ORF522 – Linear and Nonlinear Optimization

9. Sensitivity analysis for linear optimization

Bartolomeo Stellato — Fall 2020

Ed forum

- The strong duality statement doesn't state that when the primal is not feasible, the dual is unbounded (p=+∞, d=+∞). Why it's not possible to have p=+∞ and d is some finite value?

- Does the dual simplex index j refer to the basis matrix index? Or the full A?

$B^T d = e_j$

Recap

General forms

Standard form LP

Primal

minimize $c^T x$

subject to $Ax = b$

$x \geq 0$

Dual

maximize $-b^T y$

subject to $A^T y + c \geq 0$

Inequality form LP

Primal

minimize $c^T x$

subject to $Ax \leq b$

Dual

maximize $-b^T y$

subject to $A^T y + c = 0$

$y \geq 0$

Today's lecture
[Chapter 5, Bertsimas and Tsitsiklis]

Sensitivity analysis in linear optimization

- Adding new constraints and variables
- Change problem data
- Differentiable optimization

Adding new constraints and variables

Adding new variables

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to} \quad & Ax = b \\ & x \geq 0 \end{aligned}$$

Solution $x^\star, y^\star$

Adding new variables

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & x \geq 0 \end{array} \longrightarrow \begin{array}{ll} \text{minimize} & c^T x + c_{n+1} x_{n+1} \\ \text{subject to} & Ax + A_{n+1} x_{n+1} = b \\ & x, x_{n+1} \geq 0 \end{array}$$

Solution $x^\star, y^\star$

Adding new variables

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & x \geq 0 \end{array} \longrightarrow \begin{array}{ll} \text{minimize} & c^T x + c_{n+1} x_{n+1} \\ \text{subject to} & Ax + A_{n+1} x_{n+1} = b \\ & x, x_{n+1} \geq 0 \end{array}$$

Solution $x^\star, y^\star$

Solution $(x^\star, 0), y^\star$ **optimal** for the new problem?

Adding new variables

Optimality conditions

minimize	$c^T x + c_{n+1} x_{n+1}$
subject to	$Ax + A_{n+1} x_{n+1} = b$
	$x, x_{n+1} \geq 0$

$\longrightarrow$ Solution $(x^\star, 0)$ is still **primal feasible**

Adding new variables

Optimality conditions

minimize	$c^T x + c_{n+1} x_{n+1}$
subject to	$Ax + A_{n+1} x_{n+1} = b$
	$x, x_{n+1} \geq 0$

$\longrightarrow$ Solution $(x^\star, 0)$ is still **primal feasible**

$A^T y + c \geq 0$

Is $y^\star$ still **dual feasible**?

$$A_{n+1}^T y^\star + c_{n+1} \geq 0$$

Adding new variables

Optimality conditions

$$\begin{array}{ll} \text{minimize} & c^T x + c_{n+1} x_{n+1} \\ \text{subject to} & Ax + A_{n+1} x_{n+1} = b \\ & x, x_{n+1} \geq 0 \end{array}$$

⟶ Solution $(x^\star, 0)$ is still **primal feasible**

Is $y^\star$ still **dual feasible**?

$$A_{n+1}^T y^\star + c_{n+1} \geq 0$$

Yes	**Otherwise**
$(x^\star, 0)$ still **optimal** for new problem	Primal simplex

Adding new variables

Example

$$
\begin{aligned}
\text{minimize} \quad & -60x_1 - 30x_2 - 20x_3 \\
\text{subject to} \quad & 8x_1 + 6x_2 + x_3 \leq 48 \\
& 4x_1 + 2x_2 + 1.5x_3 \leq 20 \\
& 2x_1 + 1.5x_2 + 0.5x_3 \leq 8 \\
& x \geq 0
\end{aligned}
$$

Adding new variables

Example

minimize $-60x_1 - 30x_2 - 20x_3$ -profit

subject to
$$8x_1 + 6x_2 + x_3 \leq 48$$
$$4x_1 + 2x_2 + 1.5x_3 \leq 20$$
$$2x_1 + 1.5x_2 + 0.5x_3 \leq 8$$
$$x \geq 0$$

Adding new variables

Example

minimize	$-60x_1 - 30x_2 - 20x_3$	-profit
subject to	$8x_1 + 6x_2 + x_3 \leq 48$	material
	$4x_1 + 2x_2 + 1.5x_3 \leq 20$	
	$2x_1 + 1.5x_2 + 0.5x_3 \leq 8$	
	$x \geq 0$	

Adding new variables

Example

minimize	$-60x_1 - 30x_2 - 20x_3$	-profit
subject to	$8x_1 + 6x_2 + x_3 \leq 48$	material
	$4x_1 + 2x_2 + 1.5x_3 \leq 20$	production
	$2x_1 + 1.5x_2 + 0.5x_3 \leq 8$	
	$x \geq 0$	

Adding new variables

Example

minimize	$-60x_1 - 30x_2 - 20x_3$	-profit
subject to	$8x_1 + 6x_2 + x_3 \leq 48$	material
	$4x_1 + 2x_2 + 1.5x_3 \leq 20$	production
	$2x_1 + 1.5x_2 + 0.5x_3 \leq 8$	quality control
	$x \geq 0$	

Adding new variables

Example

minimize	$-60x_1 - 30x_2 - 20x_3$	-profit
subject to	$8x_1 + 6x_2 + x_3 \leq 48$	material
	$4x_1 + 2x_2 + 1.5x_3 \leq 20$	production
	$2x_1 + 1.5x_2 + 0.5x_3 \leq 8$	quality control
	$x \geq 0$	

$$c = (-60, -30, -20, 0, 0, 0)$$

minimize	$c^T x$
subject to	$Ax = b$
	$x \geq 0$

$$A = \begin{bmatrix} 8 & 6 & 1 & 1 & 0 & 0 \\ 4 & 2 & 1.5 & 0 & 1 & 0 \\ 2 & 1.5 & 0.5 & 0 & 0 & 1 \end{bmatrix}$$

$$b = (48, 20, 8)$$

Adding new variables

Example

minimize	$-60x_1 - 30x_2 - 20x_3$	-profit
subject to	$8x_1 + 6x_2 + x_3 \leq 48$	material
	$4x_1 + 2x_2 + 1.5x_3 \leq 20$	production
	$2x_1 + 1.5x_2 + 0.5x_3 \leq 8$	quality control
	$x \geq 0$	

$$c = (-60, -30, -20, 0, 0, 0)$$

minimize	$c^T x$
subject to	$Ax = b$
	$x \geq 0$

$$A = \begin{bmatrix} 8 & 6 & 1 & 1 & 0 & 0 \\ 4 & 2 & 1.5 & 0 & 1 & 0 \\ 2 & 1.5 & 0.5 & 0 & 0 & 1 \end{bmatrix}$$

$$b = (48, 20, 8)$$

$x^\star = (2, 0, 8, 24, 0, 0), \quad y^\star = (0, 10, 10), \quad c^T x^\star = -280,$ basis $\{1, 3, 4\}$

Adding new variables

Example: add new product?

minimize $\quad c^T x + c_{n+1} x_{n+1}$

subject to $\quad Ax + A_{n+1} x_{n+1} = b$

$\quad x, x_{n+1} \geq 0$

$$c = (-60, -30, -20, 0, 0, 0, -15)$$

$$A = \begin{bmatrix} 8 & 6 & 1 & 1 & 0 & 0 & 1 \\ 4 & 2 & 1.5 & 0 & 1 & 0 & 1 \\ 2 & 1.5 & 0.5 & 0 & 0 & 1 & 1 \end{bmatrix}$$

$$b = (48, 20, 8)$$

Adding new variables

Example: add new product?

