5. Duality

Outline

Lagrangian

- **standard form problem** (not necessarily convex)

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

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

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

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

  - weighted sum of objective and constraint functions
  - λ_i is **Lagrange multiplier** associated with $f_i(x) \leq 0$
  - ν_i is Lagrange multiplier associated with $h_i(x) = 0$

Lagrange dual function

Recall. conjugate $f^*(y) = \sup_x y^T x - f(x)$.

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

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

- g is concave, can be $-\infty$ for some λ, ν
- **lower bound property:** if $\lambda \succeq 0$, then $g(\lambda, \nu) \leq p^\star$

($f_i(x) \leq 0$, $h_i(x) = 0$)

- proof: if $\tilde{x}$ is feasible and $\lambda \succeq 0$, then

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

minimizing over all feasible $\tilde{x}$ gives $p^\star \geq g(\lambda, \nu)$

Least-norm solution of linear equations

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

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

$$\nabla_x L(x, \nu) = 2x + A^T \nu = 0 \quad \Longrightarrow \quad x = -(1/2) A^T \nu$$

- plug x into L to obtain

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

$x^T x = \frac{1}{4} \upsilon^T A A^T \upsilon$.

$\upsilon^T A x = -\frac{1}{2} \upsilon^T A A^T \upsilon$

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

Sanity check: when $A = I$.

$Ax = b \Leftrightarrow x = b$.

So $p^* = b^T b = \|b\|_2^2$.

$-\frac{1}{4} \upsilon^T \upsilon - b^T \upsilon$ is maximized

when $\upsilon = -2b$.

$\Rightarrow -b^T b + 2b^T b = b^T b$.

Standard form LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b, \quad x \succeq 0 \end{array}$$

- Lagrangian is

$$L(x, \lambda, \nu) = c^T x + \nu^T (Ax - b) - \lambda^T x = -b^T \nu + (c + A^T \nu - \lambda)^T x$$

- L is affine in x, so

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

because x can take any value in $\mathbb{R}^n$.

- g is linear on affine domain $\{(\lambda, \nu) \mid A^T \nu - \lambda + c = 0\}$, hence concave
- lower bound property: $p^\star \geq -b^T \nu$ if $A^T \nu + c \succeq 0$ (i.e. $\lambda \geq 0$).

Sanity check: $A = I$.
then $p^* = c^T b$. ($b \geq 0$).
lower bound: $p^* \geq -b^T v$
$v + c \geq 0$.

$\Rightarrow$. $-b^T v = -\sum \underbrace{b_i}_{\geq 0} v_i$, $v_i \geq -c_i$.
decrease in each v_i.
$\Rightarrow$. max when $v_i = -c_i$.

Equality constrained norm minimization

$$\inf_x \|x\| - v^T A x$$
$$= \sup_x \; v^T A x - \|x\|$$
$$= \sup_x \; \|x\| \left(v^T A \frac{x}{\|x\|} - 1 \right).$$
$$= \sup_\alpha \alpha \left(\underbrace{\sup_{\|z\|=1} (v^T A z)}_{\|A^T v\|_*} - 1 \right).$$
$$\left(z = \frac{x}{\|x\|} \right)$$

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

- dual function is

$\|x\| + v^T(b - Ax) = \underline{\|x\| - v^T A x} + v^T b.$

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

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

(dual norm of 2-norm is 2 norm
dual norm of l_1-norm is ∞-norm)

- lower bound property: $p^\star \ge b^T \nu$ if $\|A^T \nu\|_* \le 1$

Sanity check: $A = I$. then $p^* = \|b\|$.

$$\max_{\|v\|_* \le 1} b^T v = \frac{b^T b}{\|b\|} = \|b\|$$

Two-way partitioning

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

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

e.g. spin glass, gene interaction etc.

- dual function is

$\sum_i v_i x_i^2 = x^T \mathrm{diag}(v)\, x.$

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

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

Sanity check: $W = I$.
then. $I + \mathrm{diag}(v) \succeq 0$
$\max_v -1^T v$ when $v = -1$
$\quad = n. = p^*$

Lagrange dual and conjugate function

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

- dual function

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

  where $f^*(y) = \sup_{x \in \mathbf{dom} f}(y^T x - f(x))$ is conjugate of f_0
- simplifies derivation of dual if conjugate of f_0 is known
- **example: entropy maximization**

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

Outline

The Lagrange dual problem

(Lagrange) **dual problem**

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

- finds best lower bound on $p^\star$, obtained from Lagrange dual function g, concave.
- a convex optimization problem, even if original **primal** problem is not
- dual optimal value denoted $d^\star$
- λ, ν are dual feasible if $\lambda \succeq 0$, $(\lambda, \nu) \in \mathbf{dom}\, g$
- often simplified by making implicit constraint $(\lambda, \nu) \in \mathbf{dom}\, g$ explicit

Example: standard form LP

(see slide 5.5)

- primal standard form LP:

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & x \succeq 0 \end{array}$$

- dual problem is

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

  with $g(\lambda, \nu) = -b^T \nu$ if $A^T \nu - \lambda + c = 0$, $-\infty$ otherwise

- make implicit constraint explicit, and eliminate λ to obtain (transformed) dual problem

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

Weak and strong duality

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

- always holds (for convex and nonconvex problems)
- can be used to find nontrivial lower bounds for difficult problems, *e.g.*, solving the SDP

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

Primal: $\min \; x^T W x$
s.t. $x_i^2 = 1$

Lagrange dual

$$g(\nu) = \begin{cases} -\mathbf{1}^T \nu & \text{if } W + \mathrm{diag}(\nu) \succeq 0 \\ -\infty & \text{else.} \end{cases}$$

gives a lower bound for the two-way partitioning problem on page 5.7

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

- does not hold in general
- (usually) holds for convex problems
- conditions that guarantee strong duality in convex problems are called **constraint qualifications**

Slater's constraint qualification

strong duality holds for a convex problem

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

if it is **strictly feasible**, *i.e.*, there is an $x \in \textbf{int}\,\mathcal{D}$ with $f_i(x) < 0$, $i = 1, \ldots, m$, $Ax = b$

- 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)
  - affine inequalities do not need to hold with strict inequality
- there are many other types of constraint qualifications

Inequality form LP

primal problem

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

dual function

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

dual problem

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

- from the sharpened Slater's condition: $p^\star = d^\star$ if the primal problem is feasible
- in fact, $p^\star = d^\star$ except when primal and dual are both infeasible

