ELG6108: Introduction to Convex Optimization

Lecture 5: Duality

Dr. Sergey Loyka

EECS, University of Ottawa

March 16, 2021

Duality[1]

- Lagrangian & dual function
- dual problem
- weak and strong duality
- geometric interpretation
- optimality (KKT) conditions
- perturbation and sensitivity analysis
- examples

[1] adapted from Boyd & Vandenberghe, Convex Optimization, Lecture slides.

Lagrangian

Standard form problem (not necessarily convex)

$$\begin{aligned} \min \quad & f_0(\mathbf{x}) \\ \text{s.t.} \quad & f_i(\mathbf{x}) \leq 0, \quad i = 1, \ldots, m \\ & h_i(\mathbf{x}) = 0, \quad i = 1, \ldots, p \end{aligned} \tag{1}$$

domain $\mathcal{D}$, optimal value $p^\star$

Lagrangian:

$$L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}) + \sum_{i=1}^{p} \nu_i h_i(\mathbf{x}) \tag{2}$$

- weighted sum of objective and constraint functions
- λ_i is Lagrange multiplier resp. for $f_i(\mathbf{x}) \leq 0$
- ν_i is Lagrange multiplier resp. for $h_i(\mathbf{x}) = 0$

Lagrange dual function

- **Lagrange dual function**:

$$g(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \min_{\mathbf{x} \in \mathcal{D}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = \min_{\mathbf{x} \in \mathcal{D}} \left(f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}) + \sum_{i=1}^{p} \nu_i h_i(\mathbf{x}) \right) \tag{3}$$

- unconstraint minimization in $\min_{\mathbf{x} \in \mathcal{D}}$
- $g(\boldsymbol{\lambda}, \boldsymbol{\nu})$ is (jointly) concave (can be $-\infty$ for some $\boldsymbol{\lambda}, \boldsymbol{\nu}$);
- Q: why?
- fundamental for optimality conditions
- also used by many algorithms

Lagrange dual function & fundamental LB

Fundamental lower bound (LB):

$$\text{if } \boldsymbol{\lambda} \succeq 0 \text{ then } g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \le p^\star \tag{4}$$

proof:

1. if $\mathbf{x}$ is feasible and $\boldsymbol{\lambda} \succeq 0$, then

$$f_0(\mathbf{x}) \ge L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \ge \min_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \tag{5}$$

Q: explain (5)

2. minimizing over all feasible $\mathbf{x}$ gives $p^\star \ge g(\boldsymbol{\lambda}, \boldsymbol{\nu})$

LB holds even if not convex

Example: least-norm solution of linear equations

$$\min \quad \mathbf{x}^T\mathbf{x} \quad \text{s.t.} \quad \mathbf{A}\mathbf{x} = \mathbf{b} \tag{6}$$

dual function

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

$$\nabla_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\nu}) = 2\mathbf{x} + \mathbf{A}^T\boldsymbol{\nu} = 0 \quad \Longrightarrow \quad \mathbf{x} = -(1/2)\mathbf{A}^T\boldsymbol{\nu} \tag{7}$$

- plug in in L to obtain g :

$$g(\boldsymbol{\nu}) = L\left((-1/2)\mathbf{A}^T\boldsymbol{\nu}, \boldsymbol{\nu}\right) = -\frac{1}{4}\boldsymbol{\nu}^T\mathbf{A}\mathbf{A}^T\boldsymbol{\nu} - \mathbf{b}^T\boldsymbol{\nu} \tag{8}$$

- $g(\boldsymbol{\nu})$ is concave in $\boldsymbol{\nu}$

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

Example: standard form LP

$$\min\ \mathbf{c}^T\mathbf{x} \quad \text{s.t. } \mathbf{A}\mathbf{x} = \mathbf{b},\ \mathbf{x} \succeq 0 \tag{9}$$

dual function

- the Lagrangian is

$$\begin{aligned} L(\mathbf{x},\boldsymbol{\lambda},\boldsymbol{\nu}) &= \mathbf{c}^T\mathbf{x} + \boldsymbol{\nu}^T(\mathbf{A}\mathbf{x}-\mathbf{b}) - \boldsymbol{\lambda}^T\mathbf{x} \\ &= -\mathbf{b}^T\boldsymbol{\nu} + \left(\mathbf{c} + \mathbf{A}^T\boldsymbol{\nu} - \boldsymbol{\lambda}\right)^T\mathbf{x} \end{aligned} \tag{10}$$