$$c = (-60, -30, -20, 0, 0, 0, -15)$$

$$\begin{aligned} \text{minimize} \quad & c^T x + c_{n+1} x_{n+1} \\ \text{subject to} \quad & Ax + A_{n+1} x_{n+1} = b \\ & x, x_{n+1} \geq 0 \end{aligned}$$

$$A = \begin{bmatrix} 8 & 6 & 1 & 1 & 0 & 0 & 1 \\ 4 & 2 & 1.5 & 0 & 1 & 0 & 1 \\ 2 & 1.5 & 0.5 & 0 & 0 & 1 & 1 \end{bmatrix}$$

$$b = (48, 20, 8)$$

Previous solution

$x^\star = (2, 0, 8, 24, 0, 0), \quad y^\star = (0, 10, 10), \quad c^T x^\star = -280,$ basis $\{1, 3, 4\}$

Adding new variables

Example: add new product?

$$c = (-60, -30, -20, 0, 0, 0, -15)$$

$$A = \begin{bmatrix} 8 & 6 & 1 & 1 & 0 & 0 & 1 \\ 4 & 2 & 1.5 & 0 & 1 & 0 & 1 \\ 2 & 1.5 & 0.5 & 0 & 0 & 1 & 1 \end{bmatrix}$$

$$b = (48, 20, 8)$$

$$\begin{array}{ll} \text{minimize} & c^T x + c_{n+1} x_{n+1} \\ \text{subject to} & Ax + A_{n+1} x_{n+1} = b \\ & x, x_{n+1} \geq 0 \end{array}$$

Previous solution

$x^\star = (2, 0, 8, 24, 0, 0), \quad y^\star = (0, 10, 10), \quad c^T x^\star = -280,$ basis $\{1, 3, 4\}$

Still optimal (do not add new product)

$$A_{n+1}^T y^\star + c_{n+1} = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 10 \\ 10 \end{bmatrix} - 15 = 5 \geq 0$$

Adding new constraints

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

Solution $x^\star, y^\star$

Adding new constraints

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & x \geq 0 \end{array} \longrightarrow \begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & a_{m+1}^T x = b_{m+1} \\ & x \geq 0 \end{array}$$

Solution $x^\star, y^\star$

Adding new constraints

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & x \geq 0 \end{array} \longrightarrow \begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & a_{m+1}^T x = b_{m+1} \\ & x \geq 0 \end{array}$$

Solution $x^\star, y^\star$

Dual

$$\begin{array}{ll} \text{maximize} & -b^T y \\ \text{subject to} & A^T y + a_{m+1} y_{m+1} + c \geq 0 \end{array}$$

Adding new constraints

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & x \geq 0 \end{array} \longrightarrow \begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & a_{m+1}^T x = b_{m+1} \\ & x \geq 0 \end{array}$$

Solution $x^\star, y^\star$

Dual

$$\begin{array}{ll} \text{maximize} & -b^T y \\ \text{subject to} & A^T y + a_{m+1} y_{m+1} + c \geq 0 \end{array}$$

Solution $x^\star, (y^\star, 0)$ **optimal** for the new problem?

Adding new constraints

Optimality conditions

maximize	$-b^T y$
subject to	$A^T y + a_{m+1} y_{m+1} + c \geq 0$

$\longrightarrow$ Solution $(y^\star, 0)$ is still **dual feasible**

Adding new constraints

Optimality conditions

maximize $\quad -b^T y$

subject to $\quad A^T y + a_{m+1} y_{m+1} + c \geq 0$ ⟶ Solution $(y^\star, 0)$ is still **dual feasible**

Is $x^\star$ still **primal feasible**?

$$
\begin{aligned}
&Ax = b \\
&a_{m+1}^T x = b_{m+1} \\
&x, x_{n+1} \geq 0
\end{aligned}
$$

Adding new constraints

Optimality conditions

maximize $-b^T y$

subject to $A^T y + a_{m+1} y_{m+1} + c \geq 0$ ⟶ Solution $(y^\star, 0)$ is still **dual feasible**

Is $x^\star$ still **primal feasible**?

$$Ax = b$$

$$a_{m+1}^T x = b_{m+1}$$

x [redacted] ≥ 0

Yes

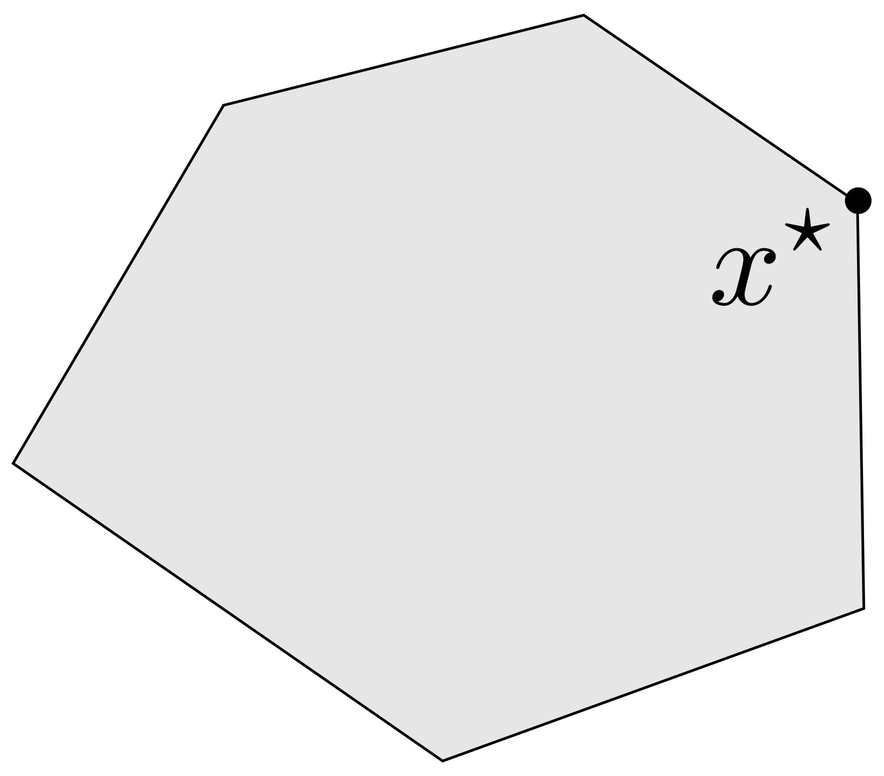
[redacted] $x^\star$ [redacted] still **optimal** for new problem