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

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

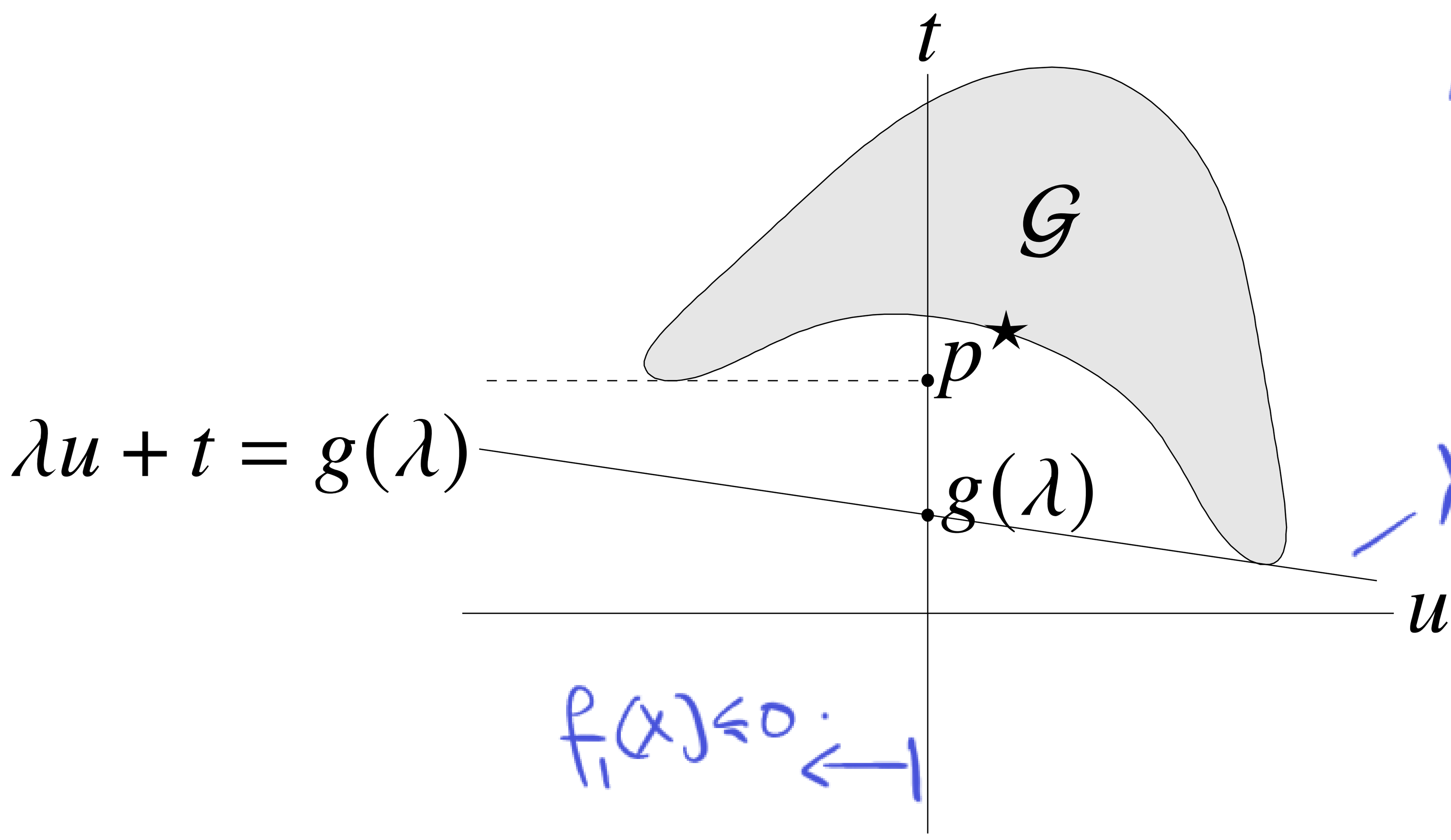
dual problem

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

- from the sharpened Slater's condition: $p^\star = d^\star$ if the primal problem is feasible
- in fact, $p^\star = d^\star$ always (because $p^* > -\infty$ always. In fact $p^* \geq 0$)

Geometric interpretation

- for simplicity, consider problem with one constraint $f_1(x) \leq 0$
- $\mathcal{G} = \{(f_1(x), f_0(x)) \mid x \in \mathcal{D}\}$ is set of achievable (constraint, objective) values
- **interpretation of dual function:** $g(\lambda) = \inf_{(u,t)\in\mathcal{G}} (t + \lambda u)$

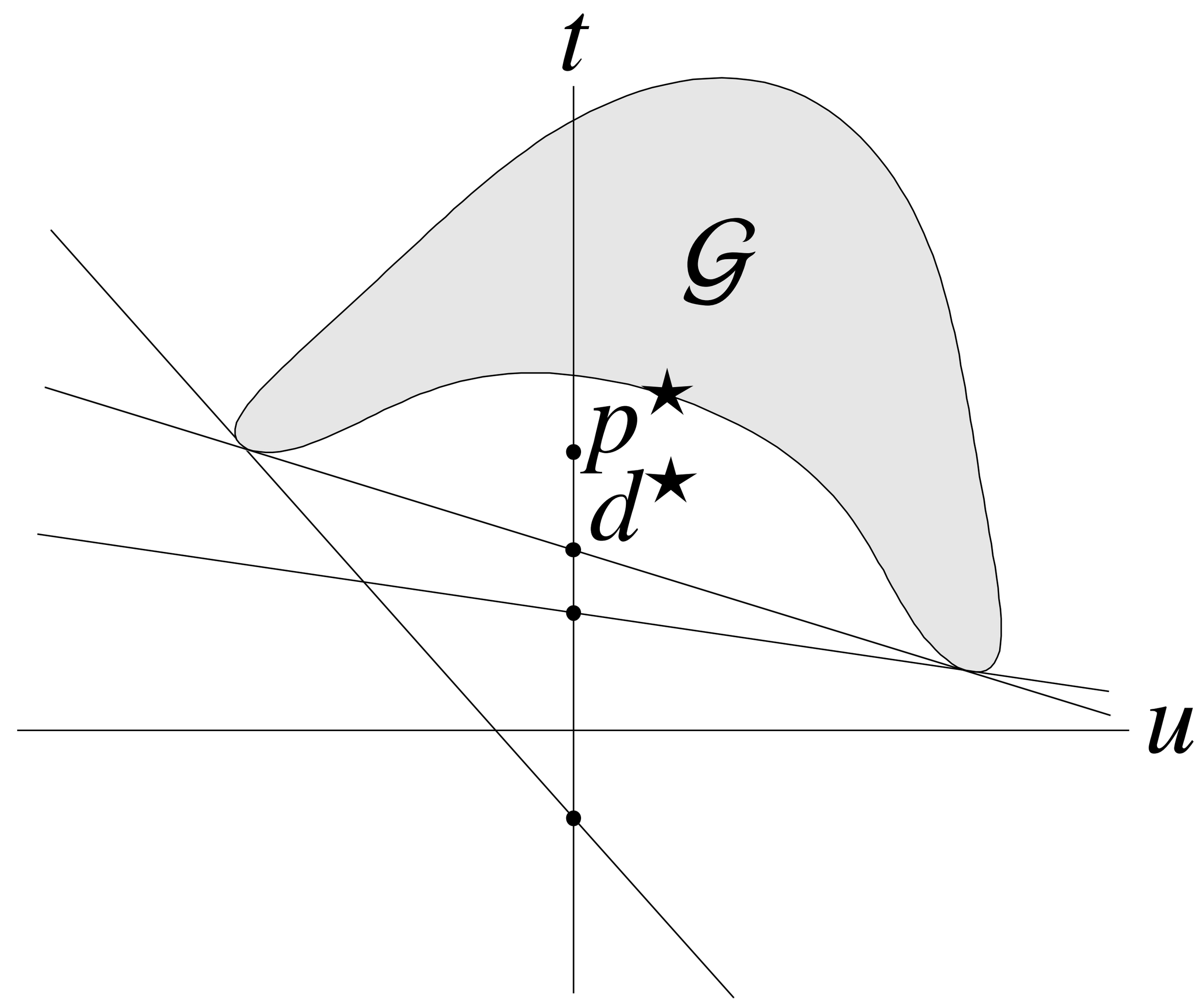


$f_1(x)$

$f_0(x)$.

λ is slope.

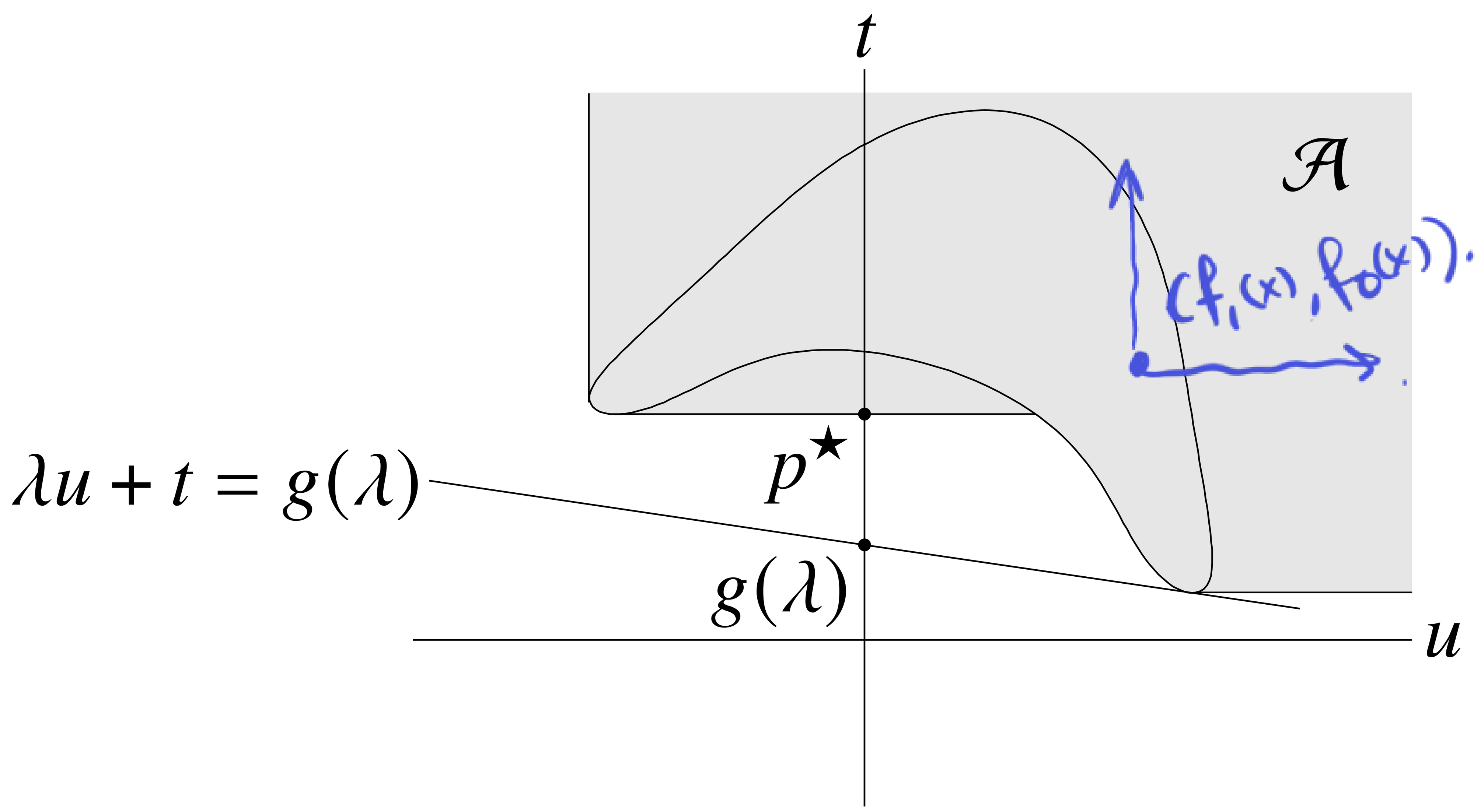
$f_1(x) \leq 0$.