Example: standard form LP

- $L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$ is affine in $\mathbf{x}$, hence

$$g(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \min_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = \begin{cases} -\mathbf{b}^T\boldsymbol{\nu} & \mathbf{A}^T\boldsymbol{\nu} - \boldsymbol{\lambda} + \mathbf{c} = 0 \\ -\infty & \text{otherwise} \end{cases} \quad (11)$$

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

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

Two-way partitioning

$$\min \;\; \mathbf{x}^T \mathbf{W} \mathbf{x} \;\; \text{s.t. } x_i^2 = 1, \;\; i = 1..n \tag{12}$$

- convex ?

Two-way partitioning

$$\min \ \mathbf{x}^T\mathbf{W}\mathbf{x} \ \text{ s.t. } x_i^2 = 1, \ i = 1..n \tag{12}$$

- convex ?
- feasible set contains 2^n 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

Two-way partitioning: dual function, LB

- **dual function**

$$g(\boldsymbol{\nu}) = \min_{\mathbf{x}} \left(\mathbf{x}^T \mathbf{W} \mathbf{x} + \sum_i \nu_i \left(x_i^2 - 1 \right) \right)$$
$$= \min_{\mathbf{x}} \mathbf{x}^T (\mathbf{W} + \operatorname{diag}(\boldsymbol{\nu}))\mathbf{x} - \mathbf{1}^T \boldsymbol{\nu}$$
$$= \begin{cases} -\mathbf{1}^T \boldsymbol{\nu} & \mathbf{W} + \operatorname{diag}(\boldsymbol{\nu}) \succeq 0 \\ -\infty & \text{otherwise} \end{cases} \tag{13}$$

- Q: prove (13)

Two-way partitioning: dual function, LB

- **dual function**

$$
\begin{aligned}
g(\boldsymbol{\nu}) &= \min_{\mathbf{x}} \left(\mathbf{x}^T \mathbf{W} \mathbf{x} + \sum_i \nu_i \left(x_i^2 - 1 \right) \right) \\
&= \min_{\mathbf{x}} \mathbf{x}^T (\mathbf{W} + \mathrm{diag}(\boldsymbol{\nu})) \mathbf{x} - \mathbf{1}^T \boldsymbol{\nu} \\
&= \begin{cases} -\mathbf{1}^T \boldsymbol{\nu} & \mathbf{W} + \mathrm{diag}(\boldsymbol{\nu}) \succeq 0 \\ -\infty & \text{otherwise} \end{cases}
\end{aligned} \tag{13}
$$

- Q: prove (13)
- **lower bound:** $p^\star \geq -\mathbf{1}^T \boldsymbol{\nu}$ if $\mathbf{W} + \mathrm{diag}(\boldsymbol{\nu}) \succeq 0$
- example:

$$
\boldsymbol{\nu} = -\lambda_{\min}(\mathbf{W}) \mathbf{1} \to p^\star \geq n \lambda_{\min}(\mathbf{W}) \tag{14}
$$

The dual problem: best LB

- **Lagrange dual problem**

$$\max_{\boldsymbol{\lambda},\boldsymbol{\nu}} \; g(\boldsymbol{\lambda},\boldsymbol{\nu}) \;\; \text{s.t. } \boldsymbol{\lambda} \succeq 0 \tag{15}$$

- best LB on $p^\star$ via Lagrange dual function

The dual problem: best LB

- **Lagrange dual problem**

$$\max_{\boldsymbol{\lambda},\boldsymbol{\nu}} \; g(\boldsymbol{\lambda},\boldsymbol{\nu}) \;\; \text{s.t. } \boldsymbol{\lambda} \succeq 0 \tag{15}$$

- best LB on $p^\star$ via Lagrange dual function
- convex problem?

