Lecture 4: Lagrangian duality and optimality conditions

Yongwei Huang
Hong Kong University of Science and Technology (HKUST)

University of Napoli "Federico II"

June 25, 2009

Outline

- Lagrangian dual problem
- weak and strong duality
- geometric interpretation
- optimality conditions
- examples

Lagrangian

standard form problem (not necessarily convex)

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

variable $\boldsymbol{x} \in \mathbf{R}^n$, domain $\mathcal{D} = \bigcap_{i=0}^m \mathbf{dom}\, f_i \cap (\bigcap_{i=1}^p \mathbf{dom}\, h_i)$, optimal value $p^\star$

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

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

- weighted sum of objective and constraint functions
- λ_i is Lagrange multiplier associated with $f_i(\boldsymbol{x}) \leq 0$
- ν_i is Lagrange multiplier associated with $h_i(\boldsymbol{x}) = 0$
- $L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$ is affine function with respect to $\boldsymbol{\lambda}$ and $\boldsymbol{\nu}$

Lagrange dual function

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

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

$g(\boldsymbol{\lambda}, \boldsymbol{\nu})$ is concave, can be $-\infty$ for some $\boldsymbol{\lambda}$, $\boldsymbol{\nu}$

lower bound property: if $\boldsymbol{\lambda} \geq \mathbf{0}$ (i.e., $\lambda_i \geq 0$, $\forall i$), then $p^\star \geq g(\boldsymbol{\lambda}, \boldsymbol{\nu})$

proof: if $\boldsymbol{\lambda} \geq \mathbf{0}$, and $\tilde{\boldsymbol{x}}$ is feasible, namely,

$$\tilde{\boldsymbol{x}} \in X = \{\boldsymbol{x} | f_i(\boldsymbol{x}) \leq 0,\ i = 1, \ldots, m,\ h_i(\boldsymbol{x}) = 0,\ i = 1, \ldots, p\} \cap \textbf{dom}\, f_0,$$

then

$$f_0(\tilde{\boldsymbol{x}}) \geq L(\tilde{\boldsymbol{x}}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \geq \inf_{\boldsymbol{x} \in \mathcal{D}} L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = g(\boldsymbol{\lambda}, \boldsymbol{\nu})$$

minimizing over all feasible $\tilde{\boldsymbol{x}}$ gives

$$p^\star \geq g(\boldsymbol{\lambda}, \boldsymbol{\nu})$$

the "best" lower bound: $p^\star \geq d^\star = \begin{array}{ll} \underset{\boldsymbol{\lambda} \geq \mathbf{0},\, \boldsymbol{\nu} \in \mathbf{R}^p}{\text{maximize}} & g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \end{array}$

Proof that $-g(\boldsymbol{\lambda}, \boldsymbol{\nu})$ is convex

proof: the epigraph

$$
\begin{aligned}
\text{epi}(-g(\boldsymbol{\lambda}, \boldsymbol{\nu})) &= \{(\boldsymbol{\lambda}, \boldsymbol{\nu}, t) | -g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \leq t\} = \{(\boldsymbol{\lambda}, \boldsymbol{\nu}, t) | -\inf_{\boldsymbol{x} \in \mathcal{D}} L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \leq t\} \\
&= \{(\boldsymbol{\lambda}, \boldsymbol{\nu}, t) | L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \geq -t, \forall \boldsymbol{x} \in \mathcal{D}\} \\
&= \bigcap_{\boldsymbol{x} \in \mathcal{D}} \{(\boldsymbol{\lambda}, \boldsymbol{\nu}, t) | L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \geq -t\}
\end{aligned}
$$

Notice that given $\boldsymbol{x} \in \mathcal{D}$, Lagrangian $L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$ is affine function of $\boldsymbol{\lambda}$ and $\boldsymbol{\nu}$, thus the set

$$\{(\boldsymbol{\lambda}, \boldsymbol{\nu}, t) \in \mathbf{R}^{m+p+1} | L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \geq -t\}$$

is a half space in $\mathbf{R}^{m+p+1}$.

Therefore epi$(-g)$ is intersection of infinitely many convex sets (half spaces), and then is a convex set.

general result: given $L(\boldsymbol{x}, \boldsymbol{y})$ is affine, then

- $g(\boldsymbol{x}) = \inf_{\boldsymbol{y} \in \mathcal{D}} L(\boldsymbol{x}, \boldsymbol{y})$ is concave
- $h(\boldsymbol{x}) = \sup_{\boldsymbol{y} \in \mathcal{D}} L(\boldsymbol{x}, \boldsymbol{y})$ is convex

Least-norm solution of linear equations

$$\begin{array}{ll} \underset{\boldsymbol{x}\in\mathbf{R}^n}{\text{minimize}} & \boldsymbol{x}^T\boldsymbol{x} \\ \text{subject to} & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \end{array}$$

dual function

- Lagrangian is $L(\boldsymbol{x}, \boldsymbol{\nu}) = \boldsymbol{x}^T\boldsymbol{x} + \boldsymbol{\nu}^T(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})$
- to minimize L over $\boldsymbol{x}$, set gradient equal to zero:

$$\nabla_{\boldsymbol{x}} L(\boldsymbol{x}, \boldsymbol{\nu}) = 2\boldsymbol{x} + \boldsymbol{A}^T\boldsymbol{\nu} = \boldsymbol{0} \Rightarrow \boldsymbol{x} = -(1/2)\boldsymbol{A}^T\boldsymbol{\nu}$$

- plug in L to obtain g

$$g(\boldsymbol{\nu}) = L(-(1/2)\boldsymbol{A}^T\boldsymbol{\nu}, \boldsymbol{\nu}) = -\frac{1}{4}\boldsymbol{\nu}^T\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{\nu} - \boldsymbol{b}^T\boldsymbol{\nu}$$

  a concave function of $\boldsymbol{\nu}$

lower bound property: $p^\star \geq -(1/4)\boldsymbol{\nu}^T\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{\nu} - \boldsymbol{b}^T\boldsymbol{\nu}$ for all $\boldsymbol{\nu}$

