Introduction to Stochastic Programming

9/4/2024
26:711:555
Stochastic Programming

Homeworks: 1-2 problems involving modeling
and programming for computation.

No in-class exams. Room 1076 before/after class in Newark.

The existance of RVs in Stochastic Programming results in many different problem formulation, unlike linear/nonlinear prog.

Applications:

1. Linear Regression Model: $Y = X\beta + \epsilon$

   $Y = (y_1, \dots, y_n)^T$ are the responses (labels),

   X is an $n \times p$ feature matrix,

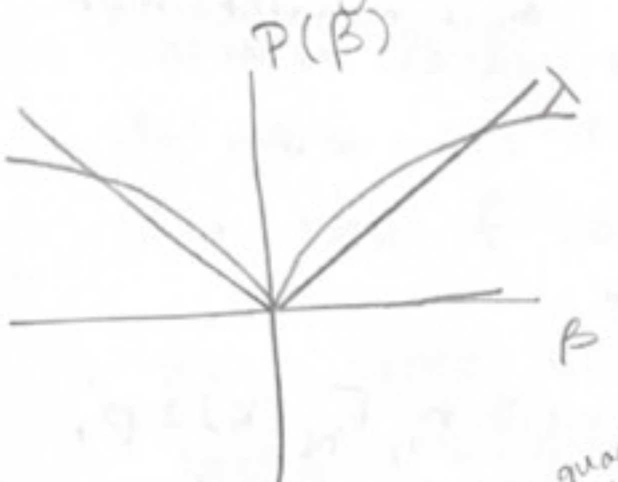
   β is vector of p unknown coefficients to be estimated.

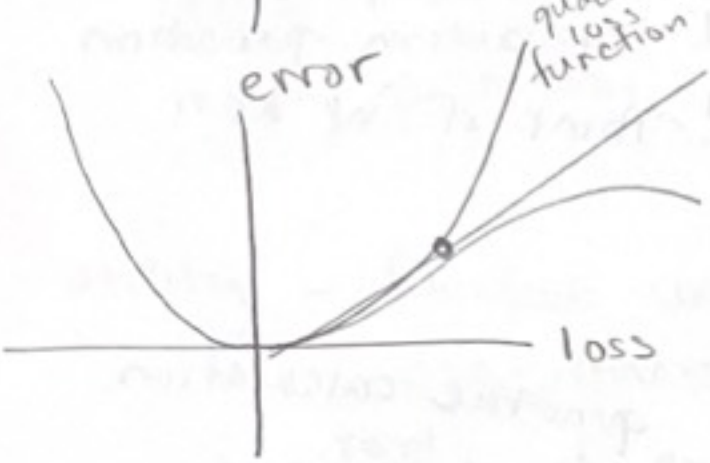
Generalized Lasso framework

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \left\{ \frac{1}{2n} \|Y - X\beta\| \right.$$

stochastic subgradient methods. (Ukrainian)

nonconvex | non differentiable penalty functions are used to bring coefficients close to zero to zero but ignore large coefficients. loss functions also to reduce bias caused by large outliers.





2 - stage Production Problem

2^{nd} stage is simple linear programming problem with 2 parameters (Demand and parts ordered)
D X

$Q(x, D)$ * - optimal value of 2^{nd} stage problem is random

$$\min_{x} \; c^T x + \sum_{k=1}^{K} p_k Q(x, d^k)$$

$Q(x, d^k)$ - optimal value of LP, usually will not have a gradient.

To solve as one large problem over all scenarios is computationally intractable except for toy examples.

So we must exploit statistical nature of problem!

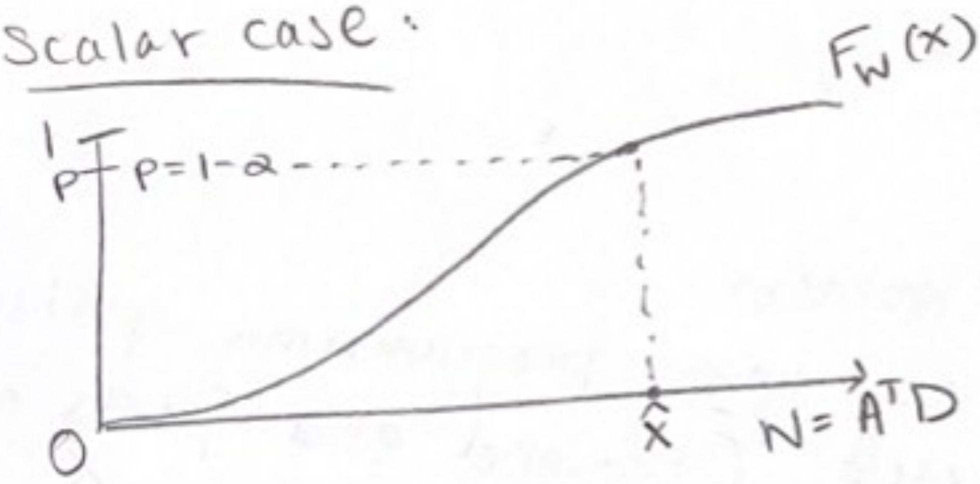
Joint-Chance constrained model.

- Manufacturer wants to satisfy the demand with some probability level $1-\alpha$ for $\alpha \in (0,1)$. $\longrightarrow$ maintaining reputation as a manufacturer.

$$\min_{x \geq 0} c^T x \quad \text{s.t.} \quad \underbrace{\Pr\{A^T D \leq x\} \geq p}_{\text{joint chance constraint}}$$

has random vector $W = A^T D$. can be rewritten $F_W(x) \geq p$, where $F_W(\cdot)$ is multidemensional distribution function of W. $\longrightarrow$ new kind of constraint! Think of W as parts required by demand.

