Optimization Lecture 6+7

Qingfu Zhang

Dept of CS , CityU

Outline

Recap

- conjugate of $f(x)$
- first order condition for a differentiable convex function
- Jensen inequality
- implicit constraints, explicit constraints
- first order sufficient and necessary optimality condition for convex optimization problem.

Lagrangian and dual function

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 $\operatorname{dom} L = \mathcal{D} \times \mathbf{R}^m \times \mathbf{R}^p$,

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

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

Lagrange dual function

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

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

- g is concave, can be $-\infty$ for some λ, μ
- lower bound property: if $\lambda \geq 0$, then $g(\lambda, \mu) \leq p^\star$
- proof: if $\tilde{x}$ is feasible and $\lambda \geq 0$, then

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

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

Least-norm solution of linear equations

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

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

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

- plug x into L to obtain

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

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

Standard form LP

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

- Lagrangian is

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

- L is affine in x, so

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

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

Equality constrained norm minimization

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

- dual function is

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

  where $\|\mu\|_* = \sup_{\|u\|\le 1} u^T\mu$ is dual norm of $\|\cdot\|$
- lower bound property: $p^\star \ge b^T\mu$ if $\left\|A^T\mu\right\|_* \le 1$

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

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

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

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

Lagrangian dual problem

The Lagrange dual problem

(Lagrange) dual problem

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

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

Example: standard form LP

- primal standard form LP:

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

- dual problem is

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

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

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

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

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.

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

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

if it is strictly feasible, i.e.,there is an $x \in \operatorname{int} \mathcal{D}$ 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 int $\mathcal{D}$ with relint $\mathcal{D}$ (interior relative to affine hull)
  - linear inequalities do not need to hold with strict inequality
- there are many other types of constraint qualifications

Inequality form LP

primal Problem

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

dual function

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

dual problem

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

- $p^\star = d^\star$ except when primal and dual are both infeasible (See books on LP)

Quadratic program

primal problem (assume $P \in \mathbf{S}^n_{++}$)

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

dual function

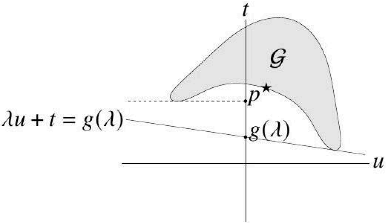
$$g(\lambda) = \inf_x [x^T P x + \lambda^T (Ax - b)] = -\frac{1}{4} \lambda^T A P^{-1} A^T \lambda - b^T \lambda$$

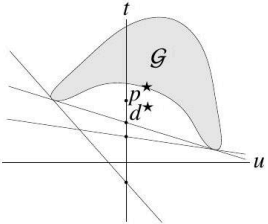
dual problem

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

- from the sharpened Slater's condition: $p^\star = d^\star$ if the primal problem is feasible
- in fact, $p^\star = d^\star$ always

Geometric interpretation

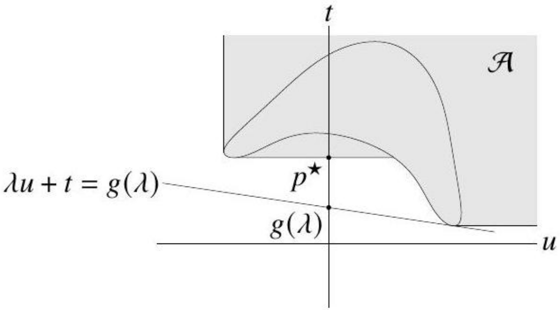




- for simplicity, consider problem with one constraint $f_1(x) \leq 0$
- $\mathcal{G} = \{(f_1(x), f_0(x)) \mid x \in \mathcal{D}\}$ is set of achievable (constraint, objective) values
- interpretation of dual function: $g(\lambda) = \inf_{(u,t)\in\mathcal{G}}(t + \lambda u)$
- $\lambda u + t = g(\lambda)$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

Epigraph variation

- same with $\mathcal{G}$ replaced with $\mathcal{A} = \{(u,t) \mid f_1(x) \leq u, f_0(x) \leq t$ 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

KKT condition

Complementary slackness

Assume (i) strong duality holds, (ii) $x^\star$ is primal optimal, and (iii) $(\lambda^\star, \mu^\star)$ is dual optimal. Then

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

The two inequalities must hold with equality. **Then:**

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

$$\lambda_i^\star > 0 \Longrightarrow f_i\left(x^\star\right) = 0, \quad f_i\left(x^\star\right) < 0 \Longrightarrow \lambda_i^\star = 0$$

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: $\lambda \geq 0$
3. complementary slackness: $\lambda_i f_i(x) = 0, i = 1, \ldots, m$
4. gradient of Lagrangian with respect to x vanishes:

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

Theorem: If strong duality holds and x, λ, μ are optimal, they satisfy the KKT conditions.
How to prove it?

