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:

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

$$\begin{aligned} \text{s.t. } \; & T_s x_s + W_s y_s = h \qquad s = 1, \dots, S \\ & A x_s = b \\ & x_s = \sum_{i=1}^{S} p_i x_i \qquad s = 1, \dots, S \\ & x_s \geq 0, \; y_s \geq 0 \qquad s = 1, \dots, S \end{aligned}$$

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

$PPx = Px$

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

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

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

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

$$\sum_{s=1}^{S} p_s \lambda_s^T \left(x_s - \sum_{i=1}^{S} p_i x_i\right) = \langle \lambda, (I-P)x \rangle = \langle (I-P)\lambda, x \rangle$$

Then

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

ty)

Two Stage Problems in finite space

$\Omega = \{\omega^1, \dots, \omega^S\}$, $\mathbb{P}(\omega^s) = p_s$, $s = 1, \dots, S$

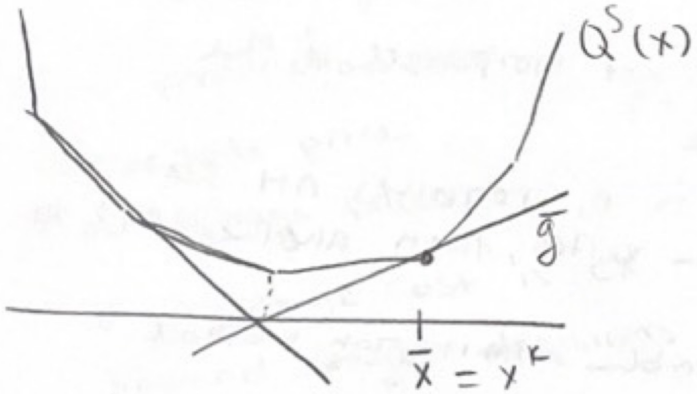
Denote $T^s = T(\omega^s)$, $W^s = W(\omega^s)$, $h^s = h(\omega^s)$, $y^s = y(\omega^s)$, $q^s = q(\omega^s)$

Linear Problem

$$\min_{x \in X} \left\{ f(x) = \langle c, x \rangle + \sum_{s=1}^{S} p_s Q^s(x) \right\}$$

$$Q^s(x) = \inf_{y \in \mathbb{R}_+^{n_2}} \{ (q^s)^T y \mid W^s y = h^s - T^s x$$

$$X = \{ x \in \mathbb{R}^{n_1} : Ax = b, x \geq 0 \}$$



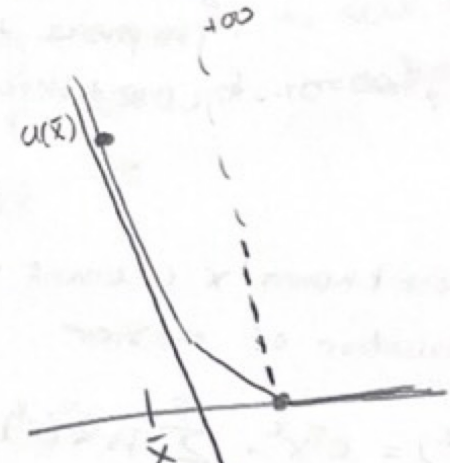
$$Q^s(x) \geq Q^s(\bar{x}) + \langle \bar{g}, x - \bar{x} \rangle$$

$$\bar{g} = -(T^s)^T \pi$$

feasibility cut $\|z\|$

$$u(x) = \sum_{i=1}^{m} z_i \quad \text{s.t.} \dots$$

$u(\bar{x}) > 0$



$$0 \geq u(x) \geq u(\bar{x}) + \langle r, x - \bar{x} \rangle$$

$$0 \geq u(\bar{x}) + \langle r, x - \bar{x} \rangle$$

$\|\eta\|^* = \max |\eta_i|$ where $\|\cdot\|^*$ is dual of $\|\cdot\|$

2nd Stage problem $\min q^T y$ s.t. $Wy = h - Tx$
$y \geq 0$

$\downarrow$

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

To avoid dealing with feasibility cuts, use penalty method.