The dual problem: best LB

- **Lagrange dual problem**

$$\max_{\boldsymbol{\lambda},\boldsymbol{\nu}} \; g(\boldsymbol{\lambda},\boldsymbol{\nu}) \;\; \text{s.t.} \;\; \boldsymbol{\lambda} \succeq 0 \tag{15}$$

- best LB on $p^\star$ via Lagrange dual function
- convex problem?
- yes, optimal value $= d^\star$:

$$d^* = \max_{\boldsymbol{\lambda},\boldsymbol{\nu}} \; g(\boldsymbol{\lambda},\boldsymbol{\nu}) \;\; \text{s.t.} \;\; \boldsymbol{\lambda} \succeq 0$$

- $\boldsymbol{\lambda},\boldsymbol{\nu}$ are dual feasible if $\boldsymbol{\lambda} \succeq 0, (\boldsymbol{\lambda},\boldsymbol{\nu}) \in \operatorname{dom} g$
- often simplified by making implicit constraint $(\boldsymbol{\lambda},\boldsymbol{\nu}) \in \operatorname{dom} g$ explicit

Example: standard form LP and its dual

- standard LP

$$\min_{\mathbf{x}} \ \mathbf{c}^T\mathbf{x} \ \text{ s.t. } \mathbf{A}\mathbf{x} = \mathbf{b}, \ \mathbf{x} \succeq 0 \tag{16}$$

Example: standard form LP and its dual

- standard LP

$$\min_{\mathbf{x}} \ \mathbf{c}^T\mathbf{x} \ \text{ s.t. } \mathbf{A}\mathbf{x} = \mathbf{b}, \ \mathbf{x} \succeq 0 \tag{16}$$

- and its dual

$$\max_{\boldsymbol{\nu}} \ -\mathbf{b}^T\boldsymbol{\nu} \ \text{ s.t. } \ \mathbf{A}^T\boldsymbol{\nu} + \mathbf{c} \succeq 0 \tag{17}$$

Weak and strong duality

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

- *always holds* (for convex and nonconvex problems)
- nontrivial lower bounds for difficult problems
- example: a lower bound for the two-way partitioning problem

$$\min_{\mathbf{x}} \ \mathbf{x}^T \mathbf{W} \mathbf{x} \ \text{ s.t. } x_i^2 = 1 \tag{18}$$

via the SDP

$$\max_{\boldsymbol{\nu}} \ -\mathbf{1}^T \boldsymbol{\nu} \ \text{ s.t. } \ \mathbf{W} + \text{diag}(\boldsymbol{\nu}) \succeq 0 \tag{19}$$

Weak and strong duality

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

$$\begin{array}{ll} \min & f_0(\mathbf{x}) \\ \text{s.t.} & f_i(\mathbf{x}) \leq 0, \quad i = 1, \dots, m \\ & \mathbf{A}\mathbf{x} = \mathbf{b} \end{array} \tag{20}$$

if it is strictly feasible, i.e.

$$\exists x \in \operatorname{int} \mathcal{D} : \quad f_i(\mathbf{x}) < 0, \quad i = 1, \dots, m, \quad \mathbf{A}\mathbf{x} = \mathbf{b} \tag{21}$$

- also guarantees that the dual optimum is attained (if $p^\star > -\infty$)
- can be sharpened: e.g., can replace int $\mathcal{D}$ with 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

Complementary slackness

Assume that strong duality holds and let $\mathbf{x}^*$ be primal optimal, $(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$ be dual optimal. Then,

$$f_0(\mathbf{x}^*) = p^* = d^* = g(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*) \tag{22}$$

$$= \min_{\mathbf{x}} \left(f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i^* f_i(x) + \sum_{i=1}^{p} \nu_i^* h_i(x) \right) \tag{23}$$

$$\leq f_0(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i^* f_i(\mathbf{x}^*) + \sum_{i=1}^{p} \nu_i^* h_i(\mathbf{x}^*) \tag{24}$$

$$\leq f_0(\mathbf{x}^*) \tag{25}$$

hence, the two inequalities hold with equality

- $\mathbf{x}^*$ minimizes $L(\mathbf{x}, \boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$
- $\lambda_i^* f_i(\mathbf{x}^*) = 0$ for all i, known as complementary slackness:

$$\lambda_i^* > 0 \Longrightarrow f_i(\mathbf{x}^*) = 0, \quad f_i(\mathbf{x}^*) < 0 \Longrightarrow \lambda_i^* = 0 \tag{26}$$

Karush-Kuhn-Tucker (KKT) conditions

The most fundamental optimality conditions:

1. **stationarity**: $\nabla_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = 0$, or

$$\nabla_{\mathbf{x}} f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i \nabla_{\mathbf{x}} f_i(\mathbf{x}) + \sum_{i=1}^{p} \nu_i \nabla_{\mathbf{x}} h_i(\mathbf{x}) = 0 \qquad (27)$$

2. **complementary slackness**: $\lambda_i f_i(\mathbf{x}) = 0$
3. **primal feasibility**: $f_i(\mathbf{x}) \leq 0, h_i(\mathbf{x}) = 0$, for all i
4. **dual feasibility**: $\lambda_i \geq 0$ (no condition on ν_i)

if strong duality holds and $\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}$ are optimal, then they must satisfy the KKT conditions, i.e. KKT conditions are necessary for optimality

KKT conditions for convex problem

If $\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\nu}^*$ satisfy KKT conditions for a convex problem, then they are optimal, i.e. **any solution of KKT is optimal** (sufficiency).