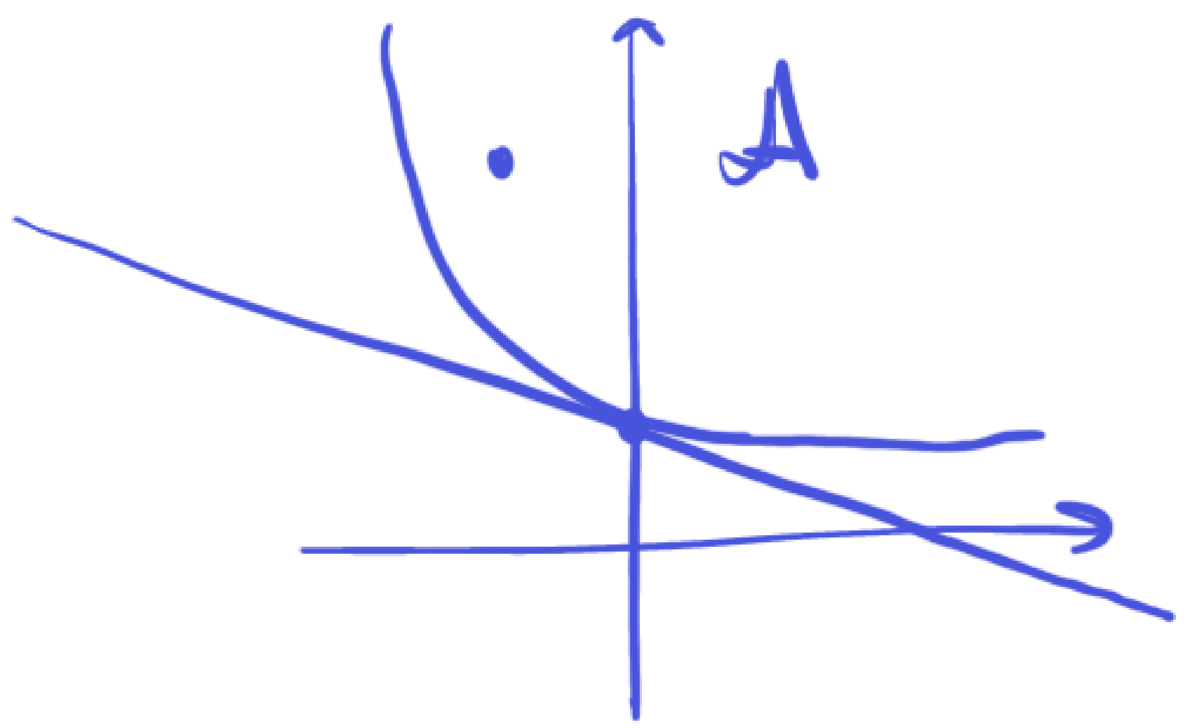
- $\lambda u + t = g(\lambda)$ is (non-vertical) supporting hyperplane to $\mathcal{G}$ (inf. so all points in $\mathcal{G}$ satisfy $\lambda u + t \geq g(\lambda)$)
- hyperplane intersects t-axis at $t = g(\lambda)$

Epigraph variation

- same with $\mathcal{G}$ replaced with $\mathcal{A} = \{(u,t) \mid f_1(x) \le u,\ f_0(x) \le t \text{ for some } x \in \mathcal{D}\}$



- 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 supporting hyperplane at $(0, p^\star)$
- Slater's condition: if there exist $(\tilde{u}, \tilde{t}) \in \mathcal{A}$ with $\tilde{u} < 0$, then supporting hyperplane at $(0, p^\star)$ must be non-vertical



Outline

Complementary slackness

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

$$
\begin{aligned}
f_0(x^\star) = g(\lambda^\star, \nu^\star) &= \inf_x \left(f_0(x) + \sum_{i=1}^m \lambda_i^\star f_i(x) + \sum_{i=1}^p \nu_i^\star h_i(x) \right) \\
&\le f_0(x^\star) + \sum_{i=1}^m \underbrace{\lambda_i^\star}_{\geq 0} \underbrace{f_i(x^\star)}_{\leq 0} + \sum_{i=1}^p \nu_i^\star \underbrace{h_i(x^\star)}_{= 0.} \\
&\le f_0(x^\star)
\end{aligned}
$$

- 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**):

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

Karush-Kuhn-Tucker (KKT) conditions

the **KKT conditions** (for a problem with differentiable f_i, h_i) are

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

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

if strong duality holds and x, λ, ν are optimal, they satisfy the KKT conditions

KKT conditions for convex problem

$L(x,\lambda,\nu) = f_0(x) + \underbrace{\sum \lambda_i f_i(x)}_{0} + \underbrace{\sum \nu_i h_i(x)}_{0}$

$= f_0(x)$ ③ Complementary slackness ①. primal constraints

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): $g(\tilde{\lambda}, \tilde{\nu}) = L(\tilde{x}, \tilde{\lambda}, \tilde{\nu})$ } $\inf_x f_0(x) + \sum \lambda_i f_i(x) + \sum \nu_i h_i(x)$ is unconstrained convex prob. in x.

hence, $f_0(\tilde{x}) = g(\tilde{\lambda}, \tilde{\nu})$ } achieves lower bound. so. optimal.

if Slater's condition is satisfied, then

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

Outline

Perturbation and sensitivity analysis

(unperturbed) optimization problem and its dual

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \dots, m \\ & h_i(x) = 0, \quad i = 1, \dots, p \end{array} \qquad \begin{array}{ll} \text{maximize} & g(\lambda, \nu) \\ \text{subject to} & \lambda \succeq 0 \end{array}$$

perturbed problem and its dual What happens if we relax / tighten the constraints?