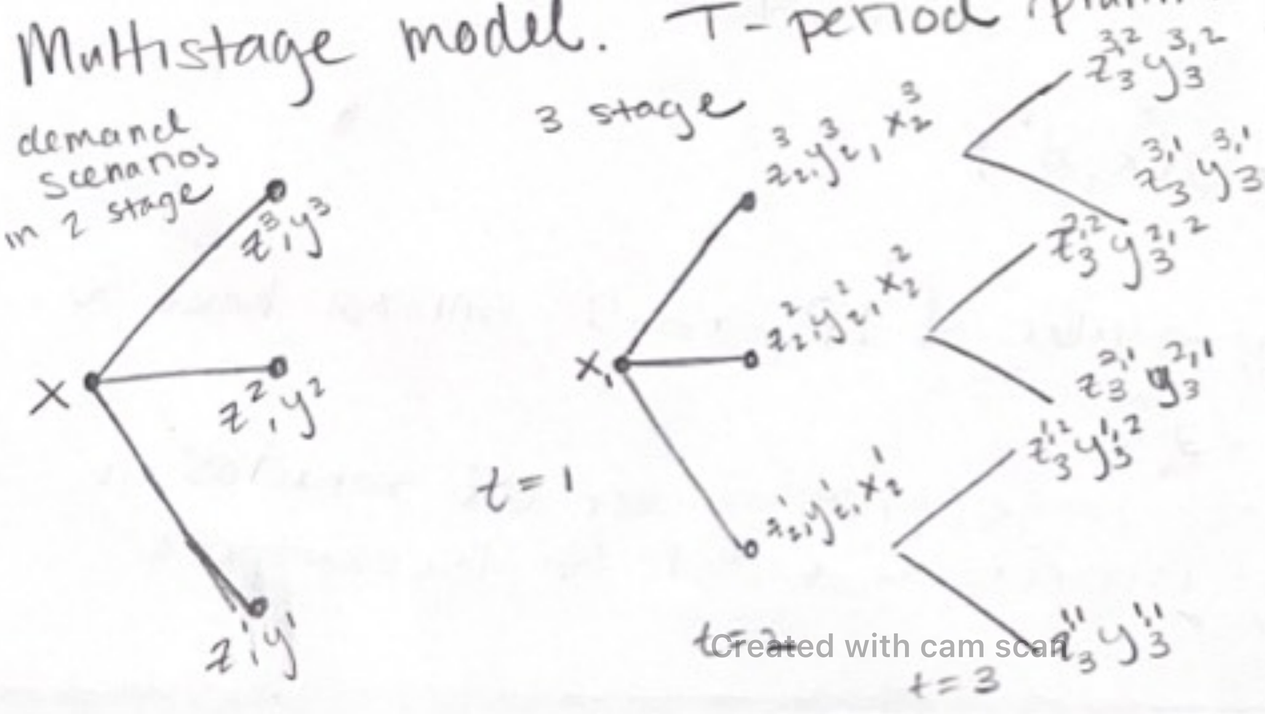
Scalar case:



quantile calculation

When x is high dimensional the calculation is very difficult.

Multistage model. T-period planning horizon.



- demand scenarios for $t=3$ are conditional on demand scenarios for $t=2$.
- number of scenarios grows exponentially with time horizon (number of stages).

Our decisions are functions of the data so we must work w/ functional spaces (group of functions w/ some properties).

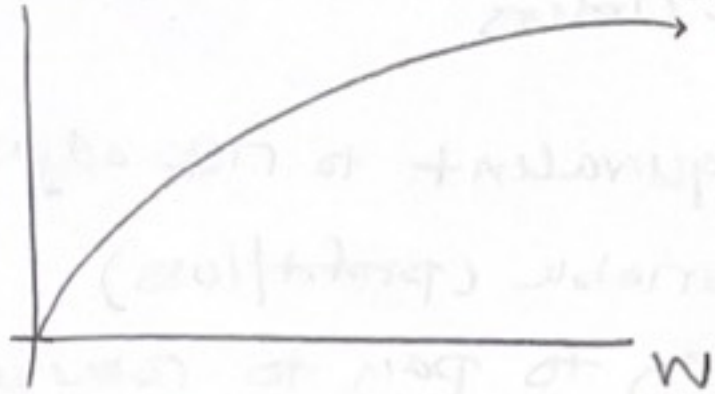
eg./ Hilbert spaces → need to study this rigorously.

Portfolio Selection –

- Expected Value model will always select a single asset to invest all capital in. This only makes sense when the Law of Large numbers applies, which is not realistic.
- makes sense in repeated manufacturing case or large insurance companies which spread risk across many policies.

utility – function converts results (economic value) to usefullness to the decision maker

↳ new proof of existance of utility function in 2024 Risk book (chapter 1). Simpler than Von Neumann's original proof.



for example

$$U(W) = \begin{cases} (W-a) & \text{if } W \geq a \\ (1+s)(W-a) & \text{if } W \leq a \end{cases}$$

w/ $s > 0,\ a > 0$

specifying utility function is difficult modeling challenge.

How many good medical treatments are worth one death?

...n nature of problem!

Mean-Risk Approach - measure result expected value and variability. 2 objectives:

- mean $E[W_1]$ maximize
- risk $Var[W_1]$ minimize $E[(W_1-\mu)^2]$

  $lsd[W_1]$ $\left(E[(\mu - W_1)_+^2]\right)^{1/2}$

- Avoids creating a utility function for decision maker.
- Weighting mean vs risk gives range of reasonable policies.

$E[W_1] = \langle P, W \rangle$ linear operator in the probability measure.

$lsd[W_1] = \left(\langle P, (\langle P, \mu\rangle - W)_+^2 \rangle\right)^{1/2}$ nonlinear operator w.r.t. probability measure.

- source of many research problems → how to deal with models w/ these risk functions

Chance constraints and Stochastic Orders

risk measure — synthetic equivalent to risk adjusted \$ fairvalue of random variable (profit/loss)
How much are you willing to pay to remove this risk? (losses)

Stochastic (Sub) Gradient Methods

Neural Networks

feature X $\rightsquigarrow$ labels Y

$X \to \boxed{NN\ (\beta)} \to$

$\Phi(X;\beta)$

some non differentiable non convex function

Training problem

$$\min_{\beta} E\left[\ell\left(\Phi(X;\beta)-Y\right)\right]$$

done w/ stochastic subgrad methods

$Z=(X,Y)$

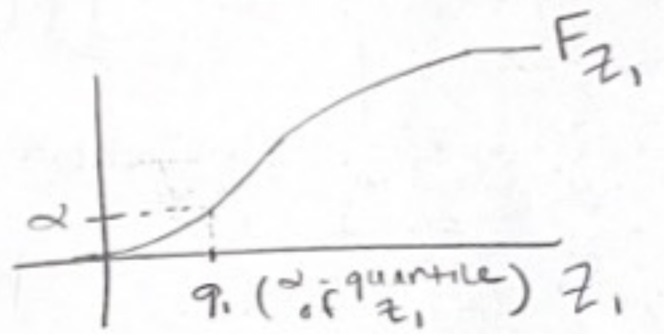
Optimization models w/ Probabilistic Constraints

$$g_j(x, z) \le 0, \; j \in J = \{1, \dots, m\}$$

- not useful to ensure constraint is satisfied in worst possible realization of z.