$Q^s(x) \geq Q^s(\bar{x}) + \langle \bar{g}^s, x - \bar{x} \rangle$ $\times p_s$ add sum over S

$$\sum_s p_s Q^s(x) \geq \sum_s p_s Q^s(\bar{x}) + \sum_s p_s \langle \bar{g}^s, x - \bar{x} \rangle$$

Regularized Decomposition

- Cutting plane methods do not work well if the dimension of X is large
  - $\hookrightarrow$ if dimension is n, requires n+1 cuts to get a stable vertex, then another n+1 to improve it.
- Adaptation of nondifferentiable optimization method.

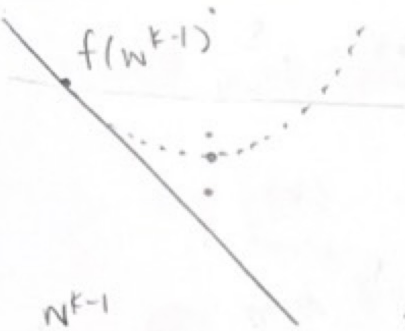
w^{k-1} - best known x (actual value, not approx) so far

x^k - solution of master

$$f(x^k) = c^T x^k + \sum_{s=1}^{S} p_s Q^s(x^k)$$

$\gamma \in (0,1)$

$$f(x^k) \leq (1-\gamma) f(w^{k-1}) + \gamma f^{k-1}(x^k)$$



$$\min \; c^T x + \sum p_s v^s + \frac{\rho}{2}\|x - w^{k-1}\|^2$$

subject to cuts $\leq v^s$

2 possibilities: · either improved over $f(w^{k-1})$ and move w^k to this point

· does not improve over $f(w^{k-1})$ and keep $w^k = w^{k-1}$ (null step)

Bundle methods → proof of convergence: (idea)

Newsvendor Problem

- x - first stage variable, production quantity to satisfy demand D - random, unknown.
- $c > 0$ production cost
- $s > 0$ sale price
- $r \geq 0$ salvage value

$r < c < q$

First stage cost is cx

Second stage decisions y - amount to sell for price q

r - amount to salvage for price r

cost is $-qy - rz$

$D \geq x$ sell all x

$D < x$ sell D and salvage $x - D$

$$\min_{x \geq 0} \; cx + E[Q(x, D)]$$

$$Q(x, D) = \min_{y \geq 0, z \geq 0} -qy - rz \quad \text{s.t. } y \leq D, \; y + z \leq x$$

Suppose we have finite scenarios for demand $s = 1, \dots, S$

$$E[Q(x,D)] = \sum_s p_s Q(x, D^s)$$

optimal value of dual problem is

$\min -qy - rz$

$y \le D \leftarrow \lambda_1$

multipliers

$y + z \le x \leftarrow \lambda_2 \ge 0$

$y, z \ge 0$

$\lambda_1 D + \lambda_2 x$

$$\frac{\partial}{\partial x}(\lambda_1 D + \lambda_2 x) = \lambda_2$$

$$Q(x, D) \ge Q(\bar{x}, D) - \langle \lambda_2, x - \bar{x} \rangle$$

$Tx + Wy = h$

$T = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$

$$Q(x, D^s) \ge Q(\bar{x}, D) - \lambda_2 x + \lambda_2 \bar{x}$$

$$g^k = -\sum_{s=1}^{S} p_s \lambda_s^k$$

$$Q(x, D^s) \ge Q(x^k, D^s) + \langle g^{k,s}, x - x^k \rangle \qquad g^{k,s} = -\lambda^{k,s}$$

$$E[Q(x,D)] \ge \sum_s p_s Q(x^k, D^s) + \langle \underbrace{-\sum_s p_s \lambda^{k,s}}_{g_s}, x - x^k \rangle$$

Slope of cut is $\left(-\sum_s p_s \lambda^{k,s}\right)$

α^k - collection of constants.