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq u_i, \quad i = 1, \dots, m \\ & h_i(x) = v_i, \quad i = 1, \dots, p \end{array} \qquad \begin{array}{ll} \text{maximize} & g(\lambda, \nu) - u^T\lambda - v^T\nu \\ \text{subject to} & \lambda \succeq 0 \end{array}$$

- x is primal variable; u, v are parameters
- $p^\star(u, v)$ is optimal value as a function of u, v
- $p^\star(0, 0)$ is optimal value of unperturbed problem

Global sensitivity via duality

- assume strong duality holds for unperturbed problem, with $\lambda^\star$, $\nu^\star$ dual optimal
- apply weak duality to perturbed problem:

$$p^\star(u,v) \geq g(\lambda^\star, \nu^\star) - u^T\lambda^\star - v^T\nu^\star = p^\star(0,0) - u^T\lambda^\star - v^T\nu^\star$$

- **implications**
  - if $\lambda_i^\star$ large: $p^\star$ increases greatly if we tighten constraint i ($u_i < 0$) $f_i(x) \leq u$.
  - if $\lambda_i^\star$ small: $p^\star$ does not decrease much if we loosen constraint i ($u_i > 0$)
  - if $\nu_i^\star$ large and positive: $p^\star$ increases greatly if we take $v_i < 0$
  - if $\nu_i^\star$ large and negative: $p^\star$ increases greatly if we take $v_i > 0$
  - if $\nu_i^\star$ small and positive: $p^\star$ does not decrease much if we take $v_i > 0$
  - if $\nu_i^\star$ small and negative: $p^\star$ does not decrease much if we take $v_i < 0$

Local sensitivity via duality

if (in addition) $p^\star(u,v)$ is differentiable at $(0,0)$, then

$$\lambda_i^\star = -\frac{\partial p^\star(0,0)}{\partial u_i}, \qquad \nu_i^\star = -\frac{\partial p^\star(0,0)}{\partial v_i}$$

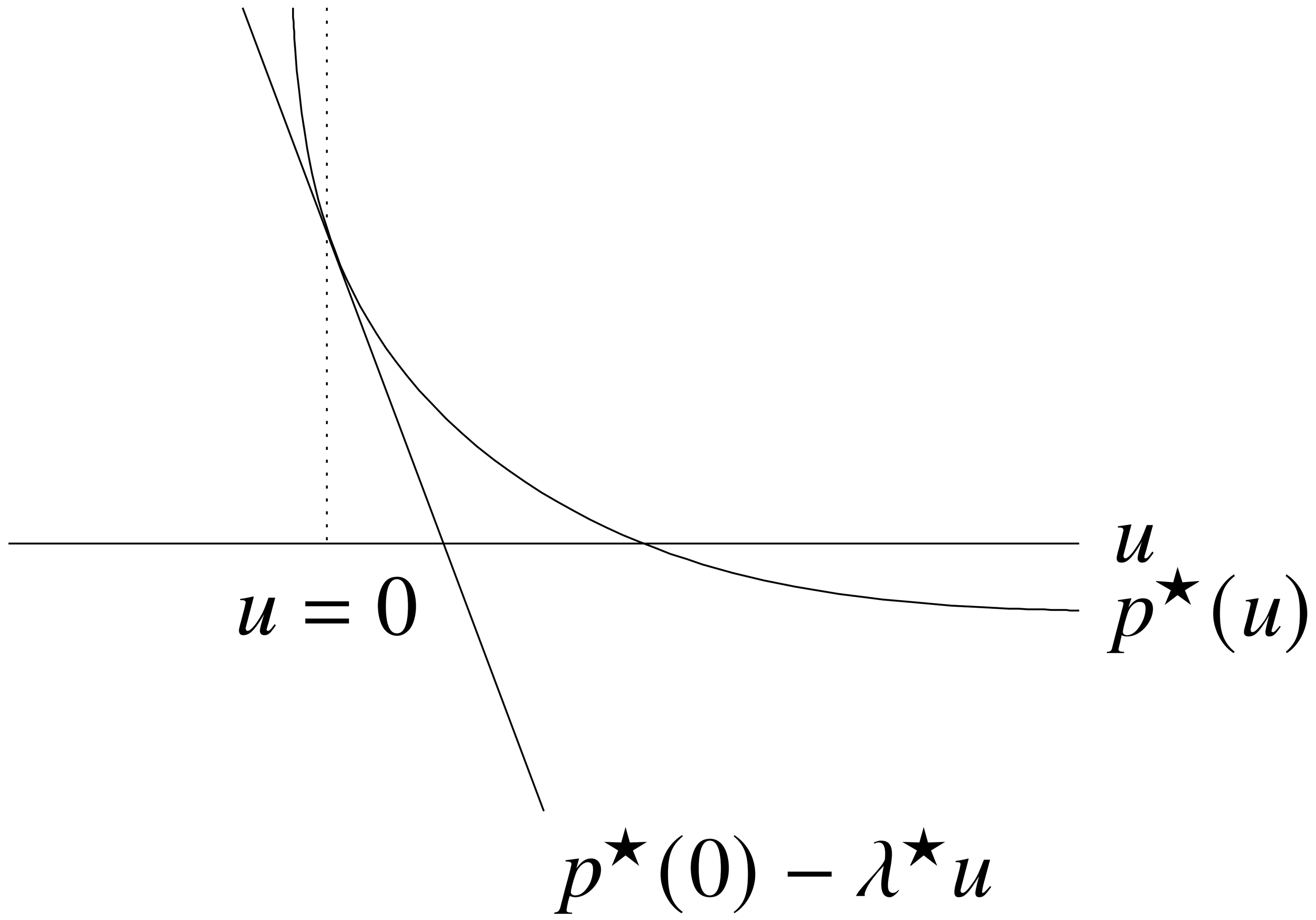
proof (for $\lambda_i^\star$): from global sensitivity result, $p^*(u,v) \geq p^*(0,0) - u^T\lambda^* - v^T\nu$

$$\frac{\partial p^\star(0,0)}{\partial u_i} = \lim_{t\searrow 0} \frac{p^\star(te_i,0) - p^\star(0,0)}{t} \geq -\lambda_i^\star \qquad \frac{\partial p^\star(0,0)}{\partial u_i} = \lim_{t\nearrow 0} \frac{p^\star(te_i,0) - p^\star(0,0)}{t} \leq -\lambda_i^\star$$

$t>0$ $\quad \geq \lim_{t\to 0^+} \frac{t e_i^T \lambda^*}{t}$ $\quad t<0$.

hence, equality

$p^\star(u)$ for a problem with one (inequality) constraint:



Outline

Lagrangian and dual function

Lagrange dual problem

KKT conditions

Sensitivity analysis

Problem reformulations

Theorems of alternatives

Duality and problem reformulations

- equivalent formulations of a problem can lead to very different duals
- reformulating primal problem can be useful when dual is difficult to derive, or uninteresting

common reformulations

- introduce new variables and equality constraints
- make explicit constraints implicit or vice-versa
- transform objective or constraint functions, *e.g.*, replace $f_0(x)$ by $\phi(f_0(x))$ with ϕ convex, increasing

Introducing new variables and equality constraints

- unconstrained problem: minimize $f_0(Ax+b)$
- dual function is constant: $g = \inf_x L(x) = \inf_x f_0(Ax+b) = p^\star$
- we have strong duality, but dual is quite useless

primal.
$\min f_0(x)$
s.t. $Ax=b$

Dual.
$\max -b^Tv - f_0^*(-A^Tv)$

- introduce new variable y and equality constraints $y = Ax+b$

$$\begin{array}{ll} \underset{x,y}{\text{minimize}} & f_0(y) \\ \text{subject to} & Ax+b-y=0 \end{array}$$

- dual of reformulated problem is $g = \inf_{x,y} f_0(y) + v^T(Ax+b-y) = \begin{cases} \inf_y b^Tv + \underbrace{f_0(y) - v^Ty}_{-f_0^*(v)}, & A^Tv = 0. \\ -\infty & A^Tv \neq 0 \end{cases}$

$$\begin{array}{ll} \text{maximize} & b^T\nu - f_0^*(\nu) \\ \text{subject to} & A^T\nu = 0 \end{array}$$

- a nontrivial, useful dual (assuming the conjugate f_0^* is easy to express)

Example: Norm approximation

Primal

$\min_{x,y} f_0(y)$

s.t. $Ax+b=y$

Dual

$\max_{\nu} b^T\nu - f_0^*(\nu)$

s.t. $A^T\nu = 0$

- minimize $\|Ax - b\|$
- reformulate as minimize $\|y\|$ subject to $y = Ax - b$
- recall conjugate of general norm:

$\underbrace{\max_{\|x\|\le 1} x^T z}$

$$\|z\|^* = \begin{cases} 0 & \|z\|_* \le 1 \\ \infty & \text{otherwise} \end{cases}$$

$\|z\|^* = \sup_x x^T z - \|x\|$

$= \sup_x \|x\| \left[\left(\frac{x}{\|x\|}\right)^T z - 1 \right]$

$= \sup_\alpha \alpha \left(\underbrace{\sup_{\|y\|\le 1} y^T z}_{\|z\|_*} - 1 \right).$

$y = \frac{x}{\|x\|}$

- dual of (reformulated) norm approximation problem:

Primal

$\min \|y\|$

s.t. $Ax-b=y$

Dual

$\max_\nu -b^T\nu - \|\nu\|^*$

s.t. $A^T\nu = 0$

$$\begin{array}{ll} \text{maximize} & b^T\nu \\ \text{subject to} & A^T\nu = 0, \quad \|\nu\|_* \le 1 \end{array}$$

