Lecture #12

concave in $\underline{\lambda}, \underline{\nu}$

$$\underbrace{g(\underline{\lambda}, \underline{\nu})} \leq p^* \qquad \forall\, \underline{\lambda} \in \mathbb{R}^m_{\geq 0}$$

$$\& \quad \forall\, \underline{\nu} \in \mathbb{R}^p$$

Dual/Dual function/Lagrange dual

Tightest lower bound:

$$\underbrace{\boxed{\sup_{\substack{\underline{\lambda} \in \mathbb{R}^m_{\geq 0} \\ \underline{\nu} \in \mathbb{R}^p}} g(\underline{\lambda}, \underline{\nu})}} \leq p^*$$

$\rightarrow$ always convex optimization problem (Dual problem)

$$\boxed{d^* \leq p^*}$$

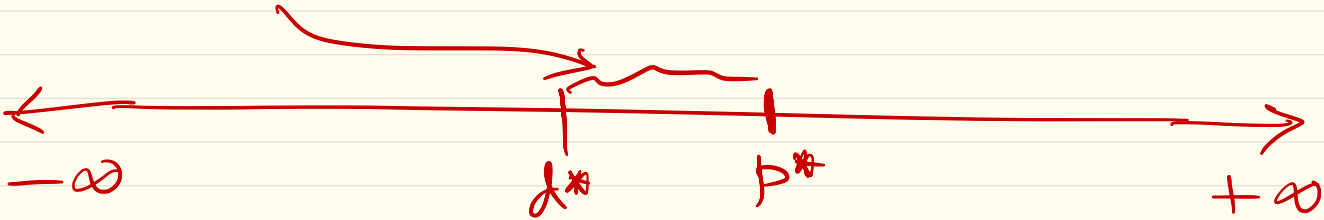
d^*: optimal value of the dual problem

p^*: optimal value of the primal problem

Weak Duality Theorem : (Primal may be non-convex)

Always. $\boxed{d^* \leq p^*}$

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



"Strong Duality" $\iff$ duality gap $= 0$

$\iff d^* = p^*$

Sufficient conditions for Strong Duality

If Primal problem is convex + (- . . . .)

constraint qualification

then **Strong duality holds.**

One "constraint qualification" condition is called "<u>Slater's condition</u>".

$\exists\ \underline{x} \in \text{relint}(\text{dom})$

s.t. $f_i(\underline{x}) < 0 \ \forall i=1,\dots,m$

(provided f_i's are nonlinear)

(strict feasibility)

Primal Problem

$$\min_{\underline{x}} f_0(\underline{x})$$

s.t. $f_i(\underline{x}) \leq 0$, $i=1,\dots,m$

$h_j(\underline{x}) = 0$, $j=1,\dots,p$

If "convex primal" + "Slater's condition"

then $d^* = p^*$

If $f_i(\underline{x})$ are linear in $\underline{x}$, then Slater's condition reduces to primal feasibility. (<u>Corollary:</u> LPs & QPs have strong duality)

Dual of LP : (Example)

Primal Problem :

LP

$$\min \; c^T x$$

$$\text{s.t. } G\underline{x} \preceq \underline{h} \in \mathbb{R}^m$$

$$A\underline{x} = \underline{b} \in \mathbb{R}^p$$

Step 1 :

Lagrangian :

$$L(\underline{x}, \underline{\lambda}, \underline{\nu}) = c^T \underline{x} + \lambda^T (G\underline{x} - \underline{h}) + \nu^T (A\underline{x} - \underline{b})$$

$$= (\underline{c}^T + \underline{\lambda}^T G + \underline{\nu}^T A)\, \underline{x} - \underline{\lambda}^T \underline{h} - \underline{\nu}^T \underline{b}$$

affine in $\underline{x}$

Step 2 :

Dual $f^{\underline{n}}$:

$$g(\underline{\lambda}, \underline{\nu}) := \inf_{\underline{x} \in \mathbb{R}^n} L(\underline{x}, \underline{\lambda}, \underline{\nu})$$

$$\Rightarrow \quad g(\underline{\lambda}, \underline{\nu}) = \begin{cases} -\underline{\lambda}^T \underline{h} - \underline{\nu}^T \underline{b} & \text{if } \underline{c} + G^T \underline{\lambda} + A^T \underline{\nu} = 0 \\ -\infty & \text{otherwise} \end{cases}$$