Standard form LP

$$\begin{array}{ll} \underset{\boldsymbol{x}\in\mathbf{R}^n}{\text{minimize}} & \boldsymbol{c}^T\boldsymbol{x} \\ \text{subject to} & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \\ & \boldsymbol{x} \geq \mathbf{0} \end{array}$$

dual function

- Lagrangian is

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

- L is linear in $\boldsymbol{x}$, hence

$$g(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \inf_{\boldsymbol{x}} L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = \begin{cases} -\boldsymbol{b}^T\boldsymbol{\nu} & \boldsymbol{A}^T\boldsymbol{\nu} - \boldsymbol{\lambda} + \boldsymbol{c} = \mathbf{0} \\ -\infty & \text{otherwise} \end{cases}$$

  g is linear on affine domain $\{(\boldsymbol{\lambda}, \boldsymbol{\nu}) | \boldsymbol{A}^T\boldsymbol{\nu} - \boldsymbol{\lambda} + \boldsymbol{c} = \mathbf{0}\}$, hence concave

lower bound property: $p^\star \geq -\boldsymbol{b}^T\boldsymbol{\nu}$, if $\boldsymbol{\lambda} = \boldsymbol{A}^T\boldsymbol{\nu} + \boldsymbol{c} \geq \mathbf{0}$

about standard form

in Lecture 3, we study LP in the form:

$$\begin{array}{ll} \underset{\boldsymbol{x}\in\mathbf{R}^n}{\text{minimize}} & \boldsymbol{c}^T\boldsymbol{x} \\ \text{subject to} & \boldsymbol{G}\boldsymbol{x} \leq \boldsymbol{h}, \\ & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \end{array}$$

LP: objective function **linear**; constraint functions: **affine**

the two standard forms are compatible to each other: introducing slack variable $\boldsymbol{s} = \boldsymbol{h} - \boldsymbol{G}\boldsymbol{x}$, and interpreting $\boldsymbol{x} = \boldsymbol{x}_1 - \boldsymbol{x}_2$, $\boldsymbol{x}_1, \boldsymbol{x}_2 \geq \mathbf{0}$, then

$$\begin{array}{ll} \underset{\boldsymbol{x}\in\mathbf{R}^n}{\text{minimize}} & \boldsymbol{c}^T\boldsymbol{x} \\ \text{subject to} & \boldsymbol{G}\boldsymbol{x} \leq \boldsymbol{h}, \\ & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \end{array} = \begin{array}{ll} \underset{\boldsymbol{x}_1,\boldsymbol{x}_2\in\mathbf{R}^n,\boldsymbol{s}\in\mathbf{R}^m}{\text{minimize}} & \boldsymbol{c}^T\boldsymbol{x}_1 - \boldsymbol{c}^T\boldsymbol{x}_2 \\ \text{subject to} & \boldsymbol{G}\boldsymbol{x}_1 - \boldsymbol{G}\boldsymbol{x}_2 + \boldsymbol{I}\boldsymbol{s} = \boldsymbol{h}, \\ & \boldsymbol{A}\boldsymbol{x}_1 - \boldsymbol{A}\boldsymbol{x}_2 + \mathbf{0}\boldsymbol{s} = \boldsymbol{b} \\ & \boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{s} \geq \mathbf{0} \end{array}$$

on the other hand,

$$\begin{array}{ll} \underset{\boldsymbol{x}\in\mathbf{R}^n}{\text{minimize}} & \boldsymbol{c}^T\boldsymbol{x} \\ \text{subject to} & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \\ & \boldsymbol{x} \geq \mathbf{0} \end{array} = \begin{array}{ll} \underset{\boldsymbol{x}\in\mathbf{R}^n}{\text{minimize}} & \boldsymbol{c}^T\boldsymbol{x} \\ \text{subject to} & -\boldsymbol{I}\boldsymbol{x} \leq \mathbf{0} \\ & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \end{array}$$

Two-way partitioning

$$\begin{array}{ll} \text{minimize} & x^T W x \\ \text{subject to} & x_i^2 = 1, \quad i = 1, \ldots, n \end{array}$$

- a nonconvex problem; feasible set contains 2^n discrete points
- interpretation: partition $\{1, \ldots, n\}$ in two sets; W_{ij} is cost of assigning i, j to the same set; $-W_{ij}$ is cost of assigning to different sets

dual function

$$\begin{aligned} g(\nu) = \inf_x (x^T W x + \sum_i \nu_i (x_i^2 - 1)) &= \inf_x x^T (W + \mathbf{diag}(\nu)) x - \mathbf{1}^T \nu \\ &= \begin{cases} -\mathbf{1}^T \nu & W + \mathbf{diag}(\nu) \succeq 0 \\ -\infty & \text{otherwise} \end{cases} \end{aligned}$$

lower bound property: $p^\star \geq -\mathbf{1}^T \nu$ if $W + \mathbf{diag}(\nu) \succeq 0$

example: $\nu = -\lambda_{\min}(W)\mathbf{1}$ gives bound $p^\star \geq n\lambda_{\min}(W)$

Semidefinite program

primal SDP ($\boldsymbol{F}_i$, $\boldsymbol{G} \in \mathbf{S}^k$, $\boldsymbol{A} = [\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n] \in \mathbf{R}^{p \times n}$)

$$\begin{array}{ll} \underset{\boldsymbol{x} \in \mathbf{R}^n}{\text{minimize}} & \boldsymbol{c}^T \boldsymbol{x} \\ \text{subject to} & x_1 \boldsymbol{F}_1 + x_2 \boldsymbol{F}_2 + \cdots + x_n \boldsymbol{F}_n + \boldsymbol{G} \preceq 0 \\ & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \end{array}$$