AMPL $\rightarrow$ online run environments or download to run locally

master problem

$\min \; cx + v$

$\langle g^j, x \rangle + \alpha_j \le v \quad j = 1, \dots, n$ cuts

$0 \le x \le u$

$-1000000 \le n$

Created with cam scan

Bundle method

① $\min f(x)$ → can't solve

② $\min f(x) + \frac{\rho}{2}\|x-w\|^2$ — moreau regularization

↳ solution x then substitute
w for x and repeat ②
(proximal point method)

For regularized cutting ~~point~~ plane
we replace f(x) with
cutting plane approx

Combination of proximal point
and cutting plane

3. $\min \langle c, x \rangle$ s.t. $P(g_i(x) \geq Y_i, \ i=1,2) \geq 0.9$

$a_i \leq x_i \leq b_i$ for $i = 1, \dots, n$

$c \in \mathbb{R}^n$, $g_i : \mathbb{R}^n \to \mathbb{R}$ concave $i = 1,2$

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \sim \text{lognormal}\left(\mu = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \Sigma = \begin{pmatrix} 0.75 & -0.1 \\ -0.1 & 0.5 \end{pmatrix} \right)$$

Let $g(x) = \begin{pmatrix} g_1(x) \\ g_2(x) \end{pmatrix}$, $g : \mathbb{R}^n \to \mathbb{R}^2$

Note $x_i \leq b_i \Leftrightarrow -x_i \geq -b_i$. $a, b \in \mathbb{R}^n$

$L(x, u, \lambda, \sigma) = \langle c, x \rangle + \langle u,$

Multistage Problems

Decision problem with form: decision(x_1) $\rightsquigarrow$ observation(ξ_2) $\rightsquigarrow$ decision (x_2) $\rightsquigarrow \dots \rightsquigarrow$ observation(ξ_T) $\rightsquigarrow$ decision (x_T)

For $t = 1, 2, \dots, T$ the sequence of data vectors $\xi_t \in \mathbb{R}^{d_t}$ is a stochastic process. Define $\xi_{[t]} \doteq (\xi_1, \dots, \xi_t)$ as the history of the process upto time t.

Nonanticipativity: decision x_t may only depend on $\xi_{[t]}$ but not future observations

T- stage stochastic Programming Problem:

$$\min_{x_1 \in \mathcal{X}_1} f_1(x_1) + E\left[\inf_{x_2 \in \mathcal{X}_2(x_1, \xi_2)} f_2(x_2, \xi_2) + E\left[\dots + E\left[\inf_{x_T \in \mathcal{X}_T(x_{T-1}, \xi_T)} f_T(x_T, \xi_T)\right]\right]\right]$$

where $x_t \in \mathbb{R}^{n_t}$, $f_t : \mathbb{R}^{n_t} \times \mathbb{R}^{d_t} \to \mathbb{R}$ are continuous functions

$\mathcal{X}_t : \mathbb{R}^{n_{t-1}} \times \mathbb{R}^{d_t} \rightrightarrows \mathbb{R}^{n_t}$ are measurable closed valued multifunctions

$f_1 : \mathbb{R}^{n_1} \to \mathbb{R}$, ξ_1, $\mathcal{X}_1 \subset \mathbb{R}^{n_1}$ are deterministic

- when we narrow the class of functions policies can take Then we can speed up computation of optimal policy significantly.

  Example → controling movement of underwater robots

- Dynamic Programming formulation of multistage problems (Expected) Cost-to-go functions $Q_t(x_{t-1}, \xi_{[t]})$ are convex and represent the (Expected) cost to operate the system from time t to T.

- Professor Lin — Stochastic models for data process (ARCH, GARCH) time series analysis

Generating Scenario Trees

- Sample distribution to get Ω_2
- for every node i in Ω_2 sample conditional distribution to get $c(i)$ and generate Ω_3
- The process grows tree size exponentially.

Ω - space of all scenarios (ξ_1, ξ_2, ξ_3)

$\mathcal{F}_T$ - collection of subsets of Ω

$\mathcal{F}_t$ - collection of subsets distinguishable at time t

$\mathcal{F}_1 = \{\phi, \Omega_T\}$ (we can not tell we are in/not in any of the possible scenarios)

$\mathcal{F}_2$ = all subsets of Ω_2

Sequence of sigma fields $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots \subset \mathcal{F}_T$

$\mathcal{F}_T$ = complete knowledge.