Step 3:

$\therefore$ Dual problem:

$$\max \; g(\underline{\lambda}, \underline{\nu})$$

$$\underline{\lambda} \in \mathbb{R}^m_{\geq 0}$$

$$\underline{\nu} \in \mathbb{R}^p$$

$$\Updownarrow$$

LP

$$\min_{\binom{\underline{\lambda}}{\underline{\nu}}} \begin{pmatrix} \underline{\lambda} \\ \underline{\nu} \end{pmatrix}^T \begin{pmatrix} \underline{h} \\ \underline{b} \end{pmatrix}$$

$$\text{s.t.} \quad \underline{c} + G^T \underline{\lambda} + A^T \underline{\nu} = 0$$

$$\underline{\lambda} \geq 0$$

This is a different LP than primal LP

but strong duality holds

Dual of QCQP (Example)

Primal: QCQP

$$\min_{\underline{x} \in \mathbb{R}^n} \underbrace{\frac{1}{2}\underline{x}^T P_0 \underline{x} + q_0^T \underline{x} + r_0}_{f_0(\underline{x})}$$

$$\text{s.t. } f_i(\underline{x}) = \frac{1}{2}\underline{x}^T P_i \underline{x} + q_i^T \underline{x} + r_i \le 0, \quad i = 1, \ldots, m$$

where $P_0 \in S^n_{++}$

and $P_i \in S^n_+ \quad \forall\, i = 1, \ldots, m$

Step 1: Lagrangian:

$$L(\underline{x}, \underline{\lambda}) = \frac{1}{2}\underline{x}^T P(\underline{\lambda})\, \underline{x} + \left(\underline{q}(\underline{\lambda})\right)^T \underline{x} + r(\underline{\lambda})$$

$$P(\underline{\lambda}) := P_0 + \sum_{i=1}^{m} \lambda_i P_i \succ 0, \quad \underline{q}(\underline{\lambda}) = \underline{q}_0 + \sum_{i=1}^{m} \lambda_i q_i$$

$$r(\underline{\lambda}) = r_0 + \sum_{i=1}^{m} \lambda_i r_i$$

Step 2 :

Dual $f^{\underline{n}}$: $g(\underline{\lambda}) = \inf_{\underline{x}\in\mathbb{R}^n} L(\underline{x}, \underline{\lambda})$

$$= -\frac{1}{2}\left(\underline{q}(\underline{\lambda})\right)^T \left(P(\underline{\lambda})\right)^{-1} \underline{q}(\underline{\lambda}) + r(\underline{\lambda})$$

Step 3 :

Dual problem:

$$d^* = \max_{\underline{\lambda}\geqslant 0} \left.\underbrace{g(\underline{\lambda})}_{\text{concave}}\right\} \text{convex}$$

In general, $d^* \leq p^*$ (Strong duality may not hold)

Slater's condition :

$\exists\, \underline{x}\in\mathbb{R}^n$, s.t. $\frac{1}{2}\underline{x}^T P_i \underline{x} + \underline{q}_i^T \underline{x} + r_i < 0$ $\forall i = 1, \ldots, m$

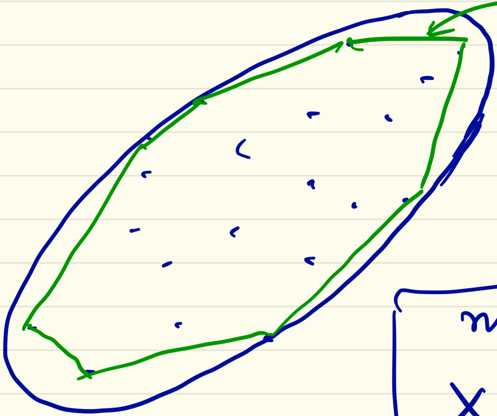
Example: $\mathcal{E}(\underline{0}_{n\times 1}, X) := \{\underline{z} \in \mathbb{R}^n \mid \underline{z}^T X \underline{z} \leq 1\}$