Outline

Theorems of alternatives

- consider two systems of inequality and equality constraints
- called **weak alternatives** if no more than one system is feasible
- called **strong alternatives** if exactly one of them is feasible
- examples: for any $a \in \mathbf{R}$, with variable $x \in \mathbf{R}$,
  - $x > a$ and $x \leq a - 1$ are weak alternatives
  - $x > a$ and $x \leq a$ are strong alternatives

- a **theorem of alternatives** states that two inequality systems are (weak or strong) alternatives
- can be considered the extension of duality to feasibility problems

Feasibility problems

- consider system of (not necessarily convex) inequalities and equalities

$$f_i(x) \leq 0, \quad i = 1, \ldots, m, \quad h_i(x) = 0, \quad i = 1, \ldots, p$$

- express as **feasibility problem**

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

- if system if feasible, $p^\star = 0$; if not, $p^\star = \infty$

Duality for feasibility problems

- dual function of feasibility problem is $g(\lambda, \nu) = \inf_x \left(\sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \nu_i h_i(x) \right)$
- for $\lambda \succeq 0$, we have $g(\lambda, \nu) \leq p^\star$
- it follows that feasibility of the inequality system

$$\lambda \succeq 0, \qquad g(\lambda, \nu) > 0$$

  implies the original system is infeasible
- so this is a weak alternative to original system
- it is strong if f_i convex, h_i affine, and a constraint qualification holds
- g is positive homogeneous so we can write alternative system as

first order in λ, ν.

$$\lambda \succeq 0, \qquad g(\lambda, \nu) \geq 1$$

$g(\alpha\lambda, \alpha\nu) = \alpha g(\lambda, \nu)$.

Primal. min 0
s.t. $f_i(x) \leq 0$
$h_i(x) = 0$

Dual max 1
s.t. $\lambda \geq 0$
$g(\lambda, \nu) \geq 1$.

Example: Nonnegative solution of linear equations

Primal.
min 0
s.t. $f_i(x) \le 0$
$h_i(x) = 0$

Dual max 1
s.t. $\lambda \succeq 0$
$g(\lambda, \nu) \ge 1$

- consider system

$$Ax = b, \qquad x \succeq 0$$

- dual function is $g(\lambda, \nu) = \begin{cases} -b^T \nu & A^T \nu = \lambda \\ -\infty & \text{otherwise} \end{cases}$

- can express strong alternative of $Ax = b$, $x \succeq 0$ as

$$A^T \nu \succeq 0, \qquad b^T \nu \le -1$$

(we can replace $b^T \nu \le -1$ with $b^T \nu = -1$)

(ν is feasible $\Rightarrow \frac{\nu}{|b^T \nu|}$ is also feasible)

Farkas' lemma

- Farkas' lemma:

$$Ax \preceq 0, \quad c^T x < 0 \quad \text{and} \quad A^T y + c = 0, \quad y \succeq 0$$

are strong alternatives

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

- proof: use (strong) duality for (feasible) LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \preceq 0 \end{array}$$

$\Rightarrow$ Dual:

$$\begin{array}{ll} \max & 0 \\ \text{s.t.} & A^T y + c = 0. \\ & y \geq 0. \end{array}$$

Investment arbitrage

- we invest x_j in each of n assets $1, \ldots, n$ with prices $p_1, \ldots, p_n$
- our initial cost is $p^T x$
- at the end of the investment period there are only m possible outcomes $i = 1, \ldots, m$
- V_{ij} is the **payoff** (获利) or final value of asset j in outcome i
- first investment is risk-free (cash): $p_1 = 1$ and $V_{i1} = 1$ for all i

套利 no cost, just reward in all outcomes, we win.

- **arbitrage** means there is x with $p^T x < 0$, $Vx \geq 0$
- arbitrage means we receive money up front, and our investment cannot lose
- standard assumption in economics: the prices are such that **there is no arbitrage**

Absence of arbitrage

Arbitrage

$\begin{cases} A^T x \leq 0, \; C^T x < 0 \\ A^T y + C = 0, \; y \geq 0 \end{cases}$

$p^T x < 0$
$Vx \geq 0$

$-V^T y + p = 0.$
$y \geq 0.$

- by Farkas' lemma, there is no arbitrage $\iff$ there exists $y \in \mathbf{R}^m_+$ with $V^T y = p$
- since first column of V is $\mathbf{1}$, we have $\mathbf{1}^T y = 1$ } A probability vector.
- y is interpreted as a **risk-neutral probability** on the outcomes $1, \ldots, m$
- $V^T y$ are the expected values of the payoffs under the risk-neutral probability
- interpretation of $V^T y = p$:

  asset prices equal their expected payoff under the risk-neutral probability

- **arbitrage theorem**: there is no arbitrage $\Leftrightarrow$ there exists a risk-neutral probability distribution under which each asset price is its expected payoff

Example

$$V = \begin{bmatrix} 1.0 & 0.5 & 0.0 \\ 1.0 & 0.8 & 0.0 \\ 1.0 & 1.0 & 1.0 \\ 1.0 & 1.3 & 4.0 \end{bmatrix}, \qquad p = \begin{bmatrix} 1.0 \\ 0.9 \\ 0.3 \end{bmatrix}, \qquad \tilde{p} = \begin{bmatrix} 1.0 \\ 0.8 \\ 0.7 \end{bmatrix}$$

- with prices p, there is an arbitrage

$$x = \begin{bmatrix} 6.2 \\ -7.7 \\ 1.5 \end{bmatrix}, \qquad p^T x = -0.2, \qquad \mathbf{1}^T x = 0, \qquad Vx = \begin{bmatrix} 2.35 \\ 0.04 \\ 0.00 \\ 2.19 \end{bmatrix}$$

- with prices $\tilde{p}$, there is no arbitrage, with risk-neutral probability } proof that there is no arbitrage.

$$y = \begin{bmatrix} 0.36 \\ 0.27 \\ 0.26 \\ 0.11 \end{bmatrix} \qquad V^T y = \begin{bmatrix} 1.0 \\ 0.8 \\ 0.7 \end{bmatrix}$$

Example. KKT Conditions for convex problems.

① Primal / Dual feasible
② Complementary Slackness
③ First order condition on x.

- Equality constrained convex quadratic.

$$\min \ \frac{1}{2} x^T P x + q^T x + r \qquad P \in \mathbb{S}^n_+ .$$
$$\text{s.t.} \quad Ax = b.$$

- KKT: ① $Ax^* = b$ i.e. $[A \ \ 0] \begin{bmatrix} x^* \\ v^* \end{bmatrix} = b$.

③ $\nabla_x L(x, \lambda^*, v^*) = \nabla_x \left(\frac{1}{2} x^T P x + q^T x + r + v^{*T}(Ax - b) \right)$

$= Px + q + A^T v^* . = 0$ at x^*.

i.e. $[P \ \ A^T] \begin{bmatrix} x \\ v^* \end{bmatrix} = -q$

$$\Rightarrow \begin{bmatrix} P & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x^* \\ v^* \end{bmatrix} = \begin{bmatrix} -q \\ b \end{bmatrix}$$

} Solve primal and dual varibles simultaneously.

<u>Example</u>. KKT conditions for convex problems.

① Primal / Dual feasible
② Complementary Slackness
③ First order condition on x.

Water-filling. / information flow.

(1). $\min \; -\sum_{i=1}^{n} \log(\alpha_i + x_i)$, $\alpha_i > 0$

s.t. $\quad x \geq 0$

$\quad\quad 1^T x = 1$

(2). Allocating power to n information channels.

– x_i – power allocated to transmitter at channel i.

$\log(\alpha_i + x_i)$ – capacity of channel i

– Max total communication rate with $1^T x = 1$ total power.

(3).

∘ $L(x, \lambda, \nu) = -\sum_{i=1}^{n} \log(\alpha_i + x_i) - \lambda^T x + \nu(1^T x - 1)$

KKT conditions. ① : $x^* \succeq 0$. $1^T x^* = 1$

$\lambda^* \succeq 0$

② . $\lambda_i^* x_i = 0$. $i = 1, \dots, n$

③ $\frac{\partial}{\partial x_i}$: $-\frac{1}{\alpha_i + x_i^*} - \lambda_i^* + \nu^* = 0 \quad i = 1, \dots, n$.

(4). Solve directly!

λ^* is slack variable. in ①, ③, eliminate.