- $\boldsymbol{Z} \in \mathbf{S}^k$: Lagrange multiplier for the constraint $x_1 \boldsymbol{F}_1 + \cdots + x_n \boldsymbol{F}_n + \boldsymbol{G} \preceq 0$
- $\boldsymbol{\nu} \in \mathbf{R}^p$: Lagrange multipliers for $\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} = 0$

- Lagrangian

$$\begin{aligned} L(\boldsymbol{x}, \boldsymbol{Z}, \boldsymbol{\nu}) &= \boldsymbol{c}^T \boldsymbol{x} + \text{tr}(\boldsymbol{Z}(x_1 \boldsymbol{F}_1 + \cdots + x_n \boldsymbol{F}_n + \boldsymbol{G})) + \boldsymbol{\nu}^T (\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}) \\ &= \sum_{i=1}^{n} (c_i + \text{tr}(\boldsymbol{Z}\boldsymbol{F}_i) + \boldsymbol{\nu}^T \boldsymbol{a}_i) x_i + \text{tr}(\boldsymbol{Z}\boldsymbol{G}) - \boldsymbol{\nu}^T \boldsymbol{b} \end{aligned}$$

- dual function

$$g(\boldsymbol{Z}, \boldsymbol{\nu}) = \inf_{\boldsymbol{x} \in \mathbf{R}^n} L(\boldsymbol{x}, \boldsymbol{Z}, \boldsymbol{\nu}) = \begin{cases} \text{tr}(\boldsymbol{Z}\boldsymbol{G}) - \boldsymbol{\nu}^T \boldsymbol{b} & c_i + \text{tr}(\boldsymbol{Z}\boldsymbol{F}_i) + \boldsymbol{\nu}^T \boldsymbol{a}_i = 0, \ \forall i \\ -\infty & \text{otherwise} \end{cases}$$

- lower bound property: $p^\star \geq \text{tr}(\boldsymbol{Z}\boldsymbol{G}) - \boldsymbol{\nu}^T \boldsymbol{b}$, if $\boldsymbol{Z} \succeq 0$ and $c_i + \text{tr}(\boldsymbol{Z}\boldsymbol{F}_i) + \boldsymbol{\nu}^T \boldsymbol{a}_i = 0$

Lagrange dual and conjugate function

$$\begin{array}{ll} \underset{\boldsymbol{x}\in\mathbf{R}^n}{\text{minimize}} & f_0(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}, \\ & \boldsymbol{C}\boldsymbol{x} = \boldsymbol{d} \end{array}$$

dual function

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

- $f^*(\boldsymbol{y}) = \sup_{\boldsymbol{x}\in\textbf{dom}\, f}(\boldsymbol{y}^T\boldsymbol{x} - f(\boldsymbol{x}))$ is the **conjugate function** of f
- simplifies derivation of dual if conjugate of f_0 is known

example: entropy maximization

$$f_0(\boldsymbol{x}) = \sum_{i=1}^{n} x_i \log x_i, \qquad f_0^*(\boldsymbol{y}) = \sum_{i=1}^{n} e^{y_i - 1}$$

The dual problem

Lagrange dual problem

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

- finds best lower bound on $p^\star$, obtained from Lagrange dual function
- a convex optimization problem (g is concave); optimal value denoted $d^\star$
- $\boldsymbol{\lambda}, \boldsymbol{\nu}$ are (dual) feasible, if $\boldsymbol{\lambda} \geq \mathbf{0}$, $(\boldsymbol{\lambda}, \boldsymbol{\nu}) \in \mathbf{dom}\, g$
- often simplified by making implicit constraint $(\boldsymbol{\lambda}, \boldsymbol{\nu}) \in \mathbf{dom}\, g$ explicit

example standard form LP and its dual

$$\begin{array}{ll} \underset{\boldsymbol{x} \in \mathbf{R}^n}{\text{minimize}} & \boldsymbol{c}^T \boldsymbol{x} \\ \text{subject to} & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \\ & \boldsymbol{x} \geq \mathbf{0} \end{array} \qquad \begin{array}{ll} \underset{\boldsymbol{\nu} \in \mathbf{R}^p}{\text{maximize}} & -\boldsymbol{b}^T \boldsymbol{\nu} \\ \text{subject to} & \boldsymbol{A}^T \boldsymbol{\nu} + \boldsymbol{c} \geq \mathbf{0} \end{array}$$

examples standard form SDP and its dual

$$\begin{array}{ll} \underset{\boldsymbol{x}\in\mathbf{R}^n}{\text{minimize}} & \boldsymbol{c}^T\boldsymbol{x} \\ \text{subject to} & x_1\boldsymbol{F}_1+\cdots+x_n\boldsymbol{F}_n+\boldsymbol{G}\preceq 0 \\ & \boldsymbol{A}\boldsymbol{x}=\boldsymbol{b} \end{array}$$

$$\begin{array}{ll} \underset{\boldsymbol{Z},\boldsymbol{\nu}}{\text{maximize}} & \text{tr}(\boldsymbol{Z}\boldsymbol{G})-\boldsymbol{\nu}^T\boldsymbol{b} \\ \text{subject to} & c_i+\text{tr}(\boldsymbol{Z}\boldsymbol{F}_i)+\boldsymbol{\nu}^T\boldsymbol{a}_i=0,\forall i \\ & \boldsymbol{Z}\succeq 0 \end{array}$$

$$\begin{array}{ll} \underset{\boldsymbol{x}\in\mathbf{R}^n}{\text{minimize}} & \boldsymbol{c}^T\boldsymbol{x} \\ \text{subject to} & x_1\boldsymbol{F}_1+\cdots+x_n\boldsymbol{F}_n+\boldsymbol{G}\preceq 0 \end{array}$$

$$\begin{array}{ll} \underset{\boldsymbol{Z}}{\text{maximize}} & \text{tr}(\boldsymbol{Z}\boldsymbol{G}) \\ \text{subject to} & c_i+\text{tr}(\boldsymbol{Z}\boldsymbol{F}_i)=0,\forall i \\ & \boldsymbol{Z}\succeq 0 \end{array}$$