(Minimum vol^{m} ellipsoid covering a finite set of points)

(Ellipsoid in $\mathbb{R}^n$) where $X \in \mathbb{S}^n_{++}$

$$\text{Vol}(\mathcal{E}(0_{n\times 1}, X)) \propto \sqrt{\det(X^{-1})}$$

convex hull of the datapoints



m datapoints in $\mathbb{R}^n$

$\underline{a}_i \in \mathbb{R}^n, \quad i = 1, \ldots, m$

$$\boxed{\begin{aligned} \min_{X \in \mathbb{S}^n_{++}} \quad & f_0(X) = \log\det(X^{-1}) \\ \text{s.t.} \quad & \underline{a}_i^T X \underline{a}_i \leq 1, \quad i = 1, \ldots, m \end{aligned}}$$

Step 1 :

Lagrangian:

$$L(X,\underline{\lambda}) = \underbrace{\log\det(X^{-1})}_{f_0(X)} + \sum_{i=1}^{m} \lambda_i \Big(\underbrace{a_i^T X a_i}_{\substack{\shortparallel \\ \mathrm{tr}(a_i^T X a_i)}} - 1\Big)$$

$$= \mathrm{tr}(\underbrace{a_i a_i^T}_{\shortparallel} X)$$

$$= \mathrm{tr}(A_i X)$$

where $A_i = \underbrace{a_i a_i^T}_{n\times n}$

Step 2 :

Dual/Dual $f^{\underline{n}}$:

$$g(\underline{\lambda}) = \inf_{X\in S_{++}^n} L(X,\underline{\lambda})$$

$$\frac{\partial L}{\partial X} = 0 \Rightarrow \text{solve for } X:$$

$$\Rightarrow \underbrace{\frac{\partial f_0}{\partial X}} + \frac{\partial}{\partial X}\sum_{i=1}^{m}\lambda_i\Big(\mathrm{tr}(A_i X) - 1\Big) = 0$$

$$\Rightarrow -X^{-T}_{\text{(Lec. 10's last page)}} + \sum_{i=1}^{m}\lambda_i \underbrace{\frac{\partial\, \mathrm{tr}(A_i X)}{\partial X}}_{\text{(Lec. 10 last page)}} = 0$$

$$\Rightarrow -X^{-T} + \sum_{i=1}^{m} \lambda_i A_i^T = 0$$

$A_i = a_i a_i^T$

$\Rightarrow A_i \in S^n$

$$\Rightarrow -X^{-1} + \sum_{i=1}^{m} \lambda_i A_i = 0$$

$$\Rightarrow \boxed{X^{-1} = \sum_{i=1}^{m} \lambda_i A_i} \Leftrightarrow I = \sum_{i=1}^{m} \lambda_i A_i X$$

$$\Leftrightarrow n = \sum_{i=1}^{m} \lambda_i \operatorname{tr}(A_i X)$$

step 2:

Dual f^n:

$$g(\lambda) = L(X = fn(\underline{\lambda}), \underline{\lambda})$$

$$= \log\det\left(\sum_{i=1}^{m} \lambda_i A_i\right) - \mathbb{1}^T \underline{\lambda} + \sum_{i=1}^{m} \lambda_i \operatorname{tr}(A_i X)$$

n

Step 3

Dual Problem: $\min_{\underline{\lambda} \geq 0} \mathbb{1}^T \underline{\lambda} + \log\det\left(\sum_{i=1}^{m} \lambda_i A_i\right)^{-1} n$

(convex)

Primal convex + Slater's condition holds

$\Rightarrow d^* = p^*$

Non-convex but Strong duality holds (example)

$$\min_{\underline{x} \in \mathbb{R}^n} \underline{x}^T A \underline{x} + 2\underline{b}^T \underline{x}$$

$$\text{s.t.} \quad \underline{x}^T \underline{x} \leq 1$$

Trust-region problem (minimize non-convex quadratic fn over unit ball)