$\nu^* - \frac{1}{\alpha_i + x_i^*} = \lambda_i^* \geq 0$. , $\lambda_i^* x_i = 0$.

$$\Leftrightarrow \quad \begin{cases} \nu^* \geq \frac{1}{\alpha_i + x_i^*} & ③ \\ 0 = x_i^* \left(\nu^* - \frac{1}{\alpha_i + x_i}\right) & ② \\ x_i^* \geq 0 & \\ 1^T x^* = 1. & ① \end{cases}$$

Example. KKT conditions for convex problems.

① Primal / Dual feasible
② Complementary Slackness
③ First order condition on x.

Water-filling. / information flow.

(1). $\min \ -\sum_{i=1}^{n} \log(\alpha_i + x_i)$, $\alpha_i > 0$

s.t. $x \geq 0$

$\mathbf{1}^T x = 1$

(3).

$\circ \ L(x, \lambda, v) = -\sum_{i=1}^{n} \log(\alpha_i + x_i) - \lambda^T x + v(\mathbf{1}^T x - 1)$

KKT conditions. ① : $x^* \succeq 0$. $\mathbf{1}^T x^* = 1$

$\lambda^* \succeq 0$

② . $\lambda_i^* x_i = 0$. $i = 1, \ldots, n$

③ $\frac{\partial}{\partial x_i}$: $-\frac{1}{\alpha_i + x_i^*} - \lambda_i^* + v^* = 0 \quad i = 1, \ldots, n$.

(4). Solve directly!

λ^* is slack variable. in ①, ③, eliminate.

$v^* - \frac{1}{\alpha_i + x_i^*} = \lambda_i^* \geq 0$, $\lambda_i^* x_i = 0$.

$$\Leftrightarrow \begin{cases} v^* \geq \frac{1}{\alpha_i + x_i^*} & ③ \\ 0 = x_i^* \left(v^* - \frac{1}{\alpha_i + x_i}\right) & ② \\ x_i^* \geq 0 & ① \\ \mathbf{1}^T x^* = 1. & \end{cases}$$

(4). cont'd.

If $v^* < \frac{1}{\alpha_i}$. $v^* \geq \frac{1}{\alpha_i + x_i} \Rightarrow x_i^* > 0$.

$\overset{②}{\Rightarrow} v^* = \frac{1}{\alpha_i + x_i^*} \Rightarrow x_i^* = \frac{1}{v^*} - \alpha_i$

If $v^* \geq \frac{1}{\alpha_i}$, then $v^* \geq \frac{1}{\alpha_i + x_i} \Rightarrow x_i^* < 0$ is impossible.

So by ②. $x_i^* = 0$.

$$\Rightarrow x_i^* = \begin{cases} \frac{1}{v^*} - \alpha_i & \text{if } v^* < \frac{1}{\alpha_i} \\ 0 & \text{if } v^* \geq \frac{1}{\alpha_i} \end{cases}$$

i.e. $x_i^* = \max\left\{0, \frac{1}{v^*} - \alpha_i\right\}$

Example. KKT conditions for convex problems.

1. Primal / Dual feasible
2. Complementary Slackness
3. First order condition on x.

Water-filling / information flow.

(1). $\min \; -\sum_{i=1}^{n} \log(\alpha_i + x_i)$, $\alpha_i > 0$

s.t. $\quad x \geq 0$

$\quad \mathbf{1}^T x = 1$

(4). cont'd.

If $v^* < \frac{1}{\alpha_i}$. $\; v^* \geq \frac{1}{\alpha_i + x_i} \Rightarrow x_i^* > 0$.

$$\overset{②}{\Rightarrow} v^* = \frac{1}{\alpha_i + x_i^*} \Rightarrow x_i^* = \frac{1}{v^*} - \alpha_i$$

If $v^* \geq \frac{1}{\alpha_i}$, then $v^* \geq \frac{1}{\alpha_i + x_i} \Rightarrow x_i^* < 0$ is impossible.

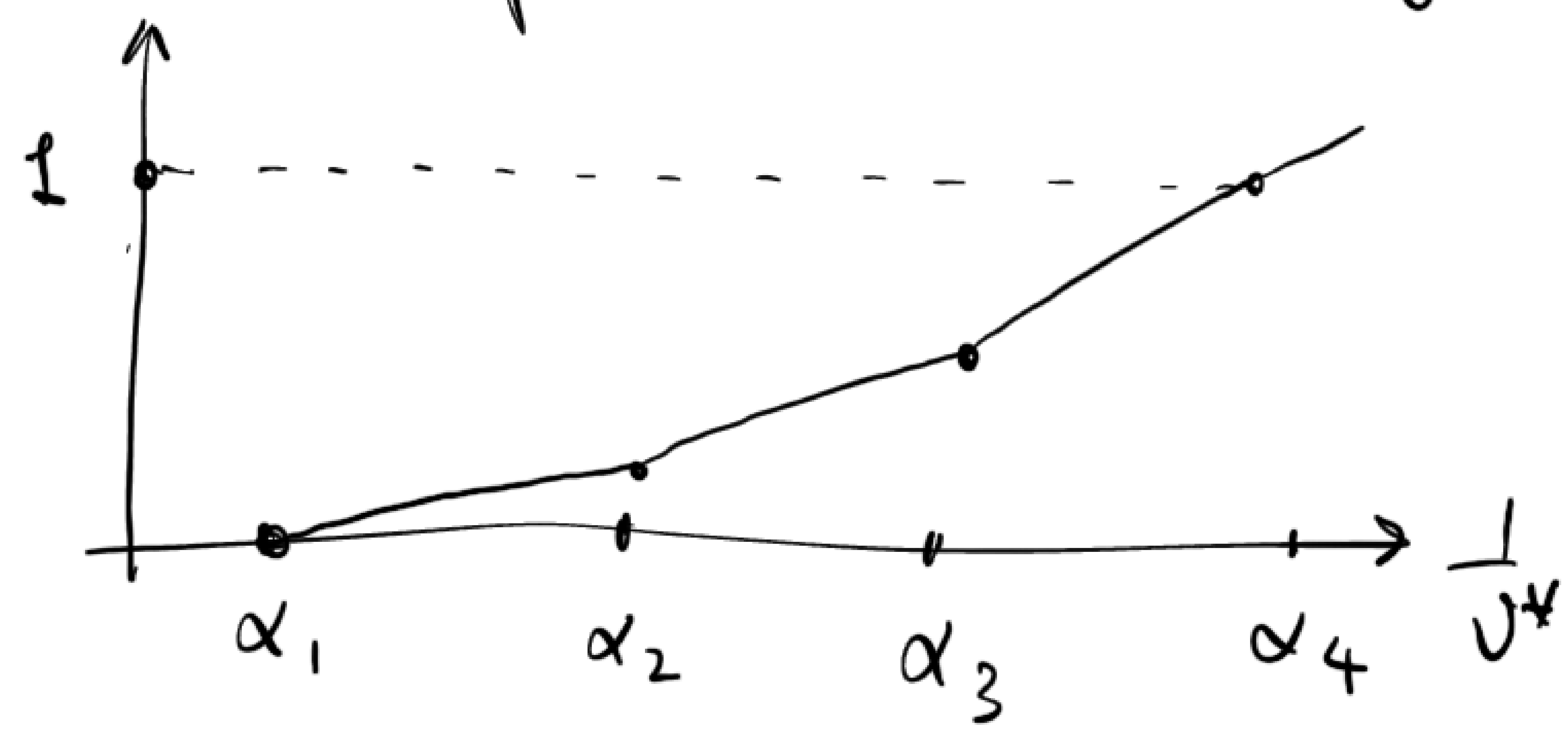
So by ②. $x_i^* = 0$.

$$\Rightarrow x_i^* = \begin{cases} \frac{1}{v^*} - \alpha_i & \text{if } v^* < \frac{1}{\alpha_i} \\ 0 & \text{if } v^* \geq \frac{1}{\alpha_i} \end{cases}$$

i.e. $x_i^* = \max\{0, \frac{1}{v^*} - \alpha_i\}$

(4) cont'd. $\mathbf{1}^T x^* = 1$. means.