- Examples: finance – Value at Risk, Engineering – water reservoirs will not overflow, Business – meet (unknown) demands of customers

Individual constraint $P\{\underbrace{g_1(x, z) \le 0}_{Y_1}\} \ge 1 - \alpha$



- can rewrite as $g_1(x) \le z_1$

$$1 - F_{Z_1}(g_1(x)) \ge 1 - \alpha$$

$$F_{Z_1}(g_1(x)) \le \alpha$$

Similiarly $F_{Z_2}(g_2(x)) \le \alpha$

$\Rightarrow$ deterministic inequality

$g_1(x) \le q_1$

$g_2(x) \le q_2$

since F_{Z_1} is monotonic increasing.

"I satisfy demand of each individual customer w/ high probability but possible that I have low probability of satisfying all customer demands"

Joint Constraint

$$P\left\{\begin{array}{l} g_1(x) \le z_1 \\ g_2(x) \le z_2 \end{array}\right\} \ge 1 - \alpha$$

$$1 - F_Z(g(x)) \ge 1 - \alpha$$

$$F_Z(g(x)) \le \alpha$$

never convex as F_Z is bounded by $[0, 1]$ and is not constant.

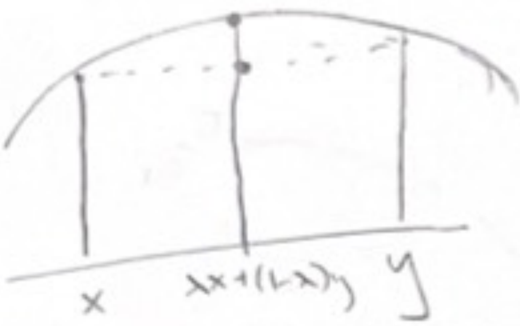
prove this!

Joint constraints consider "system state".

- GWU student/prof made early career w/ Supply chain problems with probabilistic constraints.

To deal w/ nonconvex constraint we have generalized concavity theorem.

concave $f(\lambda x + (1-\lambda)y) \geq \lambda f(x) + (1-\lambda) f(y)$



$$[f(\lambda x + (1-\lambda)y)]^{\alpha} \geq \lambda [f(x)]^{\alpha} + (1-\lambda)[f(y)]^{\alpha}$$

when $\alpha = 0$ we pass to limit and get

$$\ln[f(\lambda x + (1-\lambda)y)] \geq \lambda \ln[f(x)] + (1-\lambda) \ln[f(y)]$$

- if f is α-concave, then it is β-concave $\forall\ \beta \leq \alpha$.

proof from def:

Ex./
normal distribution

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)$$

$$\ln[f(x)] = -\ln(\sigma\sqrt{2\pi}) - \frac{(x-\mu)^2}{2\sigma^2}$$

concave quadratic function so f is log concave.

$$f(\lambda x + (1-\lambda)y) \geq \left[\lambda [f(x)]^{\alpha} + (1-\lambda)[f(y)]^{\alpha}\right]^{1-\alpha}$$

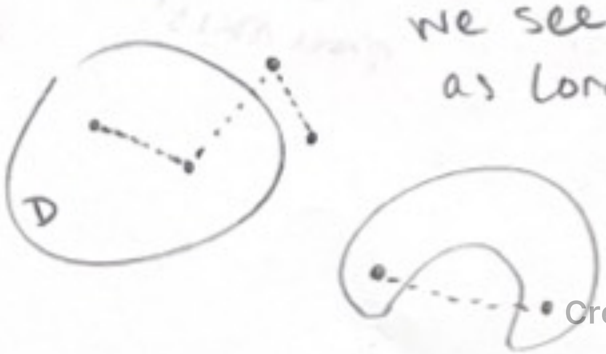
$\alpha \to -\infty$ $\quad f(\lambda x + (1-\lambda)y) \geq \min\{f(x), f(y)\}$

Ex./ Uniform distribution $D \subset \mathbb{R}^s$ convex. $f(x) = 0$ if $x \notin D$.

$f(x) = \frac{1}{V(D)}$ if $x \in D$.

$V(D)$ is volume of D.

we see f is quasi concave, as long as D is convex.



Easier to check density is α-concave then transform to get γ-concavity of probability measure.

$$\alpha = \frac{\gamma}{1 - m\gamma}$$

$$\Rightarrow \quad \alpha - m\alpha\gamma = \gamma$$

$$\Rightarrow \quad \gamma = \frac{\alpha}{1 + m\alpha}$$

$$F_Z(u) = \mathbb{P}\{Z \le u\} \quad . \; u, v$$

$$F_Z(\lambda u + (1-\lambda)v) = \mathbb{P}\{Z \le \lambda u + (1-\lambda)v\}$$

$$A = (-\infty, u], \quad B = (-\infty, v]$$

$$\lambda A + (1-\lambda)B = (-\infty, \lambda u + (1-\lambda)v]$$

$$F_Z(\lambda u + (1-\lambda)v) = \mathbb{P}\{Z \in \lambda A + (1-\lambda)B\}$$

if $\mathbb{P}$ is α-concave $\ge m_\alpha(\mathbb{P}\{A\}, \mathbb{P}\{B\}, \lambda)$

$\alpha = 0$

$$\ln F_Z(\lambda u + (1+\lambda)v) \ge \lambda \ln F_Z(u) + (1-\lambda)\ln F_Z(v)$$

Examples uniform distribution

We established density is α-concave with $\alpha = -\infty$

$$\gamma = \lim_{\alpha \to -\infty} \frac{\alpha}{1 + m\alpha} = \frac{1}{m}, \text{ where } m \text{ is dimension of } D.$$

So distribution is $\frac{1}{m}$-concave.

log normal distribution

Z has multivariate log-normal distribution if vector $Y = (\ln Z_1, \dots, \ln Z_m)^T$ has multivariate normal distribution.