$A \not\succeq 0 \Rightarrow$ non-convex

Step 1: Lagrangian:

$$L(\underline{x}, \lambda) = \underline{x}^T (A + \lambda I) \underline{x} + 2\underline{b}^T \underline{x} - \lambda$$

Step 2 Dual

$$g(\lambda) = \inf_{\underline{x} \in \mathbb{R}^n} L(\underline{x}, \lambda)$$

$$\therefore g(\lambda) = \begin{cases} -\underline{b}^T (A+\lambda I)^{+} \underline{b} - \lambda & \text{if } A+\lambda I \succcurlyeq 0,\ \underline{b} \in \text{Range}(A+\lambda I) \\ -\infty & \text{otherwise} \end{cases}$$

pseudo-inverse

Step 3: Dual problem:

convex (proof non-obvious)

$$\max \ -\underline{b}^T (A+\lambda I)^{+} \underline{b} - \lambda$$

$$\text{s.t.} \quad A+\lambda I \succcurlyeq 0$$

$$\underline{b} \in \text{Range}(A+\lambda I)$$

$d^* = p^*$ (Strong duality holds) (p. 229 of text)

Other Applications of Duality
(Algorithm to solve convex problems)

$\underline{x}^{(k)}$ — Primal feasible seq.

$\underbrace{(\underline{\lambda}^{(k)}, \underline{\nu}^{(k)})}_{\text{Dual feasible seq.}}$, $\underbrace{\text{tol} = \epsilon_{abs}}_{\text{numerical tolerance}}$

Stopping criterion:

$$f_0(\underline{x}^{(k)}) - g(\underline{\lambda}^{(k)}, \underline{\nu}^{(k)}) \leq \epsilon_{abs}$$

Complimentary Slackness :

Suppose strong duality holds :

$\underline{x}^*$ (primal optimizer)

$(\underline{\lambda}^*, \underline{\nu}^*)$ (dual optimizer)

$$p^* = d^*$$

$$\Rightarrow \quad f_0(\underline{x}^*) = g(\underline{\lambda}^*, \underline{\nu}^*)$$

$$= \inf_{\underline{x} \in \mathbb{R}^n} L(\underline{x}^*, \underline{\lambda}^*, \underline{\nu}^*)$$

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

Since $\sum_{i=1}^{m} \lambda_i^* f_i(\underline{x}^*) \leq 0$

$$\leq f_0(\underline{x}^*)$$

$\Rightarrow$ The 2 inequalities must be equalities

$$\Rightarrow \quad \sum_i \overset{\geq 0}{\lambda_i^*} \, \overset{\leq 0}{\overbrace{f_i(\underline{x}^*)}} = 0$$

$$\Leftrightarrow \quad \boxed{\lambda_i^* f_i(x_i^*) = 0}$$

i.e.,
$$\left.\begin{aligned} \lambda_i^* > 0 &\Rightarrow f_i(x^*) = 0 \\ f_i(x^*) < 0 &\Rightarrow \lambda_i^* = 0 \end{aligned}\right\} \text{Complimentary slackness}$$

Relⁿ between Lagrange & $f^*(\cdot)$ dual

Legendre-Fenchel transform

convex conjugate

Intuition

Consider $\min_{\underline{x} \in \mathbb{R}^n} f_0(\underline{x})$

s.t. $\underline{x} = 0$

$$L(\underline{x}, \underline{\nu}) = f_0(\underline{x}) + \underline{\nu}^T \underline{x}$$

$$\therefore \underbrace{g(\underline{\nu})}_{\text{Dual}} = \inf_{\underline{x}} \left(f_0(\underline{x}) + \underline{\nu}^T \underline{x} \right)$$

$$= -\sup_{\underline{x}} \left((-\underline{\nu})^T \underline{x} - f_0(\underline{x}) \right)$$

$$= - \underbrace{f_0^*(-\nu)}_{\text{Leg. Fenchel}}$$

In general,

$$\min \; f_0(\underline{x})$$
$$\text{s.t. } \quad A x \preceq b, \quad C x = d \quad \Big\} \text{ Linear constraints}$$