examples

$$\begin{array}{ll} \underset{\boldsymbol{x}\in\mathbf{R}^n}{\text{minimize}} & \boldsymbol{x}^T\boldsymbol{W}\boldsymbol{x} \\ \text{subject to} & x_i^2=1,\ i=1,\ldots,n \end{array}$$

$$\begin{array}{ll} \underset{\boldsymbol{\nu}}{\text{maximize}} & -\mathbf{1}^T\boldsymbol{\nu} \\ \text{subject to} & \boldsymbol{W}+\textbf{diag}(\boldsymbol{\nu})\succeq 0 \end{array}$$

$$\begin{array}{ll} \underset{\boldsymbol{\nu}}{\text{maximize}} & -\mathbf{1}^T\boldsymbol{\nu} \\ \text{subject to} & \boldsymbol{W}+\textbf{diag}(\boldsymbol{\nu})\succeq 0 \end{array}$$

$$\begin{array}{ll} \underset{\boldsymbol{X}\in\mathbf{S}^n}{\text{minimize}} & \text{tr}(\boldsymbol{W}\boldsymbol{X}) \\ \text{subject to} & \text{tr}(\boldsymbol{e}_i\boldsymbol{e}_i^T\boldsymbol{X})=1,\ i=1,\ldots,n \\ & \boldsymbol{X}\succeq 0 \end{array}$$

Weak and strong duality

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

- always holds (for convex and nonconvex problems)
- can be used to nontrivial lower bounds for difficult problem, for example, solving the SDP

$$\begin{array}{ll} \underset{\boldsymbol{\nu}}{\text{maximize}} & -\mathbf{1}^T\boldsymbol{\nu} \\ \text{subject to} & \boldsymbol{W} + \textbf{diag}\,(\boldsymbol{\nu}) \succeq 0 \end{array}$$

gives a lower bound for the two-way partitioning problem

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

- does not hold in general, and the quantity $p^\star - d^\star$ is called **duality gap**
- (usually) holds for convex problems
- conditions that guarantee strong duality in convex problems are called **constraint qualifications**

Slater's constraint qualification

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.*,

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

- also guarantees that the dual optimum is attained (if $p^\star > -\infty$)
- can be sharpened: *e.g.*, can replace $\textbf{int}\, \mathcal{D}$ with $\textbf{relint}\, \mathcal{D}$ (interior relative to affine hull); linear inequalities do not need to hold with strict inequality, . . .
- there exist many other types of constraint qualifications

Inequality and equality form LP

primal problem

$$\begin{array}{ll} \underset{\boldsymbol{x} \in \mathbf{R}^n}{\text{minimize}} & \boldsymbol{c}^T \boldsymbol{x} \\ \text{subject to} & \boldsymbol{G}\boldsymbol{x} \leq \boldsymbol{h} \\ & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \end{array}$$

dual function

$$\begin{aligned} g(\boldsymbol{\lambda}, \boldsymbol{\nu}) &= \inf_{\boldsymbol{x} \in \mathbf{R}^n} (\boldsymbol{c}^T \boldsymbol{x} + \boldsymbol{\lambda}^T (\boldsymbol{G}\boldsymbol{x} - \boldsymbol{h}) + \boldsymbol{\nu}^T (\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})) \\ &= \inf_{\boldsymbol{x} \in \mathbf{R}^n} ((\boldsymbol{c} + \boldsymbol{G}^T \boldsymbol{\lambda} + \boldsymbol{A}^T \boldsymbol{\nu})^T \boldsymbol{x} - \boldsymbol{h}^T \boldsymbol{\lambda} - \boldsymbol{b}^T \boldsymbol{\nu}) \\ &= \begin{cases} -\boldsymbol{h}^T \boldsymbol{\lambda} - \boldsymbol{b}^T \boldsymbol{\nu} & \boldsymbol{c} + \boldsymbol{G}^T \boldsymbol{\lambda} + \boldsymbol{A}^T \boldsymbol{\nu} = \boldsymbol{0} \\ -\infty & \text{otherwise} \end{cases} \end{aligned}$$

dual problem

$$\begin{array}{ll} \underset{\boldsymbol{\lambda} \in \mathbf{R}^m, \boldsymbol{\nu} \in \mathbf{R}^p}{\text{maximize}} & -\boldsymbol{h}^T \boldsymbol{\lambda} - \boldsymbol{b}^T \boldsymbol{\nu} \\ \text{subject to} & \boldsymbol{c} + \boldsymbol{G}^T \boldsymbol{\lambda} + \boldsymbol{A}^T \boldsymbol{\nu} = \boldsymbol{0} \\ & \boldsymbol{\lambda} \geq \boldsymbol{0} \end{array}$$

- from Slater's condition: $p^\star = d^\star$ if $\boldsymbol{G}\boldsymbol{x}_0 < \boldsymbol{h}$ for some $\boldsymbol{x}_0$

- in fact, $p^\star = d^\star$ except when primal and dual are infeasible.

- generally, strong duality for LP requires no Slater's condition, and:
  - if both primal and dual are infeasible, then $p^\star = +\infty > -\infty = d^\star$
  - if both primal and dual are feasible, then $-\infty < p^\star = d^\star < \infty)$
  - if primal is feasible and dual is infeasible, then $p^\star = d^\star = -\infty$ (i.e, primal is unbounded below)
  - if primal is infeasible and dual is feasible, then $p^\star = d^\star = +\infty$ (i.e., dual is unbounded above)

dual of dual of LP

primal problem

$$\begin{array}{ll} \underset{\boldsymbol{x}\in\mathbf{R}^n}{\text{minimize}} & \boldsymbol{c}^T\boldsymbol{x} \\ \text{subject to} & \boldsymbol{G}\boldsymbol{x} \leq \boldsymbol{h} \\ & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \end{array}$$