Otherwise

Dual simplex

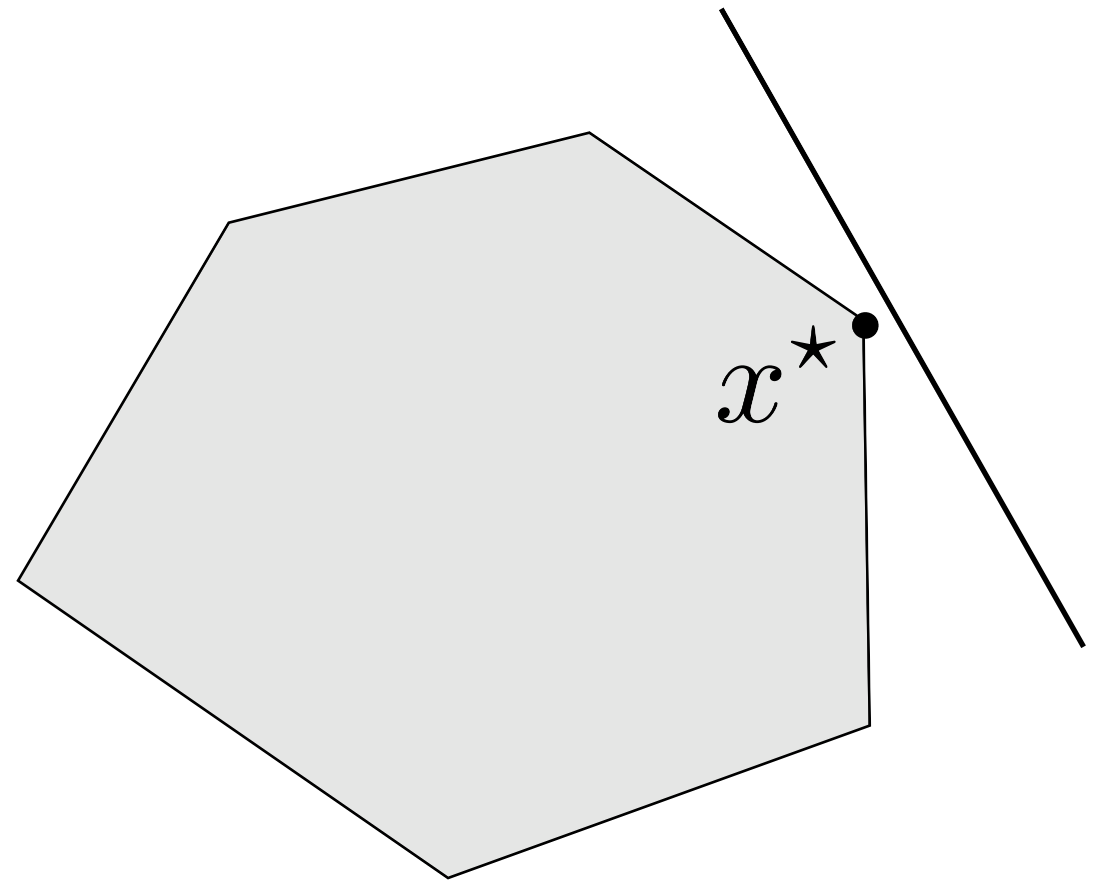
Adding new constraints

Example

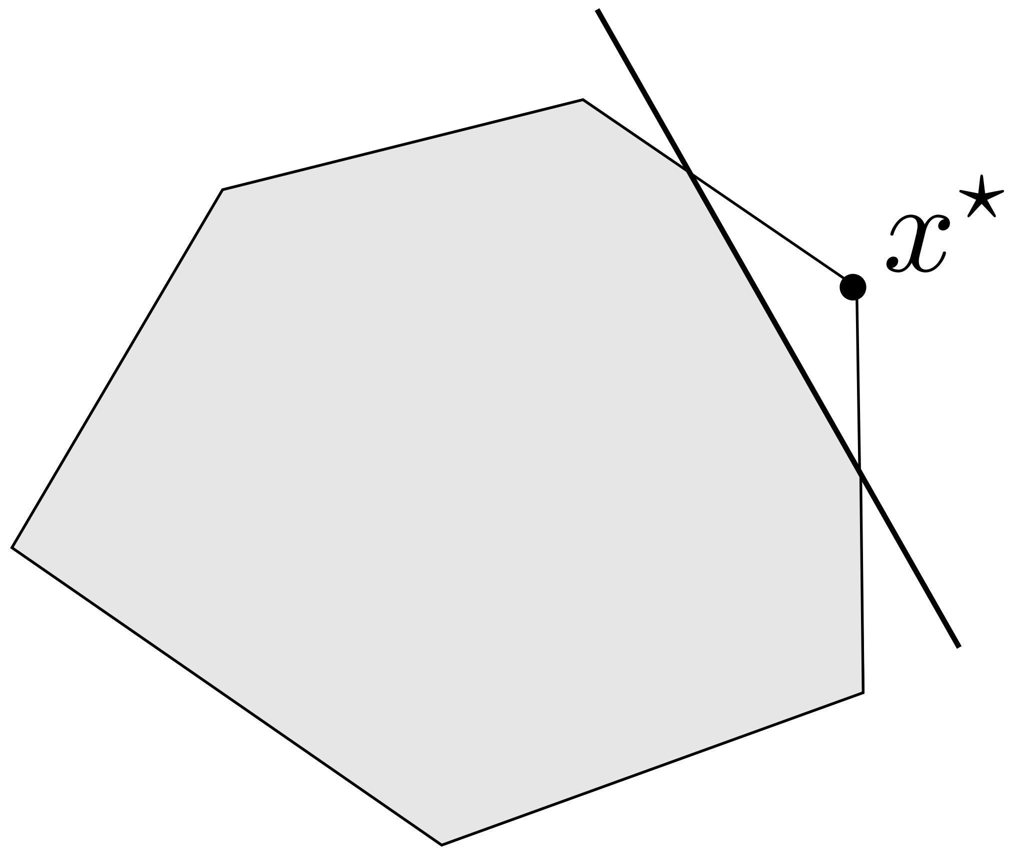


Add new constraint →

$x^\star$ still **feasible**



$x^\star$ **infeasible**



Changing problem data

Information from primal-dual solution

Goal: extract information from $x^\star, y^\star$ about their sensitivity with respect to changes in problem data

Modified LP

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

Optimal cost $p^\star(u)$

Global sensitivity

Dual of modified LP

$$\begin{array}{ll} \text{maximize} & -(b+u)^T y \\ \text{subject to} & A^T y + c \geq 0 \end{array}$$

Global sensitivity

Dual of modified LP

$$\begin{array}{ll} \text{maximize} & -(b+u)^T y \\ \text{subject to} & A^T y + c \geq 0 \end{array}$$

Global lower bound

Given $y^\star$ a dual optimal solution for $u = 0$, then

$$\begin{aligned} p^\star(u) &\geq -(b+u)^T y^\star \\ &= p^\star(0) - u^T y^\star \end{aligned}$$

(from weak duality and dual feasibility)

Global sensitivity

Dual of modified LP

$$\begin{array}{ll}\text{maximize} & -(b+u)^T y \\ \text{subject to} & A^T y + c \geq 0\end{array}$$

Global lower bound

Given $y^\star$ a dual optimal solution for $u = 0$, then

$$\begin{aligned} p^\star(u) &\geq -(b+u)^T y^\star \\ &= p^\star(0) - u^T y^\star \end{aligned}$$

(from weak duality and dual feasibility)