- commonly used in applications

$$F_Z(x) = P\{Z_1 \leq x_1, \dots, Z_m \leq z_m\}$$

$$= P\{\underbrace{x_1 - e^{Y_1} \geq 0}_{g_1(x_1, Y_1)}, \dots, \underbrace{x_m - e^{Y_m} \geq 0}_{g_m(x_m, Y_m)}\} \quad (1)$$

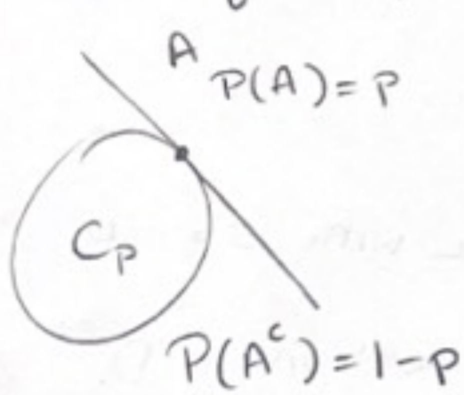
- linear in x
- concave in Y

so concave in (x, Y) $\Rightarrow$ quasiconcave in (x, Y)

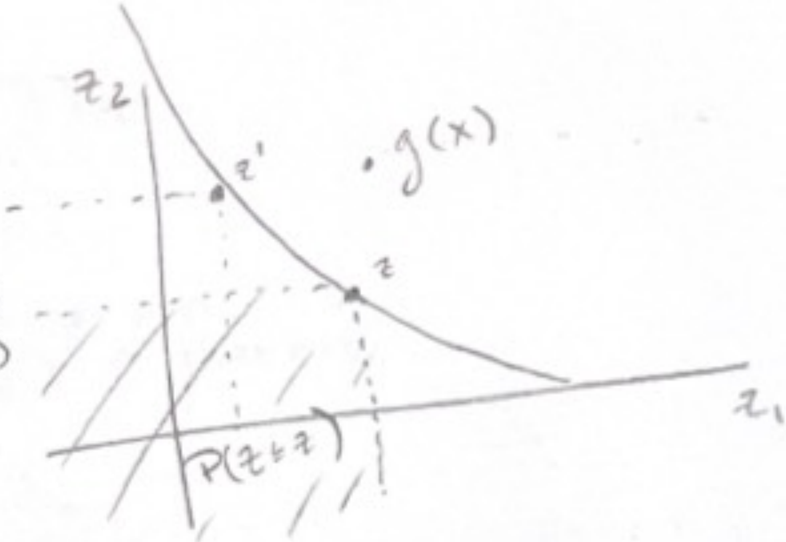
Y has normal distribution which is log concave

so (1) is logconcave, and $\{x : F_Z(x) \geq p\}$ is convex.

Floating Body



Separable Prob Constraints



$$F_Z(z) = \mathbb{P}\{Z \le z\}$$

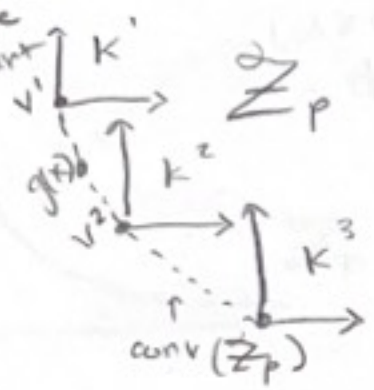
$$\mathcal{Z}_p = \{z \in \mathbb{R}^m : F_Z(z) = \mathbb{P}\{Z \le z\} \ge p\}$$

for every $p \in (0,1)$, $\mathcal{Z}_p$ is nonempty and closed

A point v is p-efficient if $F_Z(v) \ge p$ and there is no $z \le v$, $v \ne z$ such that $F_Z(z) \ge p$.

Bounded below by quantiles of marginal distribution.

Discrete distribution w/ finite p-efficient points



You will recover p efficient points from extreme points of $\mathrm{conv}(\mathcal{Z}_p)$.

$\mathcal{Z}_p$ is convex if Z has α-concave distribution function.

How to solve

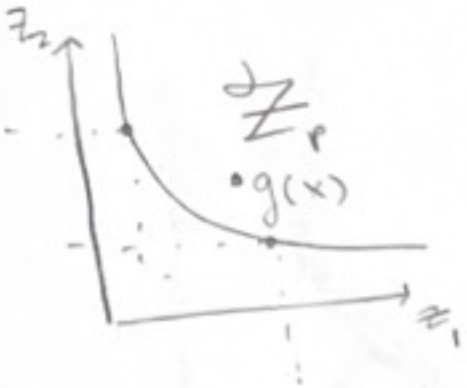
Approx $L_D(\cdot)$ with tangent ~~lines~~ planes, the slope of these planes is $g(x)$ or z.

$g(x)$ – nonlinear opt.
z – subgradient method

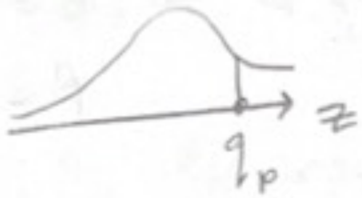
$L_D(u)$ - concave

Recall previous problem

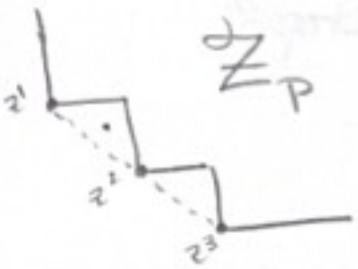
$$\min f(x) \quad \text{ST} \quad P[g(x) \geq z] \geq p$$
$$x \in D$$



In one dimension
it is just a quantile

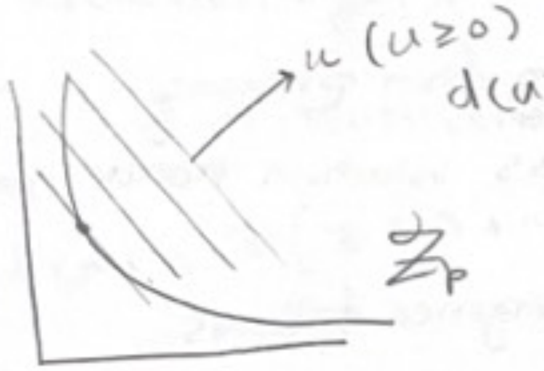


Possible situation → we take convex hull of Z_p



$z^1, z^2, z^3 \in Z_p$ (not necessarily convex)

- In practice we can recover a feasible solution by find nearest feasible point.



$$d(u) = \inf\{\langle u, z\rangle \mid z \in Z_p\}$$

* we'll recover a p-efficient point.

for every fixed $x \in D$, $z \in Z_p$ the Lagrangian

$$L(x, z, u) = f(x) + \langle u, z - g(x)\rangle$$

is a linear function of u. Then $L_D(u)$ is the minimum of infinitely many linear functions