Proof:

- from complementary slackness:

$$L(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\nu}^*) = f_0(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i^* f_i(\mathbf{x}^*) + \sum_{i=1}^{p} \nu_i^* h_i(\mathbf{x}^*) = f_0(\mathbf{x}^*) \quad (28)$$

- from stationarity and convexity:

$$g(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*) = \min_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}^*, \boldsymbol{\nu}^*) = L(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\nu}^*) = f_0(\mathbf{x}^*) \quad (29)$$

so that $f_0(\mathbf{x}^*) = p^*$, since $g(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*) = f_0(\mathbf{x}^*)$ is a certificate of optimality (via the LB). Q.E.D.

KKT conditions for convex problem

If Slater's condition is satisfied:

$\mathbf{x}$ is optimal if and only if there exist $\boldsymbol{\lambda}, \boldsymbol{\nu}$ that satisfy KKT conditions, i.e.
KKT are sufficient and necessary for optimality

- recall that Slater implies strong duality, and dual optimum is attained
- generalizes (300y old) optimality condition $\nabla f_0(\mathbf{x}) = 0$ for unconstrained problem

Example: optimal power allocation (OPA) I

Maximizing the sum rate of parallel Gaussian channels via OPA (WiFi, cellular, DSL),

$$(\text{P1}) \quad \max_{x_i} \sum_{i=1}^{n} \log(1 + x_i/\alpha_i) \ \text{ s.t. } \ x_i \geq 0, \quad \sum_i x_i = P \tag{30}$$

x_i = signal power of i-th channel, α_i = its noise power, P = total signal (Tx) power; $x_i, \alpha_i \geq 0$; $\log(1 + x_i/\alpha_i)$ = rate of i-th channel, in [b/s/Hz].

Equivalent to

$$(\text{P2}) \quad \min_{x_i} \ -\sum_{i=1}^{n} \log(x_i + \alpha_i) \ \text{ s.t. } \ -x_i \leq 0, \quad \sum_i x_i = P \tag{31}$$

Example: optimal power allocation (OPA) II

Its Lagrangian is

$$L = -\sum_{i=1}^{n} \log(x_i + \alpha_i) - \sum_i \lambda_i x_i + \nu \left(\sum_i x_i - P \right) \tag{32}$$

and the KKT conditions are

$$(a)\ \frac{1}{x_i + \alpha_i} + \lambda_i = \nu,\ (b)\ \lambda_i x_i = 0,\ (c)\ \sum_i x_i = P,\ (d)\ \lambda_i \geq 0 \tag{33}$$

from (b) and (a):

- if $x_i > 0 \rightarrow \lambda_i = 0$ and $x_i = 1/\nu - \alpha_i > 0 \rightarrow \nu > 1/\alpha_i$ (active ch.)
- if $\nu \geq 1/\alpha_i \rightarrow x_i = 0,\ \lambda_i = \nu - 1/\alpha_i$ (inactive ch.)
  so that

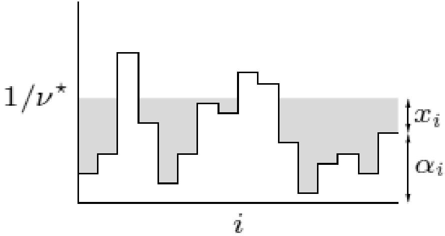
$$x_i = (1/\nu - \alpha_i)_+, \text{ where } (x)_+ = \max\{0, x\} \tag{34}$$

- find ν from (c): $\sum_i (1/\nu - \alpha_i)_+ = P$

OPA = Water Filling (WF)

water-filling interpretation

- container with n segments; floor profile: segment i is at height α_i
- flood area with P units of "water"
- "water" level is $x_i = (1/\nu - \alpha_i)_+$ at segment i