dual problem

$$\begin{array}{ll} \underset{\boldsymbol{\lambda}\in\mathbf{R}^m,\, \boldsymbol{\nu}\in\mathbf{R}^p}{\text{maximize}} & -\boldsymbol{h}^T\boldsymbol{\lambda} - \boldsymbol{b}^T\boldsymbol{\nu} \\ \text{subject to} & \boldsymbol{c} + \boldsymbol{G}^T\boldsymbol{\lambda} + \boldsymbol{A}^T\boldsymbol{\nu} = \mathbf{0} \\ & \boldsymbol{\lambda} \geq \mathbf{0} \end{array}$$

equivalent dual problem

$$- \quad \begin{array}{ll} \underset{\boldsymbol{\lambda}\in\mathbf{R}^m,\, \boldsymbol{\nu}\in\mathbf{R}^p}{\text{minimize}} & \boldsymbol{h}^T\boldsymbol{\lambda} + \boldsymbol{b}^T\boldsymbol{\nu} \\ \text{subject to} & \boldsymbol{c} + \boldsymbol{G}^T\boldsymbol{\lambda} + \boldsymbol{A}^T\boldsymbol{\nu} = \mathbf{0} \\ & \boldsymbol{\lambda} \geq \mathbf{0} \end{array}$$

compute the dual problem of LP:

$$\begin{array}{ll} \underset{\boldsymbol{\lambda}\in\mathbf{R}^m,\boldsymbol{\nu}\in\mathbf{R}^p}{\text{minimize}} & \boldsymbol{h}^T\boldsymbol{\lambda} + \boldsymbol{b}^T\boldsymbol{\nu} \\ \text{subject to} & \boldsymbol{c} + \boldsymbol{G}^T\boldsymbol{\lambda} + \boldsymbol{A}^T\boldsymbol{\nu} = \mathbf{0} \\ & \boldsymbol{\lambda} \geq \mathbf{0} \end{array}$$

dual function

$$\begin{aligned} g(\boldsymbol{x},\boldsymbol{y}) &= \inf_{\boldsymbol{\lambda}\in\mathbf{R}^m,\boldsymbol{\nu}\in\mathbf{R}^p} (\boldsymbol{h}^T\boldsymbol{\lambda} + \boldsymbol{b}^T\boldsymbol{\nu} + \boldsymbol{x}^T(\boldsymbol{c} + \boldsymbol{G}^T\boldsymbol{\lambda} + \boldsymbol{A}^T\boldsymbol{\nu}) + \boldsymbol{y}^T(-\boldsymbol{\lambda})) \\ &= \inf_{\boldsymbol{\lambda},\boldsymbol{\nu}} ((\boldsymbol{h} + \boldsymbol{G}\boldsymbol{x} - \boldsymbol{y})^T\boldsymbol{\lambda} + (\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b})^T\boldsymbol{\nu} + \boldsymbol{c}^T\boldsymbol{x} \\ &= \begin{cases} \boldsymbol{c}^T\boldsymbol{x} & \boldsymbol{h} + \boldsymbol{G}\boldsymbol{x} - \boldsymbol{y} = \mathbf{0},\ \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b} = \mathbf{0} \\ -\infty & \text{otherwise} \end{cases} \end{aligned}$$

dual problem

$$\begin{array}{ll} \underset{\boldsymbol{x},\boldsymbol{y}}{\text{maximize}} & \boldsymbol{c}^T\boldsymbol{x} \\ \text{subject to} & \boldsymbol{G}\boldsymbol{x}+\boldsymbol{h}=\boldsymbol{y} \\ & \boldsymbol{A}\boldsymbol{x}+\boldsymbol{b}=\boldsymbol{0} \\ & \boldsymbol{y}\geq \boldsymbol{0} \end{array}$$

removing $\boldsymbol{y}$, it is equivalent to

$$\begin{array}{ll} \underset{\boldsymbol{x}}{\text{maximize}} & \boldsymbol{c}^T\boldsymbol{x} \\ \text{subject to} & \boldsymbol{G}\boldsymbol{x}+\boldsymbol{h}\geq \boldsymbol{0} \\ & \boldsymbol{A}\boldsymbol{x}+\boldsymbol{b}=\boldsymbol{0} \end{array}$$

Therefore, the dual of problem:

$$\begin{array}{lll} - & \underset{\boldsymbol{\lambda}\in\mathbf{R}^m,\boldsymbol{\nu}\in\mathbf{R}^p}{\text{minimize}} & \boldsymbol{h}^T\boldsymbol{\lambda}+\boldsymbol{b}^T\boldsymbol{\nu} \\ & \text{subject to} & \boldsymbol{c}+\boldsymbol{G}^T\boldsymbol{\lambda}+\boldsymbol{A}^T\boldsymbol{\nu}=\boldsymbol{0} \\ & & \boldsymbol{\lambda}\geq\boldsymbol{0} \end{array}$$

is

$$-\begin{array}{ll} \underset{\boldsymbol{x}}{\text{maximize}} & \boldsymbol{c}^T\boldsymbol{x} \\ \text{subject to} & \boldsymbol{G}\boldsymbol{x} + \boldsymbol{h} \geq \boldsymbol{0} \\ & \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b} = \boldsymbol{0} \end{array}$$

by setting $\boldsymbol{z} := -\boldsymbol{x}$, the above problem is equivalent to

$$\begin{array}{ll} \underset{\boldsymbol{z}}{\text{minimize}} & \boldsymbol{c}^T\boldsymbol{z} \\ \text{subject to} & \boldsymbol{G}\boldsymbol{z} \leq \boldsymbol{h} \\ & \boldsymbol{A}\boldsymbol{z} = \boldsymbol{b} \end{array}$$

dual of dual of LP is itself!