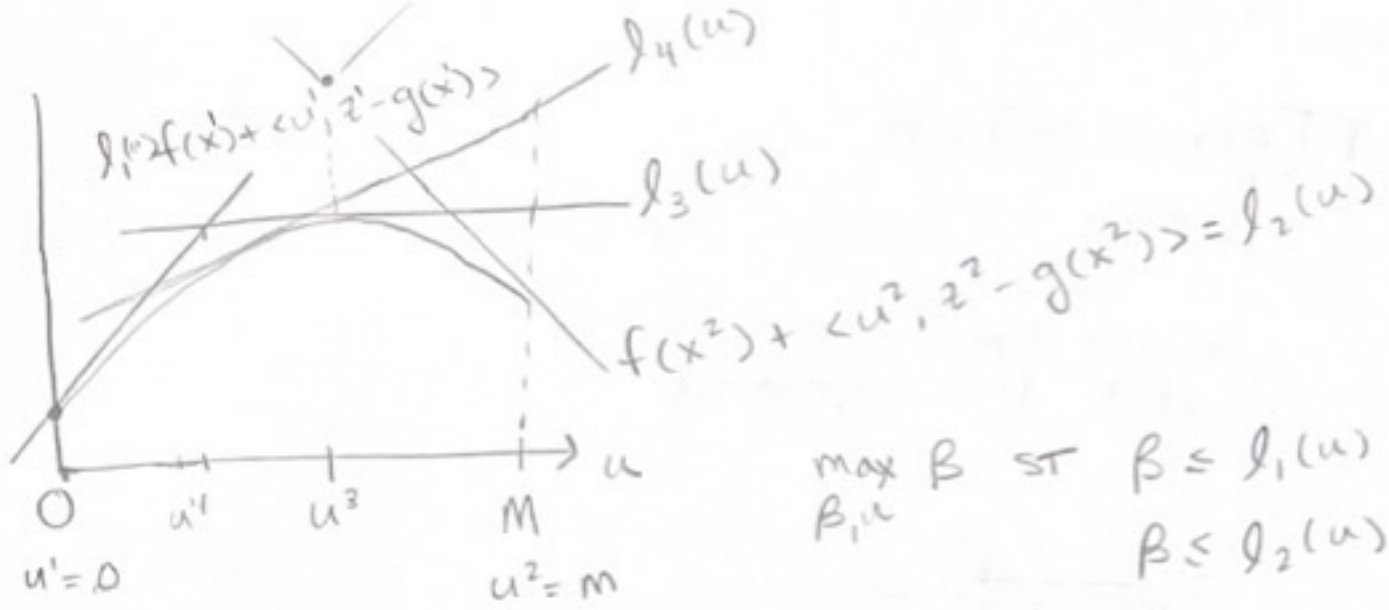
↳ concave function
(intersection of hypographs)

$$L_D(\lambda u^1 + (1-\lambda)u^2) = \min_{\alpha \in A} l_\alpha(\lambda u^1 + (1-\lambda)u^2)$$

$$= \min_{\alpha \in A} [\lambda l_\alpha(u^1) + (1-\lambda) l_\alpha(u^2)]$$ since l_α is linear wrt u.

$$\geq \lambda \underbrace{\min_{\alpha \in A} l_\alpha(u^1)}_{L_D(u^1)} + (1-\lambda) \underbrace{\min_{\alpha \in A} l_\alpha(u^2)}_{L_D(u^2)}$$

we can create a piecewise linear approximation by computing l_α for various values of u. Cutting plane method



$\max_{\beta, u} \beta$ ST $\beta \le l_1(u)$

$\beta \le l_2(u)$

gives u^3

$u^4 = \underset{\beta, u}{\operatorname{argmax}}\ \beta$ ST $\beta \le l_1(u)$

$\beta \le l_2(u)$ and so on. when to stop?

$\beta \le l_3(u)$

when $L_D(u^k) \le L_D^k(u^k) - \delta$ we stop, where

L_D^k is the binding linear inequality from previous maximization (usually l_{k-1}, unless solution nonunique)

$\|u^k - u^j\| \ge \frac{\delta}{B}$ for all $j < k$. Then proof of convergence follows (Thm 4.1)

$$L_D(u) = \underbrace{\min_{x \in D} \{f(x) - \langle u, g(x) \rangle\}}_{d(u)} + \underbrace{\min_{z \in Z_p} \{\langle u, z \rangle\}}_{h(u)}$$

primal recovery

$\max_{x,u} \beta$ ST $\beta \le l_1(u)$ $\alpha_1 \ge 0$ $\quad \sum_{i=1}^{j} \alpha_i = 1 \quad J = \{j : l_j \text{ is binding}\}$

we take convex combination to get (x, z)

$\beta \le l_2(u)$ $\alpha_2 \ge 0$

$\beta \le l_3(u)$ $\alpha_3 \ge 0$

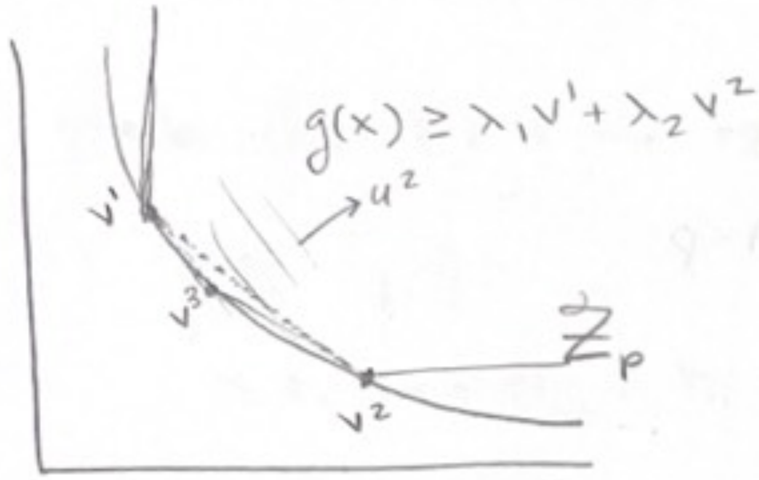
$\vdots \qquad \vdots$

$\beta \le l_j(u)$ $\alpha_j \ge 0$

$u \ge 0$

Primal-Dual method

Here we treat $d(u)$ directly since it is a deterministic NLP and approx $h(u)$.



Then next master problem has constraint $g(x) \geq \lambda_1 v^1 + \lambda_2 v^2 + \lambda_3 v^3$ which is a better approximation of Z_p.