Then

$$g(\underline{\lambda}, \underline{\nu}) = \inf_{\underline{x}} \Big(f_0(\underline{x}) + \underline{\lambda}^T (A\underline{x} - \underline{b}) + \nu^T (C\underline{x} - \underline{d}) \Big)$$

$$= -\underline{b}^T \underline{\lambda} - \underline{d}^T \underline{\nu} + \underbrace{\inf_{\underline{x}} \Big(f_0(x) + (A^T\lambda + C^T\nu)^T x \Big)}$$

$$= \boxed{-\underline{b}^T \underline{\lambda} - \underline{d}^T \underline{\nu} - f_0^* \left(-A^T \underline{\lambda} - C^T \underline{\nu}\right)}$$

For any optimization problem with linear constraints you can write the Lagrange Dual fn in terms of the convex conjugate of f_0 (primal objective)

Example:

(maximize entropy subject to halfspace constraints)

$$\max_{\underline{x} \in \mathbb{R}^n_{\geq 0}} f_0(x) = -\sum_{i=1}^n x_i \log x_i$$

s.t. $A x \preceq b$ (m inequality constraints)

$\mathbb{1}^T \underline{x} = 1$ (single equality constraint)

If $f(u) = u \log u$ then $f^*(v) = e^{v-1}$

$$\therefore \; f_0^*(\underline{y}) = \sum_{i=1}^n e^{y_i - 1}, \quad \mathrm{dom}(f_0^*) = \mathbb{R}^n$$

Then

$$g(\underline{\lambda}, \underline{\nu}) = -\underline{b}^T \underline{\lambda} - \nu - \sum_{i=1}^n \exp\left(-a_i^T \underline{\lambda} - \nu - 1\right)$$

(Lagrange dual fn.)

$$= -\underline{b}^T \underline{\lambda} - \nu - \exp(-\nu - 1) \sum_{i=1}^n e^{-a_i^T \lambda}$$

where a_i = i^{th} column of matrix A.

Notice that we were able to write $g(\underline{\lambda}, \underline{\nu})$ without taking derivative, in this example, thanks to the convex conjugate.

Therefore, the dual problem:

$$d^* = \underset{\substack{\underline{\lambda} \in \mathbb{R}^m_{\geq 0} \\ \nu \in \mathbb{R}}}{\text{minimize}} \quad \underline{b}^T \underline{\lambda} + \nu + \exp(-\nu - 1) \sum_{i=1}^{n} \exp(-\underline{a}_i^T \underline{\lambda})$$

We can simplify this further by analytically minimizing over $\nu \in \mathbb{R}$ while keeping $\underline{\lambda} \in \mathbb{R}^m_{\geq 0}$ fixed. This gives $\nu^* = \log\left(\sum_{i=1}^{n} \exp(-a_i^T \underline{\lambda})\right) - 1$

Substituting back:

$$\boxed{d^* = \underset{\underline{\lambda} \in \mathbb{R}^m_{\geq 0}}{\text{minimize}} \quad \underline{b}^T \underline{\lambda} + \log\left(\sum_{i=1}^{n} \exp(-a_i^T \underline{\lambda})\right) \qquad \text{This is a GP}}$$

KKT condition

(Karush – Kuhn – Tucker condition)
$\underbrace{\text{Karush}}_{(1939)}$ – $\underbrace{\text{Kuhn – Tucker}}_{(1951)}$

Suppose

$\underbrace{f_0,}_{\text{objective}}\ \underbrace{f_1, f_2, \ldots, f_m,}_{\text{LHS of inequality constraints}}\ \underbrace{h_1, \ldots, h_p}_{\text{LHS of equality constraints}}$ are C^1 functions (continuously differentiable)

need not be convex

KKT conditions: @ $\underline{x}^*$ (primal optimizer), $\underline{\lambda}^*, \underline{\nu}^*$ (Dual optimizer)