- recall dual problem is always convex (primal problem not necessarily convex)

- dual of dual of convex problem is itself, for example

$$\begin{array}{ll} \underset{\boldsymbol{x}\in\mathbf{R}^n}{\text{minimize}} & \boldsymbol{c}^T\boldsymbol{x} \\ \text{subject to} & \boldsymbol{G}\boldsymbol{x} \le \boldsymbol{h} \\ & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \end{array} \qquad \begin{array}{ll} \underset{\boldsymbol{\lambda},\boldsymbol{\nu}}{\text{maximize}} & -\boldsymbol{h}^T\boldsymbol{\lambda} - \boldsymbol{b}^T\boldsymbol{\nu} \\ \text{subject to} & \boldsymbol{c} + \boldsymbol{G}^T\boldsymbol{\lambda} + \boldsymbol{A}^T\boldsymbol{\nu} = \boldsymbol{0} \\ & \boldsymbol{\lambda} \ge \boldsymbol{0} \end{array} \qquad \begin{array}{ll} \underset{\boldsymbol{z}\in\mathbf{R}^n}{\text{minimize}} & \boldsymbol{c}^T\boldsymbol{z} \\ \text{subject to} & \boldsymbol{G}\boldsymbol{z} \le \boldsymbol{h} \\ & \boldsymbol{A}\boldsymbol{z} = \boldsymbol{b} \end{array}$$

- dual of dual of nonconvex problem is a convex relaxation, for example

$$\begin{array}{ll} \underset{\boldsymbol{x}\in\mathbf{R}^n}{\min} & \boldsymbol{x}^T\boldsymbol{W}\boldsymbol{x} \\ \text{s.t.} & x_i^2 = 1,\, \forall i \end{array} \qquad \begin{array}{ll} \underset{\boldsymbol{\nu}}{\max} & -\mathbf{1}^T\boldsymbol{\nu} \\ \text{s.t.} & \boldsymbol{W} + \textbf{diag}\,(\boldsymbol{\nu}) \succeq 0 \end{array} \qquad \begin{array}{ll} \underset{\boldsymbol{X}}{\min} & \text{tr}(\boldsymbol{W}\boldsymbol{X}) \\ \text{s.t.} & \text{tr}(\boldsymbol{e}_i\boldsymbol{e}_i^T\boldsymbol{W}) = 1,\, \forall i \\ & \boldsymbol{X} \succeq 0 \end{array}$$

Quadratic program

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

Semidefinite program

primal SDP ($F_i, G \in \mathbf{S}^k$)

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & x_1 F_1 + \cdots + x_n F_n \preceq G \end{array}$$

- Lagrange multiplier is matrix $Z \in \mathbf{S}^k$
- Lagrangian $L(x, Z) = c^T x + \mathbf{tr}\left(Z(x_1 F_1 + \cdots + x_n F_n - G)\right)$
- dual function

$$g(Z) = \inf_x L(x, Z) = \begin{cases} -\mathbf{tr}(GZ) & \mathbf{tr}(F_i Z) + c_i = 0, \quad i = 1, \ldots, n \\ -\infty & \text{otherwise} \end{cases}$$

dual SDP

$$\begin{array}{ll} \text{maximize} & -\mathbf{tr}(GZ) \\ \text{subject to} & Z \succeq 0, \quad \mathbf{tr}(F_i Z) + c_i = 0, \quad i = 1, \ldots, n \end{array}$$

$p^\star = d^\star$ if primal SDP is strictly feasible ($\exists x$ with $x_1 F_1 + \cdots + x_n F_n \prec G$)

A nonconvex problem with strong duality

$$\begin{array}{ll} \underset{\boldsymbol{x}}{\text{minimize}} & \boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} + 2\boldsymbol{b}^T\boldsymbol{x} \\ \text{subject to} & \boldsymbol{x}^T\boldsymbol{x} \leq 1 \end{array}$$

A is any given symmetric matrix, hence nonconvex

dual function

$$\begin{aligned} g(\lambda) &= \inf_{\boldsymbol{x}} (\boldsymbol{x}^T(\boldsymbol{A}+\lambda\boldsymbol{I})\boldsymbol{x} + 2\boldsymbol{b}^T\boldsymbol{x} - \lambda) = \inf_{\boldsymbol{x},\, t\geq 0} (\boldsymbol{x}^T(\boldsymbol{A}+\lambda\boldsymbol{I})\boldsymbol{x} + 2\boldsymbol{b}^T\boldsymbol{x} + t - t - \lambda) \\ &= \inf_{\boldsymbol{x},\, t\geq 0} \left(\begin{bmatrix} \boldsymbol{x} \\ 1 \end{bmatrix}^T \begin{bmatrix} \boldsymbol{A}+\lambda\boldsymbol{I} & \boldsymbol{b} \\ \boldsymbol{b}^T & t \end{bmatrix} \begin{bmatrix} \boldsymbol{x} \\ 1 \end{bmatrix} - t - \lambda \right) \end{aligned}$$

Observe that

$$\begin{bmatrix} \boldsymbol{C} & \boldsymbol{c} \\ \boldsymbol{c}^T & c \end{bmatrix} \succeq 0 \Leftrightarrow \begin{bmatrix} \boldsymbol{x} \\ 1 \end{bmatrix}^T \begin{bmatrix} \boldsymbol{C} & \boldsymbol{c} \\ \boldsymbol{c}^T & c \end{bmatrix} \begin{bmatrix} \boldsymbol{x} \\ 1 \end{bmatrix} = \boldsymbol{x}^T\boldsymbol{C}\boldsymbol{x} + 2\boldsymbol{c}^T\boldsymbol{x} + c \geq 0,\ \forall \boldsymbol{x}$$

Thus

$$
\begin{aligned}
g(\lambda) &= \inf_{\boldsymbol{x},\, t\geq 0} \left(\begin{bmatrix} \boldsymbol{x} \\ 1 \end{bmatrix}^T \begin{bmatrix} \boldsymbol{A}+\lambda\boldsymbol{I} & \boldsymbol{b} \\ \boldsymbol{b}^T & t \end{bmatrix} \begin{bmatrix} \boldsymbol{x} \\ 1 \end{bmatrix} - t - \lambda \right) \\
&= \begin{cases} -t-\lambda & \begin{bmatrix} \boldsymbol{A}+\lambda\boldsymbol{I} & \boldsymbol{b} \\ \boldsymbol{b}^T & t \end{bmatrix} \succeq 0 \\ -\infty & \text{otherwise} \end{cases}
\end{aligned}
$$

dual problem

$$
\begin{array}{ll}
\underset{t,\lambda}{\text{maximize}} & -t-\lambda \\
\text{subject to} & \begin{bmatrix} \boldsymbol{A}+\lambda\boldsymbol{I} & \boldsymbol{b} \\ \boldsymbol{b}^T & t \end{bmatrix} \succeq 0 \\
& \lambda \geq 0
\end{array}
$$