Generating p-efficient points

- $Z \sim$ multivariate normal (Ganz U of Washington)

$f(a) = P[Z \in A + a]$ where A is a polyhedron

- mixed integer techniques

$\min f(x)$ ST $P[g(x, Z) \geq 0] \geq p$ nonseparable constraints
$x \in D$

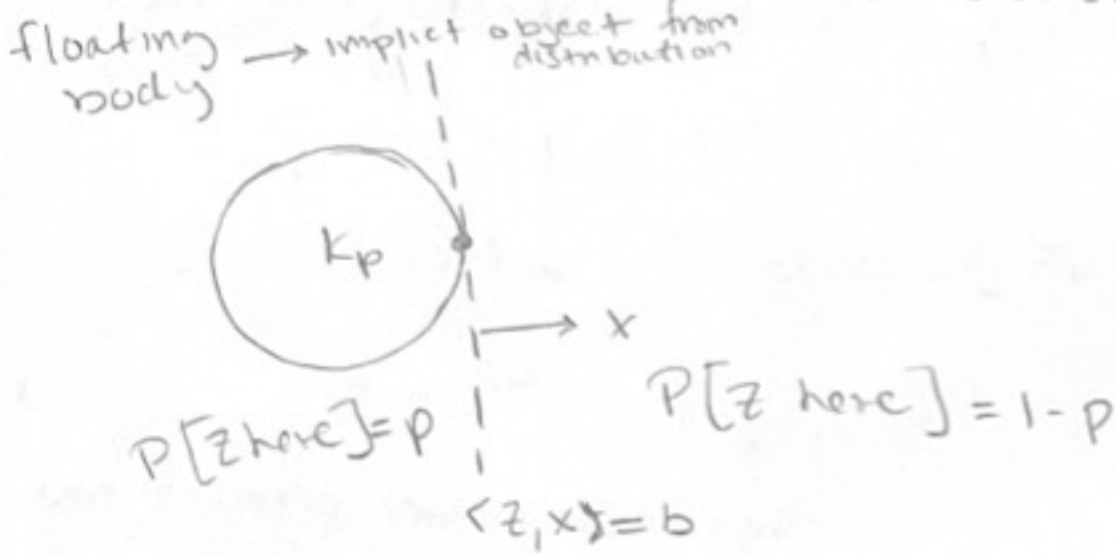
Assume Z has discrete distribution $P[Z = z_k] = p_k$ $k = 1, \dots, N$
$\sum_{k=1}^{N} p_k = 1$

$\longrightarrow$ discrete means nonconvex!

$P[\langle z, x \rangle \leq b] \geq p$

z - random

x - decision



I Two - Stage Problems

$$\min_{x \in \mathbb{R}^n} \langle c, x \rangle + E[Q(x, \xi)]$$
$$\text{s.t. } Ax = b,\ x \geq 0$$

where $Q(x, \xi)$ is the optimal value of 2^{nd} stage problem:

$$\min_{y \in \mathbb{R}^n} \langle q, y \rangle$$
$$\text{s.t. } Tx + Wy = h,\ y \geq 0$$

where $\xi = (q, h, T, W)$ are the data of 2^{nd} stage problem
$\Xi \subset \mathbb{R}^d$ denotes the support of the probability distribution of ξ.

if for some x and $\xi \in \Xi$ the second stage problem is infeasible
then by definition $Q(x, \xi) = +\infty$.

To get dual of 2^{nd} stage problem:

$$\min\ q^T y \quad \text{s.t. } \underset{(\text{multipliers})}{\pi \rightarrow} Wy = h - Tx,\quad y \geq 0$$

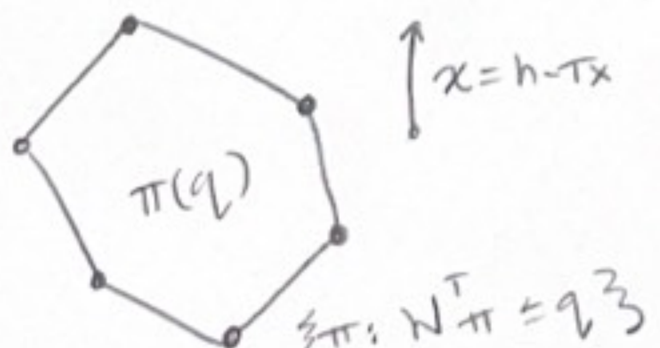
$$(DP)\quad \underbrace{\max_{\pi} \pi^T (h - Tx) \text{ s.t. } W^T \pi \leq q}_{\text{convex function (maximum of linear functions)}}$$

Define $\Pi(q) \doteq \{\pi : W^T \pi \leq q\}$

$$\sigma_q(\mathcal{X}) \doteq \sup_{\pi \in \Pi(q)} \{\pi^T \mathcal{X}\} \rightarrow \text{polyhedral function}$$

maximized at vertices of the set or its unbounded.

clearly $Q(x, \xi) = \sigma_q(h - Tx)$

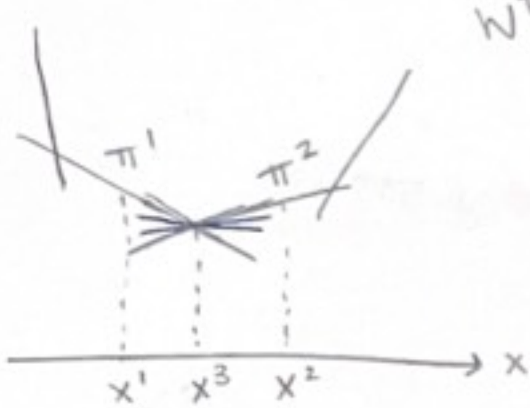


7-random

(DP) $\max \pi^T(h - Tx) = \pi^T h - \langle T^T\pi, x\rangle$

s.t. $\pi \in \Pi(q)$

$W^T\pi \leq q$



$\partial Q(x^1, \xi) = \{-T^T\pi^1\}$

$\partial Q(x^2, \xi) = \{-T^T\pi^2\}$

$\partial Q(x^3, \xi) = \{$ infinite collection of ______ $\}$ ← convex set!

$$Q(x, \xi) \geq Q(x^0, \xi)$$
$$= -\langle T^T \underset{\pi^1}{\pi}, x - x^0\rangle$$

any solution of dual problem

x^0, π^0 is an optimizer of $\pi^T(h - Tx^0)$, $\pi \in \Pi(q)$

At any other x, $Q(x, \xi) = \max\limits_{\pi\in\Pi(q)} \pi^T(h - Tx) \geq (\pi^0)^T(h - Tx)$

$$= (\pi^0)^T(h - Tx^0) - (\pi^0)^T(Tx - Tx^0)$$
$$= Q(x^0, \xi) - \underbrace{(\pi^0)^T T}_{\text{subgradient}}(x - x^0)$$

Solve the dual problem by any method, which obtains the subgradient of dual function. which can be use to optimize the dual function with a subgradent method.