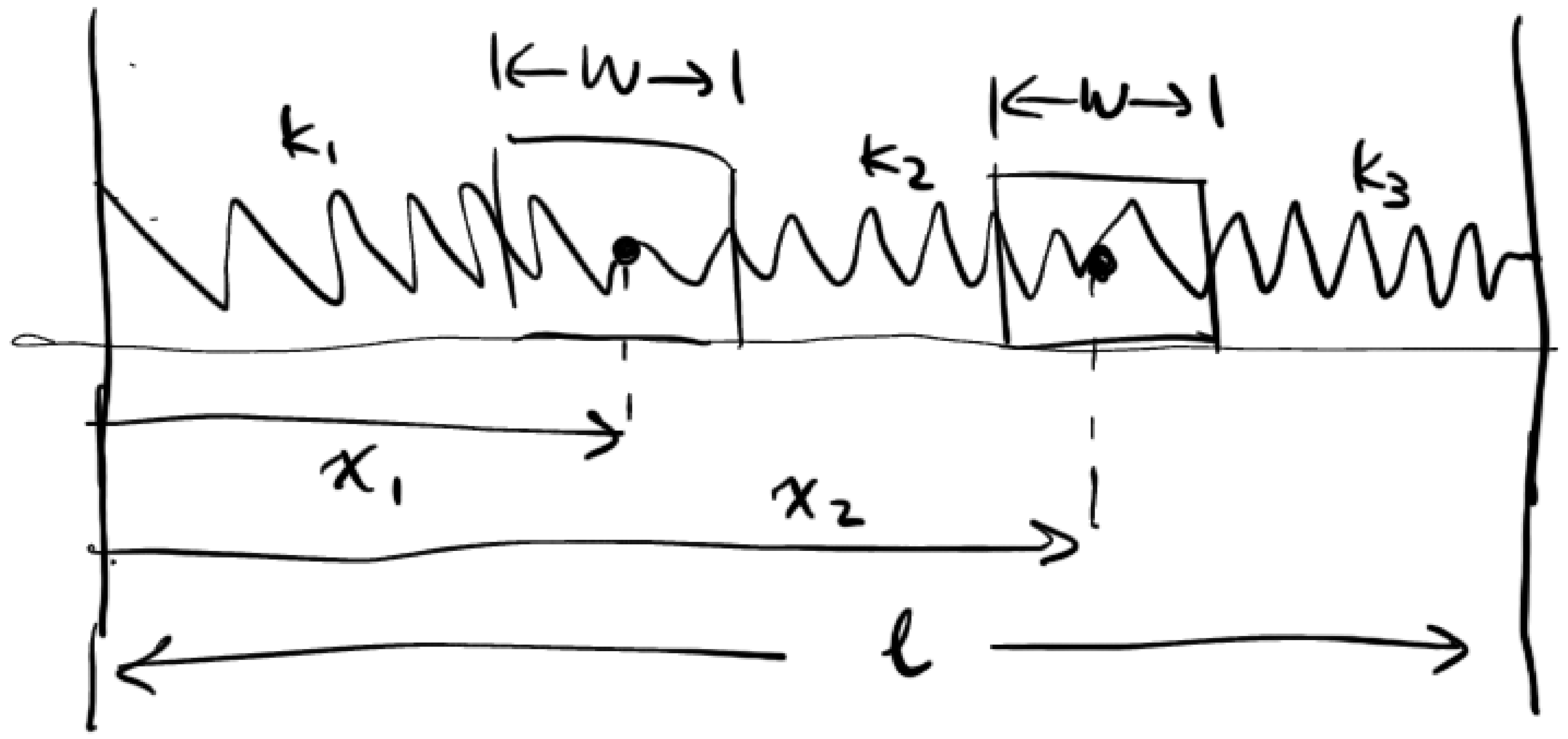
$$\sum_{i=1}^{n} x_i^* = \underbrace{\sum_{i=1}^{n} \max\{0, \frac{1}{v^*} - \alpha_i\}}_{\text{piecewise linear in } \frac{1}{v^*}} = 1.$$



Example : Mechanics Interpretation of KKT conditions — Lagrange !

Position $x \in \mathbb{R}^2$. Stiffness of springs k_i

Potential energy. $f_0(x_1, x_2) = \frac{1}{2}k_1 x_1^2 + \frac{1}{2}k_2(x_2 - x_1)^2 + \frac{1}{2}k_3(l - x_2)^2$



Equilibrium position x^* would be optimal solution to

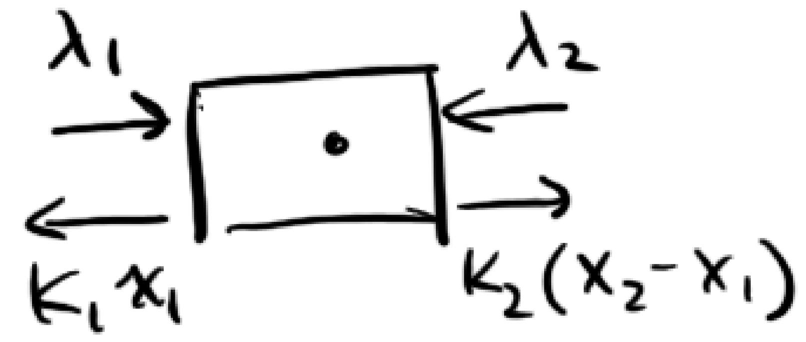
min f_0

s.t. $\frac{w}{2} - x_1 \le 0$

$w + x_1 - x_2 \le 0$

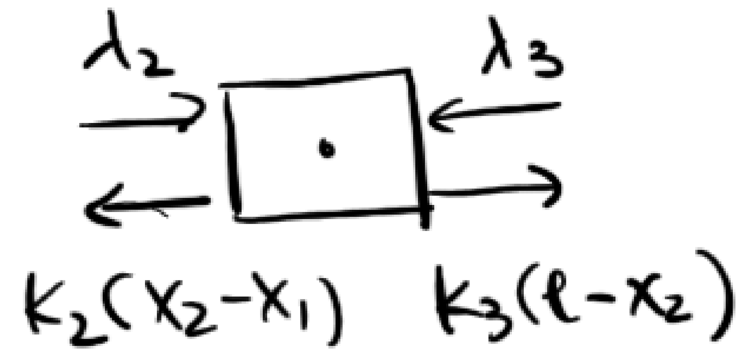
$\frac{w}{2} - l + x_2 \le 0$

} kinematic constraints



$k_1 x_1 - k_2(x_2 - x_1) - \lambda_1 + \lambda_2$

(1)



$k_2(x_2 - x_1) - k_3(l - x_2) - \lambda_2 + \lambda_3$

(2)

KKT conditions.

① . $\lambda_i \ge 0$.

② . $\lambda_1 (\frac{w}{2} - x_1) = 0$

$\lambda_2 (w + x_1 - x_2) = 0$

$\lambda_3 (\frac{w}{2} - l + x_2) = 0$

Complementary slackness : force only when contact.

$\lambda_1 \ge 0$. $\lambda_2 \le 0$.
(point away from surface)

③ . $L = \frac{1}{2}k_1 x_1^2 + \frac{1}{2}k_2(x_2 - x_1)^2 + \frac{1}{2}k_3(l - x_2)^2 + \lambda_1(\frac{w}{2} - x_1) + \lambda_2(w + x_1 - x_2) + \lambda_3(\frac{w}{2} - l + x_2)$ $\Leftrightarrow$

Force balance equation. λ_i : contact forces.

$$0 = \nabla L = \begin{bmatrix} k_1 x_1 - k_2(x_2 - x_1) - \lambda_1 + \lambda_2 \\ k_2(x_2 - x_1) - k_3(l - x_2) - \lambda_2 + \lambda_3 \end{bmatrix} \begin{matrix} (1) \\ (2) \end{matrix}$$

Examples 1. Solve the primal problem via the dual — entropy maximization

Primal:

$$\min\ f_0(x) = \sum_{i=1}^{n} x_i \log x_i$$

$$\text{s.t.}\quad Ax \preceq b$$

$$\mathbf{1}^T x = 1 \ \Big\}\ \text{probability}$$

Strong duality holds if feasible

$\Downarrow$ ①

Dual

$$\max\ g(\lambda, \upsilon) = -b^T\lambda - \upsilon - e^{-\upsilon-1}\sum_{i=1}^{n} e^{-a_i^T\lambda}$$

$$\text{s.t.}\quad \lambda \succeq 0$$

① Calculate the dual: $g(\lambda,\upsilon) = \inf_x \sum x_i \log x_i + \lambda^T(Ax-b) + \upsilon(\mathbf{1}^T x - 1)$

1.1 Conjugate of negative entropy $f(x) = x\log x$ $(x>0)$ is

$$f^*(y) = \sup_x\ xy - x\log x \quad \text{. concave.} \qquad \frac{d}{dx} = y - \log x - 1 \ \Rightarrow\ x^* = e^{y-1}$$

$$= ye^{y-1} - e^{y-1}(y-1)$$

$$= e^{y-1}.$$

So, conjugate of $f(x) = \sum_{i=1}^{n} x_i \log x_i$ is $f^*(y) = \sum_{i=1}^{n} e^{y_i - 1}$

1.2 Recall. $\min f(x)$ s.t. $x = 0$ $\Longrightarrow$

$$g(\lambda,\upsilon) = \inf_x \left(f(x) + \upsilon x\right)$$

$$= \sup_x \left(-f(x) - \upsilon x\right)$$

$$= \sup_{-x}\ \upsilon(-x) - f(x)$$

$$= -f^*(-\upsilon).$$

Similarly. $\min f(x)$ s.t. $Ax \preceq b$, $Cx = d$

$$\Rightarrow\quad g(\lambda,\upsilon) = -b^T\lambda - d^T\upsilon - f_0^*(-A^T\lambda - C^T\upsilon)$$

1.3. $g(\lambda,\upsilon) = -b^T\lambda - \upsilon - \sum_{i=1}^{n} e^{-a_i^T\lambda - \upsilon - 1}$

ith column of A.