① (Stationarity of the Lagrangian) $\nabla_{\underline{x}} L = 0$

$$\Leftrightarrow \nabla f_0(\underline{x}^*) + \sum_{i=1}^{m} \lambda_i^* \nabla_{\underline{x}} f_i(\underline{x}^*) + \sum_{j=1}^{p} \nu_j^* \nabla_{\underline{x}} h_j(\underline{x}^*) = \underline{0}$$

$$\underline{f} = \begin{pmatrix} f_1 \\ \vdots \\ f_m \end{pmatrix}, \quad \underline{h} = \begin{pmatrix} h_1 \\ \vdots \\ h_p \end{pmatrix}.$$

② (complimentary slackness) $\lambda_i^* f_i(x^*) = 0 \;\forall i = 1, \dots, m$

③ (Primal feasibility) $f_i(\underline{x}^*) \leq 0$; $h_i(\underline{x}^*) = 0$

④ (Dual feasibility) $\underline{\lambda}^* \succeq 0$.

Statement #1 (No convexity needed)

If ① $f_0, f_1, \dots, f_m, h_1, \dots, h_p$ are C^1

② Strong duality holds ($d^* = p^*$)

then the tuple $(\underline{x}^*, \underline{\lambda}^*, \underline{\nu}^*)$ must satisfy KKT condition.

Statement #2 : (convexity needed)

If ① $f_0, f_1, \ldots, f_m, h_1, \ldots, h_p$ are C^1

② Problem is convex

then any tuple $(\underline{\tilde{x}}, \underline{\tilde{\lambda}}, \underline{\tilde{\nu}})$ satisfying the KKT conditions must be primal & dual optimizers with $\underline{d^* = p^*}$.

Usage of Duality Theory for Sensitivity Analysis

Ch. 5.6 in text

Duality for SDPs

Primal Problem:

$$p^* = \min_{\underline{x} \in \mathbb{R}^n} \underline{c}^T \underline{x} \quad f_0(\underline{x})$$

$$\text{s.t.} \quad F(\underline{x}) := F_0 + x_1 F_1 + \dots + x_n F_n \succeq 0$$

where $F_0, F_1, \dots, F_n \in \mathbb{S}^n$

$$\Leftrightarrow -F(\underline{x}) \preceq 0$$

Step 1:

Lagrangian: Lagrange multiplier matrix

$$L(\underline{x}, Z) = f_0(\underline{x}) + \langle Z, -F(\underline{x}) \rangle$$

$Z \in K^* = K = \mathbb{S}^n_+$

(dual cone)

(since $\mathbb{S}^n_+$ is self dual)

$\langle A, B \rangle = \operatorname{tr}(A^T B)$

matrix inner product

$$\therefore L(\underline{x}, Z)$$

$$= x_1 \left(c_1 - \text{trace}(F_1 Z)\right) + x_2\left(c_2 - \text{trace}(F_2 Z)\right)$$

$$+ \ldots + x_n\left(c_n - \text{trace}(F_n Z)\right) - \text{trace}(F_0 Z)$$

~~Step 2~~: (Dual f^n / Lagrange dual)

$$g(Z) = \inf_{\underline{x} \in \mathbb{R}^n} L(\underline{x}, Z)$$

$$= \begin{cases} -\text{trace}(F_0 Z) & \text{if } \operatorname{tr}(F_i Z) = c_i \\ & \text{for } i = 1, \ldots, n \\ -\infty & \text{otherwise} \end{cases}$$

Step 3 : (Dual Problem)

$$d^* = \max_{Z \in S_+^n} \; -\operatorname{trace}(F_0 Z)$$

$$\operatorname{trace}(F_i Z) = c_i, \quad i = 1, \dots, n$$

$$\Updownarrow$$

$$\begin{cases} \min\limits_{Z \in S_+^n} \operatorname{trace}(F_0 Z) \\ \operatorname{trace}(F_i Z) = c_i \end{cases}$$

Strong duality holds if $\exists\, \underline{x} \in \mathbb{R}^n$ s.t. the primal is strictly feasible $\Leftrightarrow$

$\exists\, \underline{x} \in \mathbb{R}^n$ s.t. $F_0 + x_1 F_1 + \dots + x_n F_n \succ 0$.