Discrete Distributions $\quad \xi_s = (q_s, h_s, T_s, W_s)$

$\phi(x) \doteq E[Q(x,\xi)]$ Expected value function $\quad$ with probabilities p_s

Suppose finite scenarios $1, \dots, S$

$$E[Q(x,\xi)] = \sum_{s=1}^{S} p_s Q(x, \xi_s)$$

Assume $Q(x,\xi_s) = +\infty$ if x is not feasible.

Then 2 stage problem is equivalent to LP problem
(minima of minima is overall minima)

$$\min_{x, y_1, \dots, y_S} \; c^T x + \sum_{s=1}^{S} p_s q_s^T y_s$$

$$\text{s.t. } T_s x + W_s y_s = h_s, \quad s = 1, \dots, S$$

$$Ax = b,$$

$$x \geq 0, \; y_s \geq 0, \quad s = 1, \dots, S$$

$$\begin{array}{c}
\begin{array}{ccccc} x & y_1 & y_2 & \dots & y_S \end{array} \\
\left|\begin{array}{ccccc}
A & & & & \\
T_1 & W_1 & & & \\
T_2 & & W_2 & & \\
\vdots & & & \ddots & \\
T_S & & & & W_S
\end{array}\right.
\end{array}
\qquad
\begin{array}{c} b \\ h_1 \\ h_2 \\ \vdots \\ h_S \end{array}$$

Constraint dimension
(decision variables in 1^{st} stage + decision variables in 2^{nd} stage) $\times$ (# scenarios)

$(S+1)(S)$

when S grows exponentially with # of R.Vs in problem this can be huge dimension.

$\downarrow$

need another method

Suppose $\phi(\cdot)$ has a finite value in at least one point $\bar{x} \in \mathbb{R}^n$
Then $\phi(\cdot)$ is polyhedral and for any $x_0 \in \text{dom}\,\phi$

$$\partial \phi(x_0) = \sum_{s=1}^{S} p_s \partial Q(x_0, \xi_s)$$

$$\phi(x) = \sum_{s=1}^{S} p_s Q(x, \xi_s)$$

x°: find $g^\circ \in \partial \phi(x^\circ)$

1 Calculate $g_s \in \partial Q(x^\circ, \xi_s)$ $\quad s = 1, \dots, S$

$$g_s = -T^T \pi_s \qquad \pi_s = \underset{\pi \in \Pi(q_s)}{\text{argmax}}\ \pi^T (h_s - T_s x^\circ)$$

2 Set $g^\circ = \sum_{s=1}^{S} p_s g_s$

Done

we have subgrad ineq.

$$Q(x, \xi_s) \geq Q(x^\circ, \xi_s) + \langle g_s, x - x^\circ \rangle$$

$\leftarrow$ multiply by p_s and for all $s = 1, \dots, S$

$$\underbrace{\sum_{s=1}^{S} p_s Q(x, \xi_s)}_{\phi(x)} \geq \underbrace{\sum_{s=1}^{S} p_s Q(x^\circ, \xi_s)}_{\phi(x^\circ)} + \underbrace{\langle \sum_{s=1}^{S} g_s, x - x^\circ \rangle}_{\langle g_s, x - x^\circ \rangle}$$

2^{nd} stage problem

$$\min q^T y$$
$$\text{s.t. } Wy = h - Tx$$
$$y \geq 0$$

- if W is deterministic, this is a <u>fixed recourse problem</u> (RHS is random)

  Ex. | T - random accessibility of facilities
  x - location of facilities
  h - random demand
  W - recourse

- if system $Wy = \chi$ and $y \geq 0$ has a solution for every χ then recourse is <u>complete</u>
- if every $x \in \{x : Ax = b, x \geq 0\}$ the feasible set of second stage problem is nonempty a.s., the recourse is <u>relatively complete</u>.

(DP) $\max_{\pi} \pi^T (h - Tx)$

$\quad$ s.t. $\pi \in \Pi(q)$ — $\{\pi : W^T \pi \leq q\}$

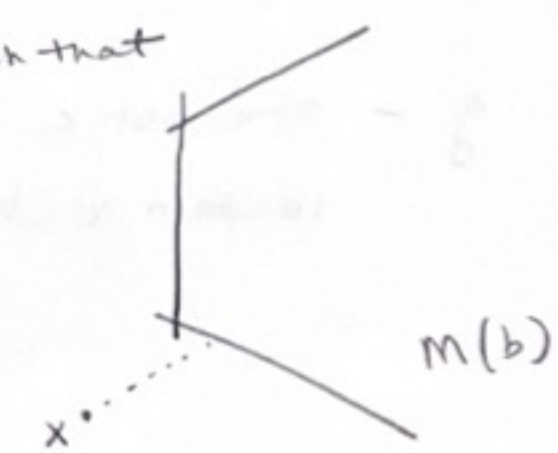
we need to check if expected value of this problem exists!

<u>Hoffman's Thm</u> Consider $M(b) \doteq \{x \in \mathbb{R}^n : Ax \leq b\}$

$\exists$ a positive constant K, depending on A, such that for any $x \in \mathbb{R}^n$ and $b \in \text{dom } M$

$$\text{dist}(x, M(b)) \leq K \|(Ax - b)_+\|$$

" distance to set is bounded by violation of constraints "



According to Hoffman's Thm, $\exists$ K, depending on W, such that if for some q_0 the set $\Pi(q_0)$ is nonempty, then for every q

$$\Pi(q) \subset \Pi(q_0) + K\|q-q_0\|B \quad \text{with} \quad B = \{\pi : \|\pi\| \le 1\}$$

<u>Thm</u> Suppose the recourse (W) is fixed, $E[\|q\|\cdot\|h\|] < +\infty$

$E[\|q\|\cdot\|T\|] < +\infty$

Then for a point $x \in \mathbb{R}^n$, $E[Q(x,\xi)_+]$ is finite if and only if $h - Tx \in \text{pos } W$ holds with probability 1.

$$\phi(x) = E[Q(x,\xi)] = \int Q(x,\xi)P(d\xi)$$

Fix x^0: $\quad g^\xi \in \partial_x Q(x^0,\xi)$ for all ξ, $g^\xi = -T^\xi \pi^\xi$

Calculate $g = \int g^\xi P(d\xi)$

$Q(x,\xi) \ge Q(x^0,\xi) + \langle g^\xi, x-x^0\rangle$ integrate both sides

$$\underbrace{\int Q(x,\xi)P(d\xi)}_{\phi(x)} \ge \underbrace{\int Q(x^0,\xi)P(d\xi)}_{\phi(x^0)} + \underbrace{\langle \int g^\xi P(d\xi), x-x^0\rangle}_{\langle g, x-x^0\rangle}$$

g - stochastic subgradient

random vector whose expectation is the real subgradient.