Examples 1. Solve the primal problem via the dual — entropy maximization

<u>Primal</u>:

$$\min f_0(x) = \sum_{i=1}^{n} x_i \log x_i$$

$$\text{s.t.} \quad Ax \leq b$$
$$1^T x = 1$$

Strong duality holds if feasible

$\Downarrow$ ①.

<u>Dual</u>

$$\max g(\lambda, v) = -b^T\lambda - v - e^{-v-1} \sum_{i=1}^{n} e^{-a_i^T \lambda}$$

$$\text{s.t.} \quad \lambda \geq 0$$

(3.1). If only equality constr. $1^T x = 1$.

— Max Entropy distr. with no constr.

e.g. $A = 0$. $b \geq 0$.

$\Rightarrow \lambda^* = 0$. $\Rightarrow g(\lambda^*, v) = -v - e^{-v-1} n$

$\Rightarrow 0 = \frac{\partial}{\partial v} = -1 + ne^{-1-v} \Rightarrow e^{-1-v^*} = \frac{1}{n}$.

$\Rightarrow \underline{x_i^* = \frac{1}{n}}$. Uniform distribution.

② If (λ^*, v^*) is optimal solution to dual.

then Lagrangian at (λ^*, v^*) is

$$L(x, \lambda^*, v^*) = \sum_{i=1}^{n} x_i \log x_i + \lambda^{*T}(Ax - b) + v^*(1^T x - 1)$$

Since x^* minimize $L(x, \lambda^*, v^*)$.

$$\frac{\partial}{\partial x_i} L = \log x_i + 1 + \frac{\partial}{\partial x_i} \sum_{k,j} \lambda_k^* a_{kj} x_j + v^*$$

$$= \log x_i + 1 + \sum_k \lambda_k^* a_{ki} + v^*.$$

Set $\nabla_x L = 0$. $\Rightarrow \underline{x_i^* = e^{-(1 + v^* + a_i^T \lambda^*)}}$

(observation)

(3.2). e.g. If x is a discrete prob. distr. random variable with distr x

and $a_i^T x \leq b_i$ is $\mathbb{E}_x(A_i) \leq b_i$.

upper bound on expectation

then what's the Max-Entropy distribution x^*?

$$\Rightarrow x_i^* = \frac{1}{Z} e^{-a_i^T \lambda^*} \quad \text{for some } \lambda^*.$$

normalization constant

Exponentially distributed! Gibbs distribution.

Examples 2 Solve the primal problem via the dual — entropy maximization

- Separable function with an equality.

$$\min \quad f_0(x) = \sum_{i=1}^{n} f_i(x_i) \qquad f_i : \mathbb{R} \to \mathbb{R}. \text{ strictly convex. differentiable.}$$

$$\text{s.t.} \quad a^T x = b \qquad a \in \mathbb{R}^n. \ b \in \mathbb{R}.$$

- f_0 is separable
- If dom f_0 intersects constraint. $\Rightarrow$ unique minima.

Lagrangian

$$L(x, \nu) = \sum_{i=1}^{n} f_i(x) + \nu(a^T x - b)$$

$$= -b\nu + \sum_{i=1}^{n} \left(f_i(x_i) + \nu a_i x_i \right)$$

So. L is also separable (in x_i)

$$g(\nu) = -b\nu + \inf_x \sum_{i=1}^{n} \left(f_i(x_i) + \nu a_i x_i \right)$$

$$= -b\nu + \sum_{i=1}^{n} \inf_{x_i} \left(f_i(x_i) + \nu a_i x_i \right)$$

$$= -b\nu + \sum_{i=1}^{n} - \sup_{x_i} \left((-\nu a_i) x_i - f_i(x_i) \right)$$

$$= -b\nu + \sum_{i=1}^{n} -f^*(-\nu a_i)$$

$\Leftarrow$

Dual: $\max\limits_{\nu} \ -b\nu - \sum_{i=1}^{n} f_i^*(-\nu a_i)$

scalar variable!

— Easy to solve. e.g. bisection

Then after obtaining ν^*. recover x^* by

$\nabla_x L(x^*, \nu^*) = 0$. i.e. $\nu a_i + f_i'(x^*) = 0$

- Saddle-point interpretation of Lagrange duality.
  - Consider no equality constraints. for simplicity.

$$\sup_{\lambda \succeq 0} L(x,\lambda) = \sup_{\lambda \succeq 0}\left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) \right) = \begin{cases} f_0(x) & f_i(x) \le 0 \quad i=1,\dots,m \\ \infty & \text{otherwise} \end{cases}$$

So. primal optimality is $p^* = \inf_x \sup_{\lambda \succeq 0} L(x,\lambda)$

Recall. def. of dual optimality is $d^* = \sup_{\lambda \succeq 0} \inf_x L(x,\lambda)$.

- So. weak duality : $d^* = \sup_{\lambda \succeq 0} \inf_x L(x,\lambda) \le \inf_x \sup_{\lambda \succeq 0} L(x,\lambda) = p^*$

Alway holds! (Max-min ineq.)

strong duality : $d^* = \sup_{\lambda \succeq 0} \inf_x L(x,\lambda) = \inf_x \sup_{\lambda \succeq 0} L(x,\lambda) = p^*$

So order can be switched

- Saddle-point interpretation of Lagrange duality.

(1) Max-min inequality in <u>zero-sum game</u>.

Player 1 chooses x, Player 2 chooses λ.

then player 1 pays $L(x,\lambda)$ to player 2.

So. player 1's goal: $\min L$

player 2's goal: $\max L$

(2) If player 1 goes first. chooses x.

then player 2's choice is

$\max_{\lambda} L(x,\lambda)$. depends on x.

(3). If player 1 knows / assumes player 2 takes this strategy. ~~then~~

$x^* = \arg\min_{x} \max_{\lambda} L(x,\lambda)$. is P1's choice.

(4). If order is reversed.

player 2 chooses first.

$\lambda^* = \arg\min_{\lambda} \max_{x} L(x,\lambda)$

(5). Max-min inequality: Better to go second.

i.e. know opponent's choice. before choosing.

(6). When strong duality holds.

going second doesn't have any advantage.

- x^*, λ^* s.t. $\inf_{\lambda} \sup_{x} L(x,\lambda) = \sup_{x} \inf_{\lambda} L(x,\lambda)$ is called <u>saddle point</u> of L.

- Price interpretation of Lagrange duality.

(1). x — a company's operation variables. e.g. warehouse size.

$f_0(x)$ — cost of operating at x

$f_i(x) \le 0$ — constraint. e.g. limit on resources / regulatory limit

$\min f_0(x)$ — optimal operating strategy.
s.t. $f_i(x) \le 0$.

(2) — If limits can be violated. Each unit of violation incurs $\lambda_i \ge 0$ cost $(f_i > 0)$
and each unit of slackness yields λ_i profit. $(f_i < 0)$.

$\Rightarrow$ violate/slack in i^{th} constr : $\lambda_i f_i(x)$ e.g. rent extra space / rent out unused space.

$\Rightarrow$ Total cost. $L(x,\lambda) = f_0(x) + \sum_{i=1}^m \lambda_i f_i(x)$.

$g(x) = \min_x L(x,\lambda)$ min cost operation

$d^* = \max g(\lambda)$ s.t. $\lambda \ge 0$. = the min cost under worst prices (λ_i) for constraint violation.
("shadow prices").

Duality : – Weak : Scenario (2) always lower cost than (1).
– Gap : smallest gain in cost when constr. are violatable.
– Strong : no gain to violate constr.