strong minimax

Def. Let ϕ be as before. A point (x^*, y^*) is a saddle-point of ϕ (min over $\mathcal{X}$ and max over $\mathcal{Y}$) **iff** the infimum in the expression

$$\inf_{x \in \mathcal{X}} \sup_{y \in \mathcal{Y}} \phi(x, y)$$

is **attained** at x^*, and the supremum in the expression

$$\sup_{y \in \mathcal{Y}} \inf_{x \in \mathcal{X}} \phi(x, y)$$

is **attained** at y^*, and these two extrema are equal.

Strong minimax

Def. Let ϕ be as before. A point (x^*, y^*) is a saddle-point of ϕ (min over $\mathcal{X}$ and max over $\mathcal{Y}$) **iff** the infimum in the expression

$$\inf_{x \in \mathcal{X}} \sup_{y \in \mathcal{Y}} \phi(x, y)$$

is **attained** at x^*, and the supremum in the expression

$$\sup_{y \in \mathcal{Y}} \inf_{x \in \mathcal{X}} \phi(x, y)$$

is **attained** at y^*, and these two extrema are equal.

$$x^* \in \operatorname*{argmin}_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} \phi(x, y) \qquad y^* \in \operatorname*{argmax}_{y \in \mathcal{Y}} \min_{x \in \mathcal{X}} \phi(x, y).$$

Optimality via minimax

$$x^* \in \underset{x \in \mathcal{X}}{\operatorname{argmin}} \max_{y \in \mathcal{Y}} \phi(x, y) \qquad y^* \in \underset{y \in \mathcal{Y}}{\operatorname{argmax}} \min_{x \in \mathcal{X}} \phi(x, y).$$

Point (x^*, y^*) is a saddle-point if and only if

$$0 \in \partial \phi(x^*, y^*) = \partial_x \phi(x^*, y^*) \times \partial_y \phi(x^*, y^*)$$

Optimality via minimax

$$x^* \in \operatorname*{argmin}_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} \phi(x, y) \qquad y^* \in \operatorname*{argmax}_{y \in \mathcal{Y}} \min_{x \in \mathcal{X}} \phi(x, y).$$

Point (x^*, y^*) is a saddle-point if and only if

$$0 \in \partial\phi(x^*, y^*) = \partial_x \phi(x^*, y^*) \times \partial_y \phi(x^*, y^*)$$

When ϕ is of "convex-concave" form, yields KKT conditions.