It holds for any u

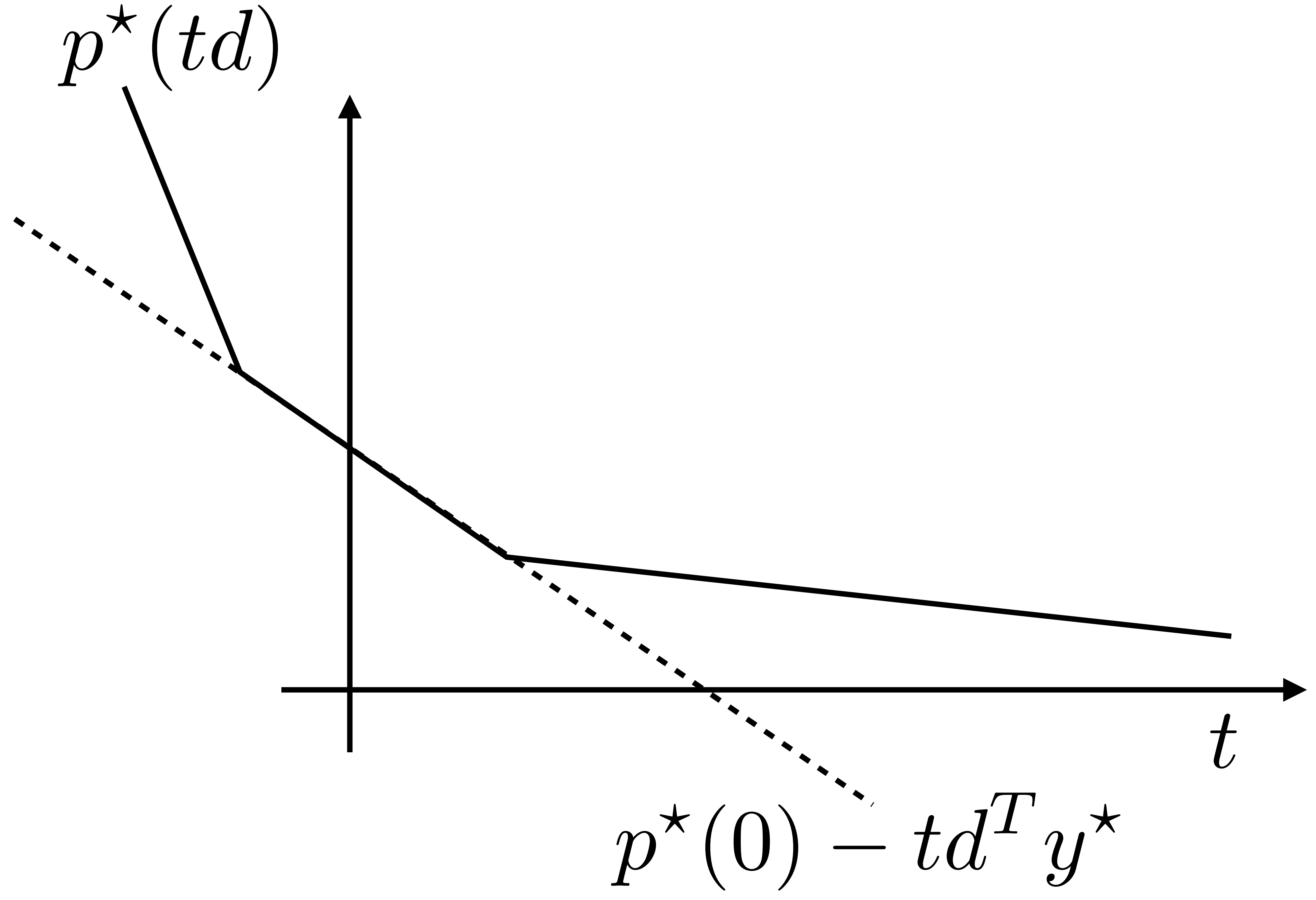
Global sensitivity

Example

Take $u = td$ with $d \in \mathbf{R}^m$ fixed

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

$p^\star(td)$ is the optimal value as a function of t

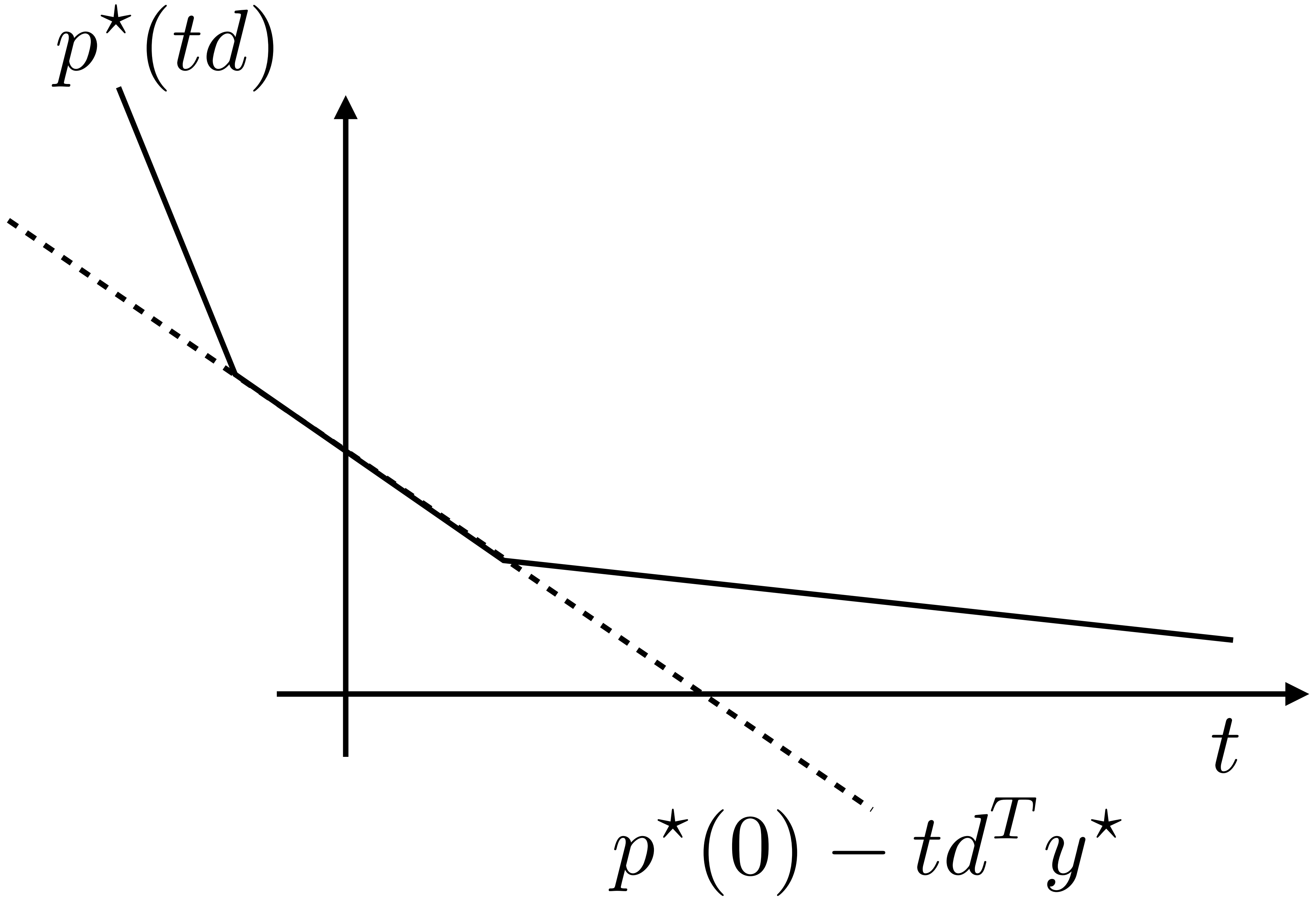


Global sensitivity

Example

Take $u = td$ with $d \in \mathbf{R}^m$ fixed

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



$p^\star(td)$ is the optimal value as a function of t

Sensitivity information (assuming $d^T y^\star \geq 0$)

- $t < 0$ the optimal value increases
- $t > 0$ the optimal value decreases (not so much if t is small)

Optimal value function

$$p^\star(u) = \min\{c^T x \mid Ax = b + u,\ x \geq 0\}$$

Assumption: $p^\star(0)$ is finite

Properties

- $p^\star(u) > -\infty$ everywhere (from global lower bound)
- the domain $\{u \mid p^\star(u) < +\infty\}$ is a polyhedron
- $p^\star(u)$ is piecewise-linear on its domain