- one of the most elegant/popular algorithms in IT, communications, signal processing, control
- widely used in practice (quantized, WiFi, 3/4/5G, DSL)

Water Filling (WF)

- Q1: find the conditions under which only 1 channel is active: $x_1^* > 0,\ x_2^* ... x_n^* = 0$
- Q2: find the conditions under which all channels are active: $x_1^* ... x_n^* > 0$
- Q3: show that the number of active streams is an increasing function of P
- Q4: find a closed-form expression for ν^* and, using it, the number of active streams
- Q5: consider a modification of (P1), where the power constraint is via an equality:

$$(\text{P3}) \ \max_{x_i} \sum_{i=1}^{n} \log(1 + x_i/\alpha_i) \ \text{ s.t. } \ x_i \geq 0, \quad \sum_i x_i \leq P \tag{35}$$

show that, at optimal point, it always holds with equality: $\sum_i x_i^* = P$, so that (P1) and (P3) are also equivalent.

Water Filling (WF) with per-channel constraint

- Q6: consider the following modification of (P3),

$$\text{(P4)} \quad \max_{x_i} \sum_{i=1}^{n} \log(1 + x_i/\alpha_i) \ \text{ s.t. } \ x_i \geq 0, \ \sum_i x_i \leq P, \ x_i \leq P_1$$

where P_1 is the maximum *per-channel* power. Find its OPA and compare it to the WF in (34). Give its geometric interpretations (similar to WF).

Perturbation and sensitivity analysis

(unperturbed) optimization problem and its dual

$$\begin{array}{lll} \min & f_0(\mathbf{x}) \\ \text{s.t.} & f_i(\mathbf{x}) \le 0, i = 1, \dots, m \\ & h_i(\mathbf{x}) = 0, \quad i = 1, \dots, p \end{array} \qquad \begin{array}{ll} \text{maximize} & g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \\ \text{s.t.} & \boldsymbol{\lambda} \succeq 0 \end{array} \tag{36}$$

perturbed problem and its dual

$$\begin{array}{lll} \min & f_0(\mathbf{x}) \\ \text{s.t.} & f_i(\mathbf{x}) \le u_i, i = 1, \dots, m \\ & h_i(\mathbf{x}) = v_i, \quad i = 1, \dots, p \end{array} \qquad \begin{array}{ll} \max & g(\boldsymbol{\lambda}, \boldsymbol{\nu}) - u^T \boldsymbol{\lambda} - v^T \boldsymbol{\nu} \\ \text{s.t.} & \boldsymbol{\lambda} \succeq 0 \end{array} \tag{37}$$

- $\mathbf{x}$ is primal variable; u, v are parameters
- $p^\star(u, v)$ is optimal value as a function of u, v
- we are interested in information about $p^\star(u, v)$ that we can obtain from the solution of the unperturbed problem and its dual

global sensitivity result assume strong duality holds for unperturbed problem, and that $\boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star$ are dual optimal for unperturbed problem apply weak duality to perturbed problem:

$$p^\star(\mathbf{u}, \mathbf{v}) \geq g\left(\boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star\right) - \mathbf{u}^T \boldsymbol{\lambda}^\star - \mathbf{v}^T \boldsymbol{\nu}^\star \tag{38}$$

$$= p^\star(0, 0) - \mathbf{u}^T \lambda^\star - \mathbf{v}^T \boldsymbol{\nu}^\star \tag{39}$$

sensitivity interpretation

- if $\lambda_i^\star$ large: $p^\star$ increases greatly if we tighten constraint $i\,(u_i < 0)$
- 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: if (in addition) $p^\star(u,v)$ is differentiable at $(0,0)$, then

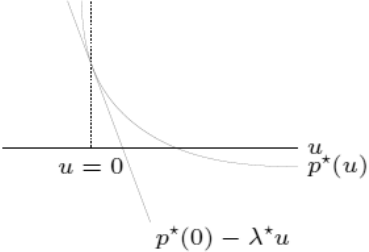
$$\lambda_i^\star = -\frac{\partial p^\star(0,0)}{\partial u_i}, \quad \nu_i^\star = -\frac{\partial p^\star(0,0)}{\partial v_i} \tag{40}$$