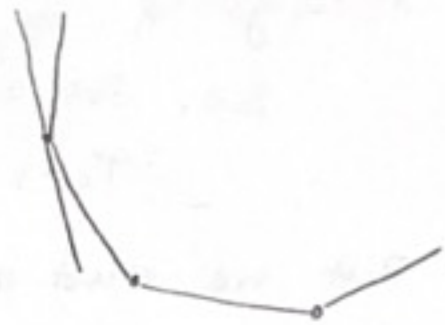
<u>Strassen Thm</u> :: $g \in \partial \phi(x^\circ) \Leftrightarrow g = \int g^{\xi} P(d\xi)$

disintegration formula $\rightarrow$ g must be expressed as the integral of random vectors

<u>Optimality Conditions</u>

in discrete case, Let $\mathcal{X} = \{ x \in \mathbb{R}^n : Ax = b, x \geq 0 \}$

we need necessary + sufficient conditions to minimize $c^T x + \phi(x)$ over $x \in \mathcal{X}$.



$\bar{x}$ is optimal, then we can write problem as

$$\min_{x \in \mathcal{X}} c^T x + g^T x$$

where $g \in \partial \phi(\bar{x})$

So conditions are $0 \in c + \partial \phi(\bar{x}) + N_{\mathcal{X}}(\bar{x})$

$g = -\sum p_s T_s^T \pi_s$. Then conditions are:

$$\sum_{s=1}^{S} p_s T_s^T \pi_s + A^T \mu \preceq c,$$

$$\bar{x}^T \left(c - \sum_{s=1}^{S} p_s T_s^T \pi_s - A^T \mu \right) = 0$$

We see this is equivalent to conditions from Large LP equivalent form.

Nonanticipativity: Scenario formulation

one x_s for each scenario

$$\min_{x_1, y_1, \dots, y_s} c^T x_s + \sum_{s=1}^{S} p_s q_s^T y_s$$

$$\text{s.t. } T_s x_s + W_s y_s = h_s \quad s=1,\dots,S$$

$$Ax_s = b$$

$$x_s \geq 0,\ y_s \geq 0 \quad s=1,\dots,S$$

is separable to S problems

$$\min_{x_s \geq 0, y_s \geq 0} c^T x_s + q_s^T y_s$$

$$\text{s.t. } Ax_s = b$$

$$T_s x_s + W_s y_s = h_s$$

But we must introduce an additional linear constraint (nonanticipativity)

$$(x_1, \dots, x_S) \in \mathcal{L} = \{ x = (x_1, \dots, x_S) \in \mathbb{R}^{nS} : x_1 = \dots = x_S \}$$

Each should equal to the average of all:

$$x_s = \sum_{i=1}^{S} p_i x_i \ , \quad s=1,\dots,S$$

which is equivalent.

we see $\mathcal{L}$ is a linear subspace of $\mathcal{X} = \underbrace{\mathbb{R}^n \times \dots \times \mathbb{R}^n}_{S \text{ times}}$

we equip with scalar product:

$$\langle x, y \rangle := \sum_{i=1}^{S} p_i x_i^T y_i$$

and define linear operator

$$x = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{pmatrix} \quad \mathbb{P}x = \begin{pmatrix} 3 \\ 3 \\ 3 \\ 3 \\ 3 \end{pmatrix}$$

$$\mathbb{P}x \doteq \begin{pmatrix} \sum_{i=1}^{S} p_i x_i \\ \vdots \\ \sum_{i=1}^{S} p_i x_i \end{pmatrix} \quad \mathbb{P}: \mathcal{X} \to \mathcal{X}$$

and we can write constraint as $x = \mathbb{P}x$.

$\mathbb{P}$ is the orthogonal projection operator of $(\mathcal{X}, \langle\cdot,\cdot\rangle)$ onto the subspace $\mathcal{L}$. Clearly $\mathbb{P}(\mathbb{P}(x)) = \mathbb{P}(x)$ and

$$\langle \mathbb{P}x, y\rangle = \left(\sum_{i=1}^{S} p_i x_i\right)^T \left(\sum_{s=1}^{S} p_s y_s\right) = \langle x, \mathbb{P}y\rangle$$

So $\mathbb{P}$ is self adjoint, and $\mathbb{P}$ is a projection.

$\mathcal{L}$ is called the nonanticipativity subspace of $\mathcal{X}$.

With this constraint:

$$\min_{x, y_1, \dots, y_S} \quad c^T x_s + \sum_{s=1}^{S} p_s q_s^T y$$

$$\text{s.t.} \quad T_s x_s + W_s y_s = h \qquad s = 1, \dots, S$$

$$A x_s = b$$

$$x_s = \sum_{i=1}^{S} p_i x_i \qquad s = 1, \dots, S$$

$$x_s \geq 0, \; y_s \geq 0 \qquad s = 1, \dots, S$$

$$L(x, \lambda) = \sum_{s=1}^{S} p_s (c^T x_s + q_s^T y_s) + \sum_{s=1}^{S} p_s \lambda_s^T \left(x_s - \sum_{i=1}^{S} p_i x_i\right)$$

$$PPx = Px$$

$$\langle Px, y\rangle = \langle x, Py\rangle$$

$$(I-P)(I-P)x = x - Px - Px + PPx$$
$$= x - Px = (I-P)x$$

$$\langle (I-P)x, y\rangle = \langle x, y\rangle - \langle Px, y\rangle$$
$$= \langle x, y\rangle - \langle x, Py\rangle$$
$$= \langle x, (I-P)y\rangle$$

So if P is a projection then $I-P$ is a projection