Optimal value function is piecewise linear

Proof

$$p^\star(u) = \min\{c^T x \mid Ax = b + u,\ x \geq 0\}$$

Optimal value function is piecewise linear

Proof

$$p^\star(u) = \min\{c^T x \mid Ax = b + u,\ x \geq 0\}$$

Dual feasible set

$$D = \{y \mid A^T y + c \geq 0\}$$

Assumption: $p^\star(0)$ is finite

Optimal value function is piecewise linear

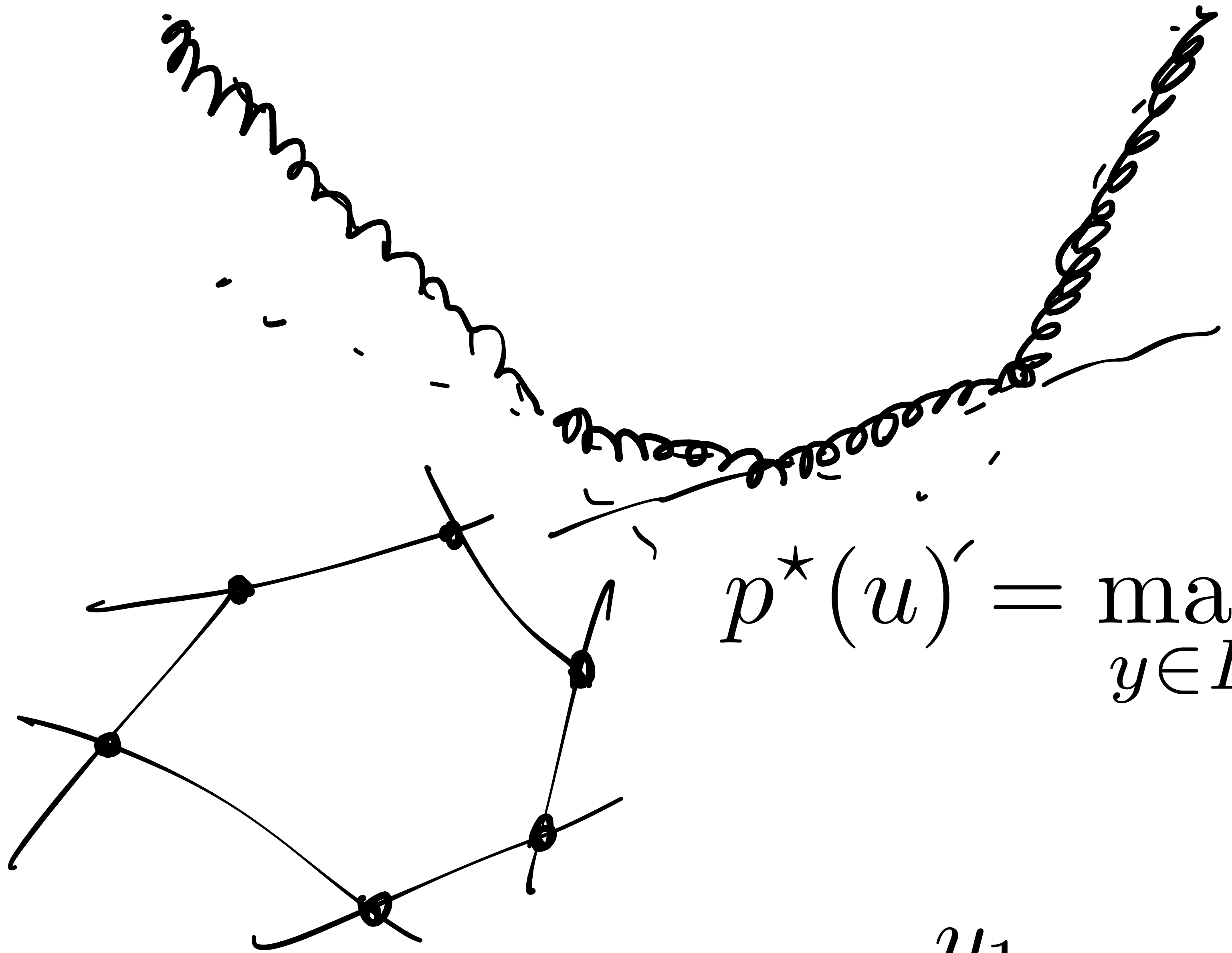
Proof

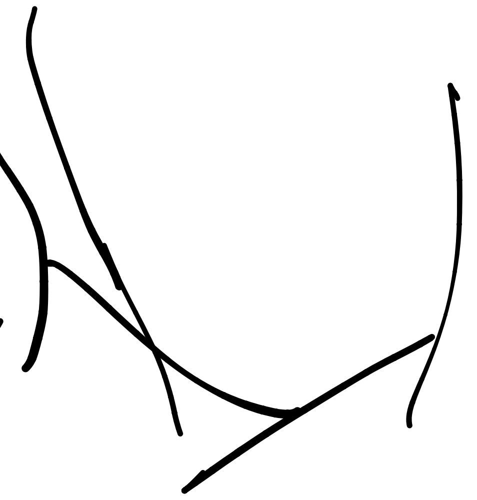
$$p^\star(u) = \min\{c^T x \mid Ax = b + u,\ x \geq 0\}$$

Dual feasible set

$$D = \{y \mid A^T y + c \geq 0\}$$

Assumption: $p^\star(0)$ is finite





If $p^\star(u)$ finite

$$p^\star(u) = \max_{y \in D} -(b-u)^T y = \max_{k=1,\dots,r} y_k^T u - b^T y_k$$

$y_1, \dots, y_r$ are the extreme points of D

Local sensitivity

u in neighborhood of the origin

Original LP

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

Local sensitivity

u in neighborhood of the origin

Original LP

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

$\longrightarrow$

Optimal solution

$$x_B^\star = B^{-1} b$$

$$y^\star = -B^{-T} c_B$$

Local sensitivity

u in neighborhood of the origin

Original LP

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

→

Optimal solution

$$x_B^\star = B^{-1} b$$

$$y^\star = -B^{-T} c_B$$

Modified LP

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

Modified dual

$$\begin{array}{ll} \text{maximize} & -(b+u)^T y \\ \text{subject to} & A^T y + c \geq 0 \end{array}$$

Optimal basis does not change

Local sensitivity

u in neighborhood of the origin

Original LP

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

→

Optimal solution

$$x_B^\star = B^{-1} b$$

$$y^\star = -B^{-T} c_B$$

Modified LP

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

Modified dual

$$\begin{array}{ll} \text{maximize} & -(b+u)^T y \\ \text{subject to} & A^T y + c \geq 0 \end{array}$$