proof (for $\lambda_i^\star$) : from global sensitivity result,

$$\begin{aligned} \tfrac{\partial p^\star(0,0)}{\partial u_i} &= \lim_{t \searrow 0} \tfrac{p^\star(te_i,0) - p^\star(0,0)}{t} \geq -\lambda_i^\star \\ \tfrac{\partial p^\star(0,0)}{\partial u_i} &= \lim_{t \nearrow 0} \tfrac{p^\star(te_i,0) - p^\star(0,0)}{t} \leq -\lambda_i^\star \end{aligned} \tag{41}$$

hence, equality

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



Duality and problem reformulations

- equivalent formulations of a problem can lead to very different duals
- reformulating the primal problem can be useful when the 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 $f_0(\mathbf{x})$ by $\phi(f_0(\mathbf{x}))$ with ϕ convex, increasing

Introducing new variables and equality constraints

$$\min \quad f_0(\mathbf{A}\mathbf{x} + \mathbf{b})$$

- dual function is constant: $g = \inf_{\mathbf{x}} L(\mathbf{x}) = \inf_{\mathbf{x}} f_0(\mathbf{A}\mathbf{x} + \mathbf{b}) = p^\star$
- we have strong duality, but dual is quite useless

reformulated problem and its dual

$$\begin{array}{llll} \min & f_0(\mathbf{y}) & \max & \mathbf{b}^T\boldsymbol{\nu} - f_0^*(\boldsymbol{\nu}) \\ \text{s.t.} & \mathbf{A}\mathbf{x} + \mathbf{b} - \mathbf{y} = 0 & \text{s.t.} & \mathbf{A}^T\boldsymbol{\nu} = 0 \end{array} \tag{42}$$

Introducing new variables and equality constraints

dual function follows from

$$g(\boldsymbol{\nu}) = \inf_{\mathbf{x},\mathbf{y}} \left(f_0(\mathbf{y}) - \boldsymbol{\nu}^T \mathbf{y} + \boldsymbol{\nu}^T \mathbf{A}\mathbf{x} + \mathbf{b}^T \boldsymbol{\nu} \right) \tag{43}$$

$$= \begin{cases} -f_0^*(\boldsymbol{\nu}) + \mathbf{b}^T \boldsymbol{\nu} & \mathbf{A}^T \boldsymbol{\nu} = 0 \\ -\infty & \text{otherwise} \end{cases} \tag{44}$$

Implicit constraints

LP with box constraints: primal and dual problem

$$\begin{array}{ll} \min & \mathbf{c}^T\mathbf{x} \\ \text{s.t.} & \mathbf{A}\mathbf{x} = \mathbf{b} \\ & -\mathbf{1} \preceq \mathbf{x} \preceq \mathbf{1} \end{array} \qquad \begin{array}{ll} \max & -\mathbf{b}^T\boldsymbol{\nu} - \mathbf{1}^T\boldsymbol{\lambda}_1 - \mathbf{1}^T\boldsymbol{\lambda}_2 \\ \text{s.t.} & \mathbf{c} + \mathbf{A}^T\boldsymbol{\nu} + \boldsymbol{\lambda}_1 - \boldsymbol{\lambda}_2 = 0 \\ & \boldsymbol{\lambda}_1 \succeq 0, \quad \boldsymbol{\lambda}_2 \succeq 0 \end{array} \tag{45}$$

reformulation with box constraints made implicit

$$\begin{array}{ll} \min & f_0(\mathbf{x}) = \begin{cases} \mathbf{c}^T\mathbf{x} & -\mathbf{1} \preceq \mathbf{x} \preceq \mathbf{1} \\ \infty & \text{otherwise} \end{cases} \\ \text{s.t.} & \mathbf{A}\mathbf{x} = \mathbf{b} \end{array} \tag{46}$$

Implicit constraints

dual function

$$\begin{aligned} g(\boldsymbol{\nu}) &= \inf_{-\mathbf{1}\preceq \mathbf{x} \preceq \mathbf{1}} \left(\mathbf{c}^T\mathbf{x} + \boldsymbol{\nu}^T(\mathbf{A}\mathbf{x} - \mathbf{b})\right) \\ &= -\mathbf{b}^T\boldsymbol{\nu} - \left|\mathbf{A}^T\boldsymbol{\nu} + \mathbf{c}\right|_1 \end{aligned} \quad (47)$$

dual problem: maximize $-\mathbf{b}^T\boldsymbol{\nu} - \left|\mathbf{A}^T\boldsymbol{\nu} + \mathbf{c}\right|_1$