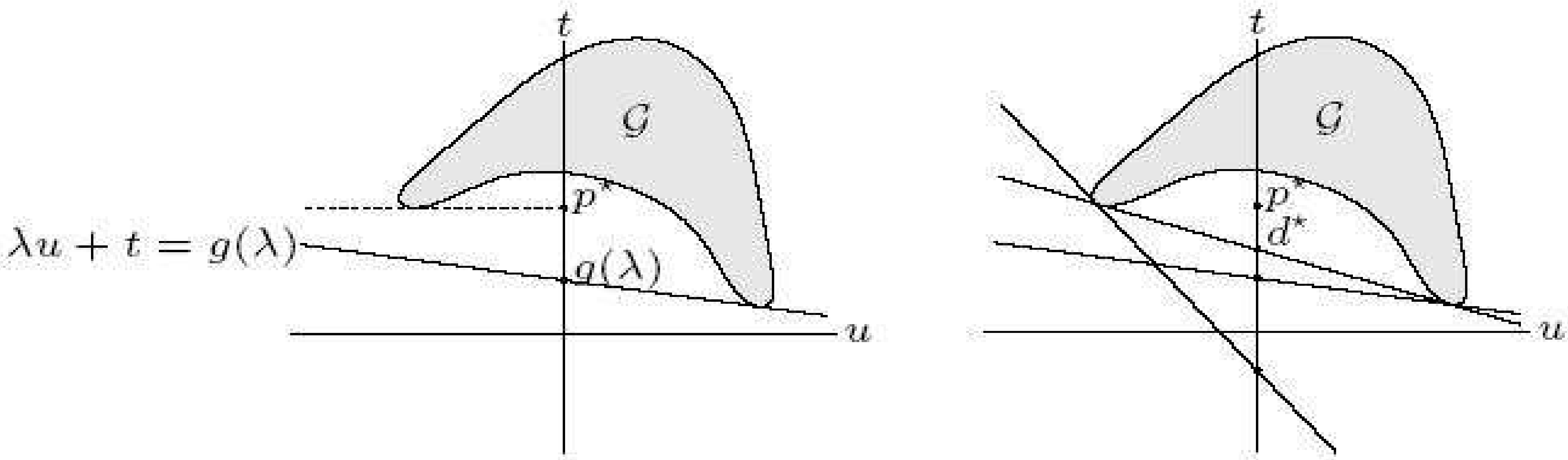
strong duality holds although primal problem is not convex (not easy to prove)

Geometric interpretation

for simplicity, consider problem with one constraint $f_1(x) \leq 0$

interpretation of dual function:

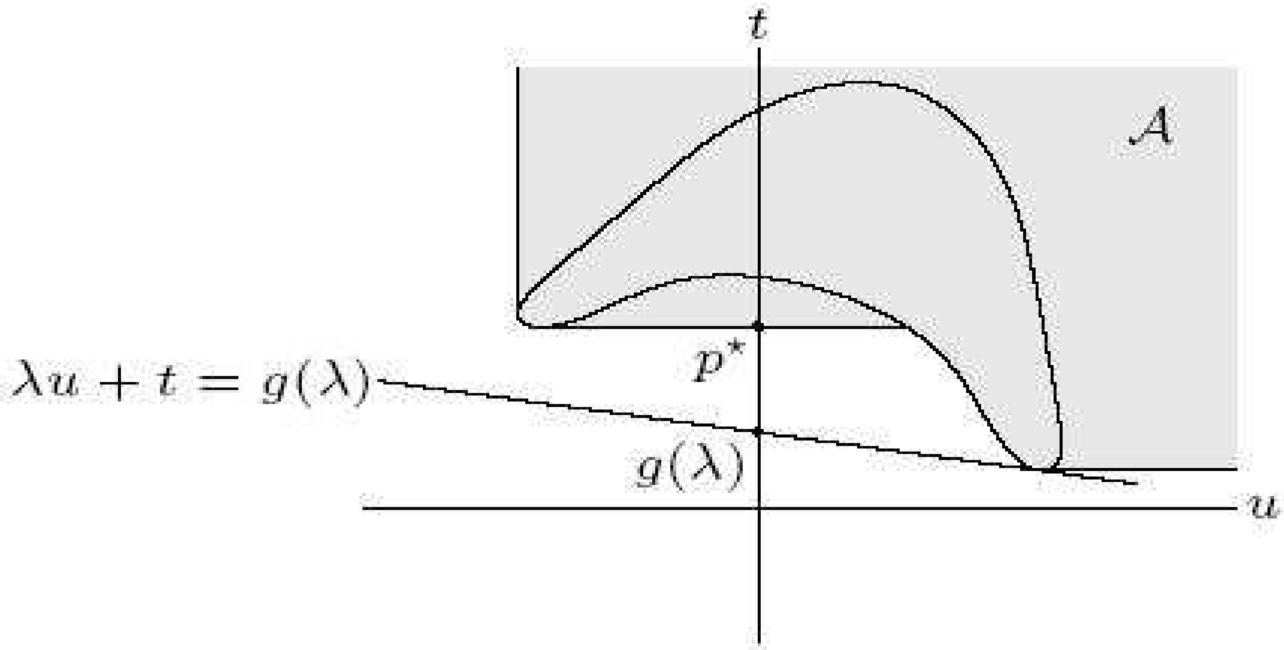
$$g(\lambda) = \inf_{(u,t)\in\mathcal{G}} (t + \lambda u), \qquad \text{where} \quad \mathcal{G} = \{(f_1(x), f_0(x)) \mid x \in \mathcal{D}\}$$