Optimal basis does not change

Modified optimal solution

$$x_B^\star(u) = B^{-1}(b+u) = x_B^\star + B^{-1} u$$

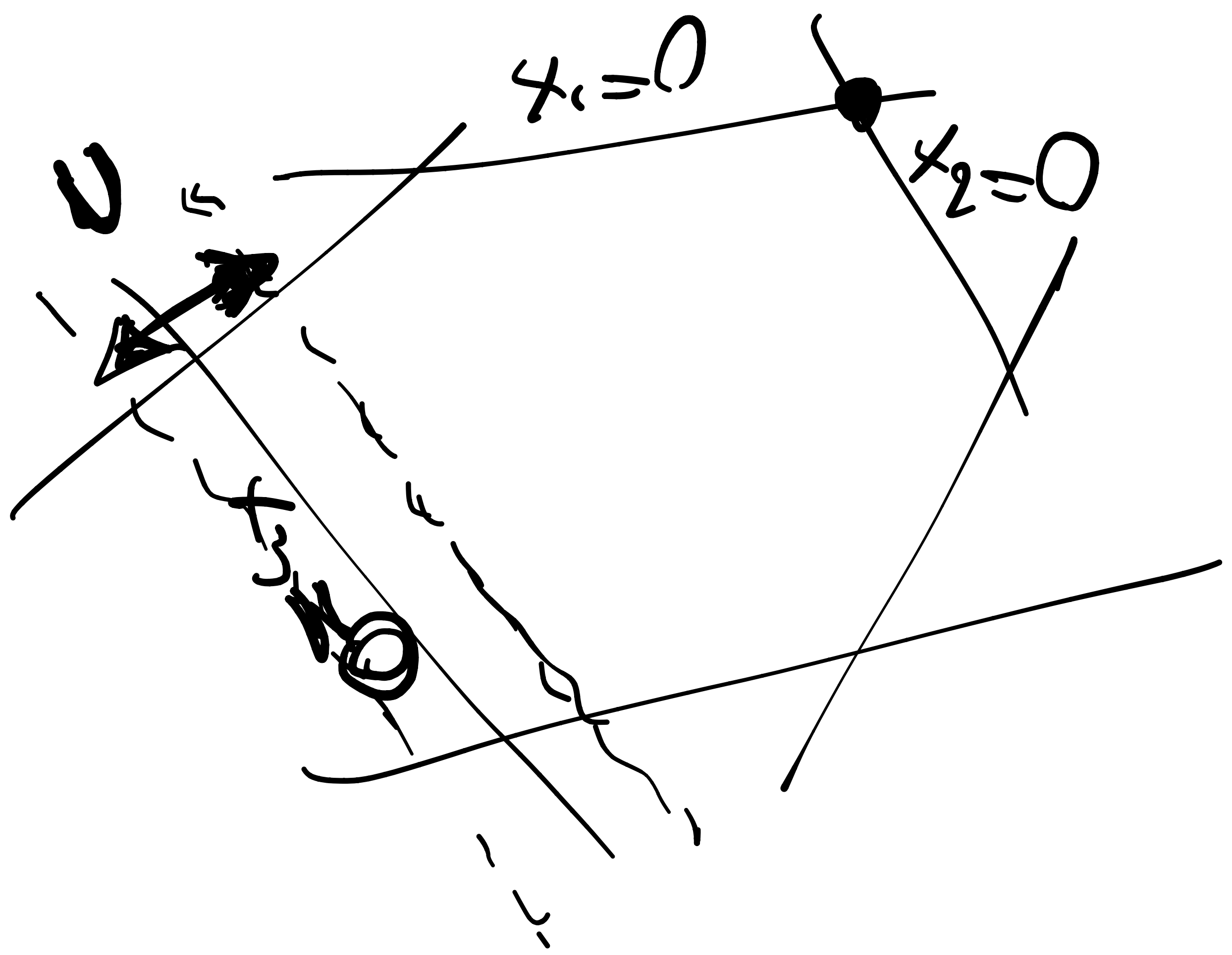
$$y^\star(u) = y^\star$$

Derivative of the optimal value function

Optimal value function

$$
\begin{aligned}
p^\star(u) &= c^T x^\star(u) \\
&= c^T x^\star + c_B^T B^{-1} u \\
&= p^\star(0) - y^{\star T} u \qquad \text{(affine for small } u\text{)}
\end{aligned}
$$

Derivative of the optimal value function



Optimal value function

$$
\begin{aligned}
p^\star(u) &= c^T x^\star(u) \\
&= c^T x^\star + c_B^T B^{-1} u \\
&= p^\star(0) - y^{\star T} u
\end{aligned}
$$

(affine for small u)

Local derivative

$$
\frac{\partial p^\star(u)}{\partial u} = -y^\star
$$

($y^\star$ are the **shadow prices**)

Sensitivity example

$$
\begin{array}{ll}
\text{minimize} & -60x_1 - 30x_2 - 20x_3 \\
\text{subject to} & 8x_1 + 6x_2 + x_3 \leq 48 \\
 & 4x_1 + 2x_2 + 1.5x_3 \leq 20 \\
 & 2x_1 + 1.5x_2 + 0.5x_3 \leq 8 \\
 & x \geq 0
\end{array}
$$

Sensitivity example

$$\begin{array}{ll} \text{minimize} & -60x_1 - 30x_2 - 20x_3 \\ \text{subject to} & 8x_1 + 6x_2 + x_3 \leq 48 \\ & 4x_1 + 2x_2 + 1.5x_3 \leq 20 \\ & 2x_1 + 1.5x_2 + 0.5x_3 \leq 8 \\ & x \geq 0 \end{array}$$

-profit

Sensitivity example

minimize	$-60x_1 - 30x_2 - 20x_3$	-profit
subject to	$8x_1 + 6x_2 + x_3 \leq 48$	material
	$4x_1 + 2x_2 + 1.5x_3 \leq 20$	
	$2x_1 + 1.5x_2 + 0.5x_3 \leq 8$	
	$x \geq 0$	

Sensitivity example

$$
\begin{array}{lll}
\text{minimize} & -60x_1 - 30x_2 - 20x_3 & \text{-profit} \\
\text{subject to} & 8x_1 + 6x_2 + x_3 \le 48 & \text{material} \\
& 4x_1 + 2x_2 + 1.5x_3 \le 20 & \text{production} \\
& 2x_1 + 1.5x_2 + 0.5x_3 \le 8 & \\
& x \ge 0 &
\end{array}
$$

Sensitivity example

minimize	$-60x_1 - 30x_2 - 20x_3$	-profit
subject to	$8x_1 + 6x_2 + x_3 \leq 48$	material
	$4x_1 + 2x_2 + 1.5x_3 \leq 20$	production
	$2x_1 + 1.5x_2 + 0.5x_3 \leq 8$	quality control
	$x \geq 0$	

Sensitivity example

minimize	$-60x_1 - 30x_2 - 20x_3$	-profit
subject to	$8x_1 + 6x_2 + x_3 \leq 48$	material
	$4x_1 + 2x_2 + 1.5x_3 \leq 20$	production
	$2x_1 + 1.5x_2 + 0.5x_3 \leq 8$	quality control
	$x \geq 0$	

$x^\star = (2, 0, 8, 24, 0, 0), \quad y^\star = (0, 10, 10), \quad c^T x^\star = -280,$ basis $\{1, 3, 4\}$

Sensitivity example

minimize	$-60x_1 - 30x_2 - 20x_3$	-profit
subject to	$8x_1 + 6x_2 + x_3 \leq 48$	material
	$4x_1 + 2x_2 + 1.5x_3 \leq 20$	production
	$2x_1 + 1.5x_2 + 0.5x_3 \leq 8$	quality control
	$x \geq 0$	

$x^\star = (2, 0, 8, 24, 0, 0), \quad y^\star = (0, 10, 10), \quad c^T x^\star = -280,$ basis $\{1, 3, 4\}$

What does $y_3^\star = 10$ mean?

Sensitivity example

minimize	$-60x_1 - 30x_2 - 20x_3$	-profit
subject to	$8x_1 + 6x_2 + x_3 \leq 48$	material
	$4x_1 + 2x_2 + 1.5x_3 \leq 20$	production
	$2x_1 + 1.5x_2 + 0.5x_3 \leq 8$	quality control
	$x \geq 0$	

$x^\star = (2, 0, 8, 24, 0, 0), \quad y^\star = (0, 10, 10), \quad c^T x^\star = -280,$ basis $\{1, 3, 4\}$

What does $y_3^\star = 10$ mean?