KKT conditions for convex problem

Theorem: If $\tilde{x}, \tilde{\lambda}, \tilde{\mu}$ satisfy KKT for a convex problem, then they are optimal.
Outline of Pf:

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

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

Theorem: If Slater's condition is satisfied. Then, x **is optimal if and only if there exist** λ, μ **that satisfy KKT conditions**
Outline of Pf:

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

Sensitivity analysis

Perturbation and sensitivity analysis

(unperturbed) optimization problem and its dual

$$\begin{array}{lll} \text{minimize} & f_0(x) & \\ \text{subject to} & f_i(x) \leq 0, & i = 1, \ldots, m \\ & h_i(x) = 0, & i = 1, \ldots, p \end{array} \qquad \begin{array}{ll} \text{maximize} & g(\lambda, \mu) \\ \text{subject to} & \lambda \geq 0 \\ & \end{array}$$

perturbed problem and its dual

$$\begin{array}{lll} \text{minimoize} & f_0(x) & \\ \text{subject to} & f_i(x) \leq u_i, & i = 1, \ldots, m \\ & h_i(x) = v_i, & i = 1, \ldots, p \end{array} \qquad \begin{array}{l} \text{maximize } g(\lambda, \mu) - u^T\lambda - v^T\mu \\ \text{subject to } \lambda \geq 0 \\ \end{array}$$

- 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 $\lambda^\star, \mu^\star$ dual optimal.Apply weak duality to perturbed problem:

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

implications:

- 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 $\mu_i^\star$ large and positive: $p^\star$ increases greatly if we take $v_i < 0$
- if $\mu_i^\star$ large and negative: $p^\star$ increases greatly if we take $v_i > 0$
- if $\mu_i^\star$ small and positive: $p^\star$ does not decrease much if we take $v_i > 0$
- if $\mu_i^\star$ 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

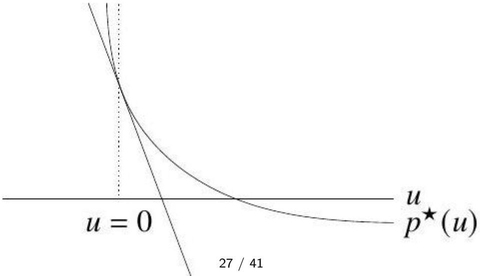
$$\lambda_i^\star = -\frac{\partial p^\star(0,0)}{\partial u_i}, \quad \mu_i^\star = -\frac{\partial p^\star(0,0)}{\partial v_i}$$

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

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

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

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



Problem reformulations

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

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

- dual of reformulated problem is

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

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

Example: Norm approximation

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

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

- dual of (reformulated) norm approximation problem:

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

Theorems of alternatives

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 $x \in \mathbf{R}$,
  - $x > a, x < b$ and $x > b, x < a$ are weak 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

System A is called a strong alternative to System B iff exactly one is feasible.
System A is called a weak alternative to System B if that system A is feasible implies B is infeasible.

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

$$\lambda \succeq 0, \quad g(\lambda, \mu) > 0$$

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

$$\lambda \succeq 0, \quad g(\lambda, \mu) \geq 1$$

Example: Nonnegative solution of linear equations

- consider system

$$Ax = b, \quad x \geq 0$$

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

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

$$A^T \mu \geq 0, \quad \mu^T b \leq -1$$

(we can replace $\mu^T b \leq -1$ with $\mu^T b = -1$)

Farkas' lemma

Farkas' lemma:

$$Ax \leq 0, c^T x < 0 \quad (1) \quad \textit{and} \quad A^T y + c = 0, \quad y \geq 0 \quad (2)$$

are strong alternatives
Proof: Consider (primal) LP and its dual

$$\begin{array}{llll} \text{minimize} & c^T x & \text{maximize} & 0 \\ \text{subject to} & Ax \leq 0 & \text{subject to} & A^T y + c = 0, y \geq 0 \end{array}$$

- $p* = 0$ or $-\infty$. And $d* = 0$ or $-\infty$.
- If (1) is infeasible, then $p^* = 0$.
- If (1) is feasible, then $p^* = -\infty$.
- If (2) is feasible, then $d^* = 0$.
- If (2) is infeasible, then $d^* = -\infty$. Noting that $x = 0$ is feasible for the primal problem, Strong duality for LP holds (see slide on inequality form LP).
- (1) and (2) are strong alternatives

another version

Let $A \in R^{m \times n}$ and $b \in R^{m}$. Either

- $Ax \leq b$ has a solution, or
- $yA = 0, y \geq 0, b^{T}y < 0$ has a solution

but not both.
How to prove it: rewrite $Ax \leq b$ as ???. (tutorial question)

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 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
- 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

- 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 $\mathbf{1}^T y = 1$
- 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

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

- with prices p, there is an arbitrage

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

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

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