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

epigraph variation: same interpretation if $\mathcal{G}$ is replaced with

$$\mathcal{A} = \{(u, t) \mid f_1(x) \leq u, f_0(x) \leq t \text{ for some } x \in \mathcal{D}\}$$



strong duality

- holds if there is a non-vertical supporting hyperplane to $\mathcal{A}$ at $(0, p^\star)$
- for convex problem, $\mathcal{A}$ is convex, hence has supp. hyperplane at $(0, p^\star)$
- Slater's condition: if there exist $(\tilde{u}, \tilde{t}) \in \mathcal{A}$ with $\tilde{u} < 0$, then supporting hyperplanes at $(0, p^\star)$ must be non-vertical

Complementary slackness

assume strong duality holds, $x^\star$ is primal optimal, $(\lambda^\star, \nu^\star)$ is dual optimal

$$\begin{aligned} f_0(x^\star) = g(\lambda^\star, \nu^\star) &= \inf_x \left(f_0(x) + \sum_{i=1}^m \lambda_i^\star f_i(x) + \sum_{i=1}^p \nu_i^\star h_i(x) \right) \\ &\le f_0(x^\star) + \sum_{i=1}^m \lambda_i^\star f_i(x^\star) + \sum_{i=1}^p \nu_i^\star h_i(x^\star) \\ &\le f_0(x^\star) \end{aligned}$$

hence, the two inequalities hold with equality

- $x^\star$ minimizes $L(x, \lambda^\star, \nu^\star)$
- $\lambda_i^\star f_i(x^\star) = 0$ for $i = 1, \ldots, m$ (known as complementary slackness):

$$\lambda_i^\star > 0 \Longrightarrow f_i(x^\star) = 0, \qquad f_i(x^\star) < 0 \Longrightarrow \lambda_i^\star = 0$$

Karush-Kuhn-Tucker (KKT) conditions

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

- primal constraints: $f_i(\boldsymbol{x}) \leq 0$, $i = 1, \ldots, m$, $h_i(\boldsymbol{x}) = 0$, $i = 1, \ldots, p$
- dual constraints: $\boldsymbol{\lambda} \geq \mathbf{0}$
- complementary slackness: $\lambda_i f_i(\boldsymbol{x}) = 0$, $i = 1, \ldots, m$
- gradient of Lagrangian with respect to $\boldsymbol{x}$ vanishes:

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

from the last page: if strong duality holds and $\boldsymbol{x}$, $\boldsymbol{\lambda}$, $\boldsymbol{\nu}$ are optimal, then they must satisfy the KKT conditions

KKT conditions for convex problem

if $\tilde{x}$, $\tilde{\lambda}$, $\tilde{\nu}$ satisfy KKT for a convex problem, then they are optimal:

- from complementary slackness: $f_0(\tilde{x}) = L(\tilde{x}, \tilde{\lambda}, \tilde{\nu})$
- from 4th condition (and convexity): $g(\tilde{\lambda}, \tilde{\nu}) = L(\tilde{x}, \tilde{\lambda}, \tilde{\nu})$

hence, $f_0(\tilde{x}) = g(\tilde{\lambda}, \tilde{\nu})$

if **Slater's condition** is satisfied:

x is optimal if and only if there exist λ, ν that satisfy KKT conditions

- recall that Slater implies strong duality, and dual optimum is attained
- generalizes optimality condition $\nabla f_0(x) = 0$ for unconstrained problem

example: water-filling (assume $\alpha_i > 0$)

$$\begin{array}{ll} \text{minimize} & -\sum_{i=1}^n \log(x_i + \alpha_i) \\ \text{subject to} & x \succeq 0, \quad \mathbf{1}^T x = 1 \end{array}$$

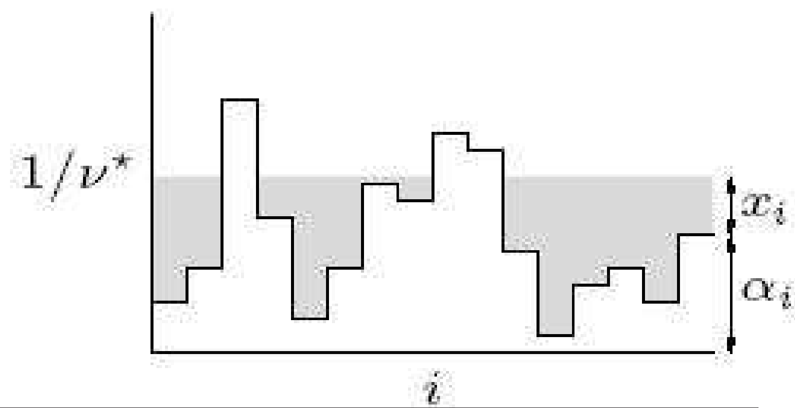
x is optimal iff $x \succeq 0$, $\mathbf{1}^T x = 1$, and there exist $\lambda \in \mathbf{R}^n$, $\nu \in \mathbf{R}$ such that

$$\lambda \succeq 0, \qquad \lambda_i x_i = 0, \qquad \frac{1}{x_i + \alpha_i} + \lambda_i = \nu$$

- if $\nu < 1/\alpha_i$: $\lambda_i = 0$ and $x_i = 1/\nu - \alpha_i$
- if $\nu \geq 1/\alpha_i$: $\lambda_i = \nu - 1/\alpha_i$ and $x_i = 0$
- determine ν from $\mathbf{1}^T x = \sum_{i=1}^n \max\{0, 1/\nu - \alpha_i\} = 1$

interpretation

- n patches; level of patch i is at height α_i
- flood area with unit amount of water
- resulting level is $1/\nu^\star$



Thank You !!

Lecture notes available at
http://www.ece.ust.hk/$\sim$eeyw/opt09/lts.htm