Let's increase the ~~production budget~~ QUALITY CONTROL BUDGET by 1, i.e., $u = (0, 0, 1)$

$$p^\star(10) = p^\star(0) - y^{\star T} u = -280 - 10 = -290$$

Differentiable optimization

Training a neural network

Single layer model

Datapoints (x^i, y^i)

$$z = f(x;\theta)$$

Training a neural network

Single layer model

Datapoints (x^i, y^i)

$$z = f(x;\theta)$$

Training

minimize $\quad \mathcal{L}(\theta) = \sum_{i=1}^{n} \ell(z^i, y^i)$

Gradient descent (more on this later)

$$\theta \leftarrow \theta - t\nabla_\theta \mathcal{L}(\theta)$$

Training a neural network

Single layer model

Datapoints (x^i, y^i)

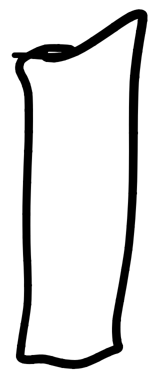
$$z = f(x; \theta)$$

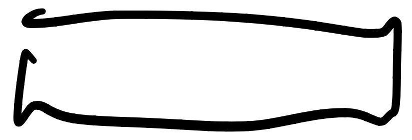
Training

minimize $\quad \mathcal{L}(\theta) = \sum_{i=1}^{n} \ell(z^i, y^i)$

Gradient descent (more on this later)

$$\theta \leftarrow \theta - t \nabla_\theta \mathcal{L}(\theta)$$





Sensitivity

$$\nabla_\theta \mathcal{L} = \left(\frac{\partial \mathcal{L}}{\partial \theta}\right)^T = \left(\frac{\partial \ell}{\partial z}\frac{\partial z}{\partial \theta}\right)^T = \left(\frac{\partial z}{\partial \theta}\right)^T \nabla_z \mathcal{L}$$

Training a neural network

Single layer model

Datapoints (x^i, y^i)

$$z = f(x; \theta)$$

Training

minimize $\quad \mathcal{L}(\theta) = \sum_{i=1}^{n} \ell(z^i, y^i)$

Gradient descent (more on this later)

$$\theta \leftarrow \theta - t\nabla_\theta \mathcal{L}(\theta)$$

Sensitivity

$$\nabla_\theta \mathcal{L} = \left(\frac{\partial \mathcal{L}}{\partial \theta}\right)^T = \left(\frac{\partial \ell}{\partial z}\frac{\partial z}{\partial \theta}\right)^T = \left(\frac{\partial z}{\partial \theta}\right)^T \nabla_z \mathcal{L}$$

Can f be an **optimization layer**?

Optimization layers

$$z = x^\star = \begin{array}{ll} \text{argmin}_x & c^T x \\ \text{subject to} & Ax \le b \end{array}$$

Parameters: $\theta = \{c, A, b\}$
Solution $x^\star(\theta)$

Optimization layers

$$z = x^\star = \underset{x}{\text{argmin}} \quad c^T x \quad \text{subject to} \quad Ax \le b$$

Parameters: $\theta = \{c, A, b\}$
Solution $x^\star(\theta)$

Features

- Add **domain knowledge** and **hard constraints**
- **End-to-end** training and optimization
- Nice theory and algorithms for general **convex optimization**
- **Applications** in RL, control, meta-learning, game theory, etc.

Optimization layers

$$z = x^\star = \begin{array}{ll} \underset{x}{\text{argmin}} & c^T x \\ \text{subject to} & Ax \leq b \end{array}$$

Parameters: $\theta = \{c, A, b\}$
Solution $x^\star(\theta)$

Features

- Add **domain knowledge** and **hard constraints**
- **End-to-end** training and optimization
- Nice theory and algorithms for general **convex optimization**
- **Applications** in RL, control, meta-learning, game theory, etc.

Goal

Compute $\dfrac{\partial x^\star}{\partial \theta}$

Implicit function theorem

Given θ and $x(\theta)$ satisfying

$$r(\theta, x(\theta)) = 0 \qquad \text{(}x(\theta)\text{ is implicitly defined by }r\text{)}$$

Implicit function theorem

Given θ and $x(\theta)$ satisfying

$$r(\theta, x(\theta)) = 0 \qquad \text{(}x(\theta)\text{ is implicitly defined by }r\text{)}$$

Then, under mild assumptions (non-singularity),

$$\frac{\partial x(\theta)}{\partial \theta} = -\left(\frac{\partial r(\theta, x(\theta))}{\partial x}\right)^{-1} \frac{\partial r(\theta, x(\theta))}{\partial \theta}$$

[Theorem 1. B.1, Dontchev and Rockafellar 2009]

Optimality conditions

minimize $c^T x$
subject to $Ax \leq b$

Parameters: $\theta = \{c, A, b\}$
Solution $x^\star(\theta)$

Solve and obtain primal-dual pair $x^\star, y^\star$ (forward-pass)

Optimality conditions

$$A^T y + c = 0$$
$$\mathbf{diag}(y)(Ax - b) = 0$$
$$y \geq 0,\ b - Ax \geq 0$$

Optimality conditions

minimize $c^T x$
subject to $Ax \leq b$

Parameters: $\theta = \{c, A, b\}$
Solution $x^\star(\theta)$

Solve and obtain primal-dual pair $x^\star, y^\star$ (forward-pass)

Optimality conditions

$$A^T y + c = 0$$
$$\mathbf{diag}(y)(Ax - b) = 0$$
$$y \geq 0,\ b - Ax \geq 0$$

Mapping $r(\theta, x(\theta)) = 0$

Computing derivatives

Take differentials

$$A^T y^\star + c = 0$$
$$\mathbf{diag}(y^\star)(Ax^\star - b) = 0$$

$\longrightarrow$

$$\mathrm{d}A^T y^\star + A^T \mathrm{d}y = 0$$
$$\mathbf{diag}(Ax - b)\mathrm{d}y + \mathbf{diag}(y^\star)(\mathrm{d}Ax^\star + A\mathrm{d}x - \mathrm{d}b) = 0$$

Computing derivatives

Take differentials

$$A^T y^\star + c = 0$$
$$\mathbf{diag}(y^\star)(Ax - b) = 0$$

$$\longrightarrow$$

$$\mathrm{d}A^T y^\star + A^T \mathrm{d}y = 0$$
$$\mathbf{diag}(Ax - b)\mathrm{d}y + \mathbf{diag}(y^\star)(\mathrm{d}Ax^\star + A\mathrm{d}x - \mathrm{d}b) = 0$$

Linear system

$$\begin{bmatrix} 0 & A^T \\ \mathbf{diag}(y^\star)A & \mathbf{diag}(Ax^\star - b) \end{bmatrix} \begin{bmatrix} \mathrm{d}x \\ \mathrm{d}y \end{bmatrix} = - \begin{bmatrix} \mathrm{d}A^T y^\star + \mathrm{d}c \\ \mathbf{diag}(y^\star)(\mathrm{d}Ax^\star - \mathrm{d}b) \end{bmatrix}$$

Computing derivatives

Take differentials

$$A^T y^\star + c = 0 \qquad \longrightarrow \qquad \mathrm{d}A^T y^\star + A^T \mathrm{d}y = 0$$

$$\mathbf{diag}(y^\star)(Ax - b) = 0 \qquad \qquad \mathbf{diag}(Ax - b)\mathrm{d}y + \mathbf{diag}(y^\star)(\mathrm{d}Ax^\star + A\mathrm{d}x - \mathrm{d}b) = 0$$

Linear system

$$\begin{bmatrix} 0 & A^T \\ \mathbf{diag}(y^\star)A & \mathbf{diag}(Ax^\star - b) \end{bmatrix} \begin{bmatrix} \mathrm{d}x \\ \mathrm{d}y \end{bmatrix} = - \begin{bmatrix} \mathrm{d}A^T y^\star + \mathrm{d}c \\ \mathbf{diag}(y^\star)(\mathrm{d}Ax^\star - \mathrm{d}b) \end{bmatrix}$$

Example: How does x change with b_1?

Set $\mathrm{d}b = e_1, \mathrm{d}A = 0, \mathrm{d}c = 0$ and solve the linear system.

The solution $\mathrm{d}x$ will correspond to $\frac{\partial x}{\partial b}$

Is it always differentiable?

The linear system matrix must be invertible
(the problem must have unique solution)

$$M \begin{bmatrix} \mathrm{d}x \\ \mathrm{d}u \end{bmatrix} = -q$$

Remember: implicit function theorem

$$\frac{\partial x(\theta)}{\partial \theta} = -\left(\frac{\partial r(\theta, x(\theta))}{\partial x}\right)^{-1} \frac{\partial r(\theta, x(\theta))}{\partial \theta}$$

Is it always differentiable?

The linear system matrix must be invertible
(the problem must have unique solution)

$$M \begin{bmatrix} \mathrm{d}x \\ \mathrm{d}u \end{bmatrix} = -q$$

Remember: implicit function theorem

$$\frac{\partial x(\theta)}{\partial \theta} = -\left(\frac{\partial r(\theta, x(\theta))}{\partial x}\right)^{-1} \frac{\partial r(\theta, x(\theta))}{\partial \theta}$$

If not, **least squares** (subdifferential)

minimize $\quad \left\| M \begin{bmatrix} \mathrm{d}x \\ \mathrm{d}u \end{bmatrix} + q \right\|_2^2$

Example

Learning to play Sudoku

			3
1			
		4	
4			1

2	4	1	3
1	3	2	4
3	1	4	2
4	2	3	1

Sudoku constraint satisfaction problem

$$\begin{array}{ll} \text{minimize} & 0 \\ \text{subject to} & Ax = b \\ & x \in \mathbf{Z}^d \end{array}$$

[OptNet: Differentiable Optimization as a Layer in Neural Networks, B. Amos and J. Z. Kolter ICML 2017]

Example

Learning to play Sudoku

			3
1			
		4	
4			1

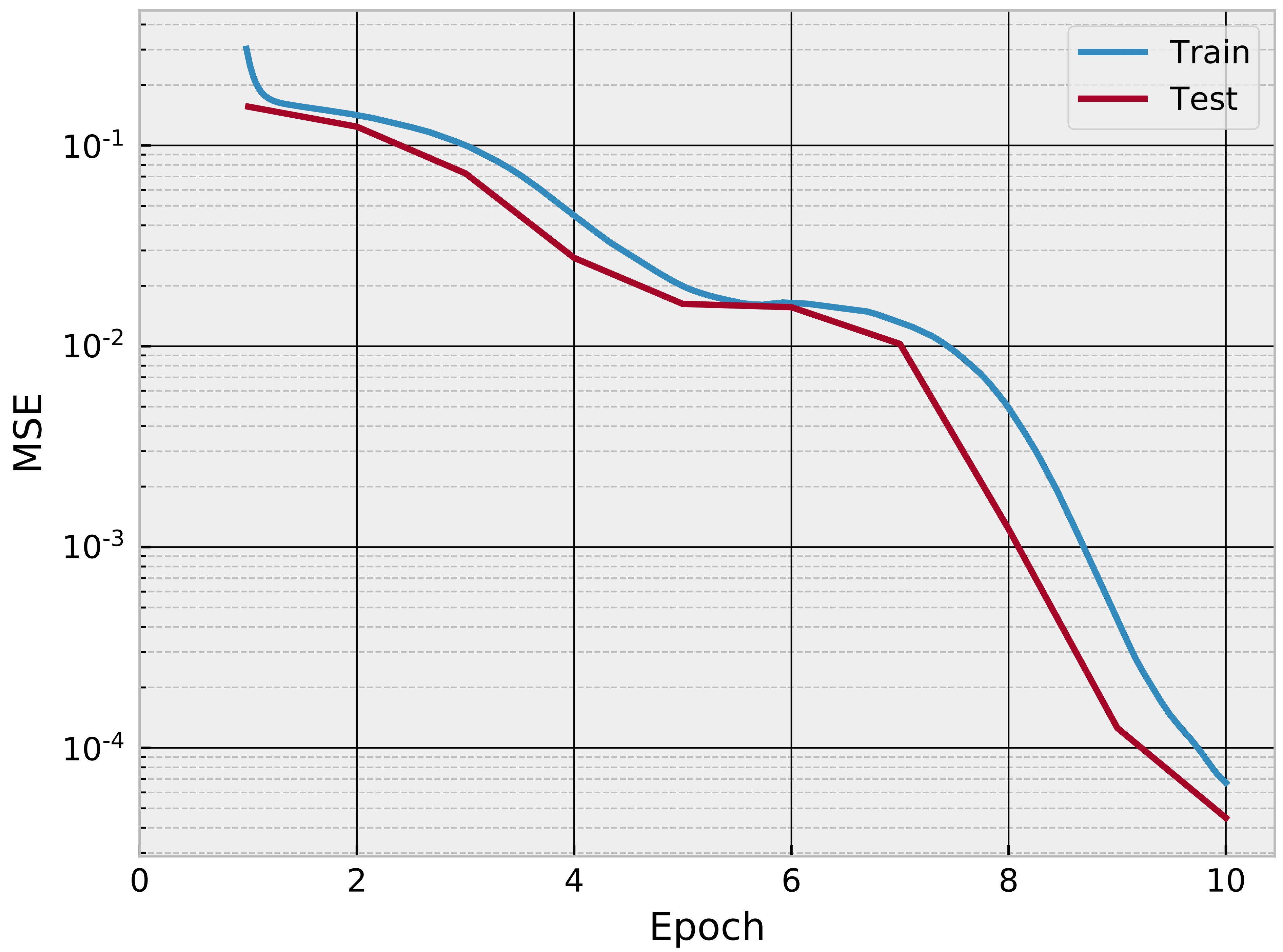
2	4	1	3
1	3	2	4
3	1	4	2
4	2	3	1

Sudoku constraint satisfaction problem

$$\begin{array}{ll} \text{minimize} & 0 \\ \text{subject to} & Ax = b \\ & x \in \mathbf{Z}^d \end{array}$$

Linear optimization layer (parameters $\theta = \{A, b\}$)

$$z = x^\star = \begin{array}{ll} \underset{x}{\text{argmin}} & 0 \\ \text{subject to} & Ax = b \\ & x \geq 0 \end{array}$$



[OptNet: Differentiable Optimization as a Layer in Neural Networks, B. Amos and J. Z. Kolter ICML 2017]

Sensitivity analysis in linear optimization

Today, we learned to:

- **Use** the most appropriate primal/dual simplex algorithm when variables and/or constraints are added

- **Analyze** sensitivity of the cost with respect to change in the data

- **Apply** sensitivity analysis to differentiable linear optimization layers

Next lecture

- Barrier methods for linear optimization

Next lecture

- Barrier methods for linear optimization