Equivalent Statements:

- x_t is $\mathcal{F}_t$ - measurable
- x_t is a function of $\xi_{[t]}$. (measurable function but every function is measurable in finite probability space
- x_t is a function of the node at time t.
- $x_t = E[x_t \mid \mathcal{F}_t]$ (linear constraint useful in building theory of optimality conditions)
- $x \in \mathcal{Y}$

$$\langle x, y \rangle \doteq \sum_{s=1}^{S} \sum_{t=1}^{T} p_s (x_t^s)^T y_t^s$$

orthogonal projection wrt this scalar product

$$x = Px$$

Computationally Convenient representation

Dense constraint: Every decision must be equal to the average

Sparse Constraint: Every decision must be equal to its "neighbor"

$$x_1^1 - x_1^2 = 0, \; x_1^2 - x_1^3 = 0, \ldots, x_1^7 - x_1^8 = 0$$

$$Q_t(x_{t-1}, \xi_{[t]}) = \inf_{x_t} \left\{ f_t(x_t, \xi_t) + \mathcal{Q}_{t+1}(x_t, \xi_{[t]}) : B_t x_{t-1} + A_t x_t = b_t \right\}$$

where $\mathcal{Q}_{t+1}(x_t, \xi_{[t]}) \doteq E\left[Q_{t+1}(x_t, \xi_{[t+1]}) \mid \xi_{[t]}\right]$

is convex.

Lagrangian

$$L_t(x_t, \pi_t) \doteq f_t(x_t, \xi_t) + \mathcal{Q}_{t+1}(x_t, \xi_{[t]}) + \pi_t^T\left(b_t - B_t x_{t-1} - A_t x_t\right)$$

(Adjoint equations)

Adjoint system that goes backwards in time w.r.t. dynamical system that goes forward in time. (control theory)

Current sensitivity is function of future sensitivity plus

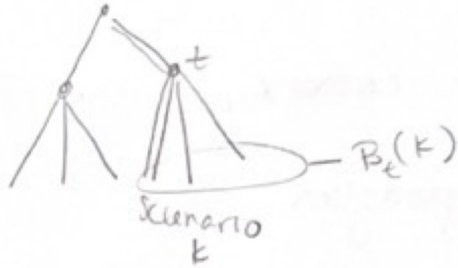
$$\langle x, y \rangle_* = \sum_{k=1}^{K} p_k \sum_{t=1}^{T} (x_t^k)^T y_t^k = \sum_{k=1}^{K} \sum_{t=1}^{T} p_k (x_t^k)^T (y_t^k)$$

- $P^2 = P$
  $P(Px) = P(x)$
- $\langle Px, y \rangle_* = \langle x, Py \rangle_*$

$$Px \leftrightarrow x_t = E[x_t \mid \mathcal{F}_t] \quad \forall t = 1, \dots, T$$

Bundle (scenarios that share same history upto t)

$$\bar{x} = Px \iff \bar{x}_t^k = \frac{\sum_{l \in B_t(k)} P_l x_t^l}{\sum_{l \in B_t(k)} P_l} \qquad B_t(k) \doteq \{ l : \xi^l_{[t]} = \xi^k_{[t]} \}$$



Let $\bar{y} = Py$

Then we must check

$$\langle \bar{x}, y \rangle_* = \langle x, \bar{y} \rangle$$

$$\bar{x}_t = E[x_t \mid \mathcal{F}_t]$$

$$\langle \bar{x}, y \rangle_* = E\left[\sum_{t=1}^{T} \bar{x}_t^T y_t\right] = E\left[\sum_{t=1}^{T} \left(E[x_t \mid \mathcal{F}_t]\right)^T y_t\right]$$

$$= \sum_{t=1}^{T} E\left[E[x_t \mid \mathcal{F}_t]^T y_t\right]$$

by tower property

$$= \sum_{t=1}^{T} E\left[E\left[E[x_t \mid \mathcal{F}_t]^T y_t \mid \mathcal{F}_t\right]\right]$$

$$= \sum_{t=1}^{T} E\left[E[x_t \mid \mathcal{F}_t]^T E[y_t \mid \mathcal{F}_t]\right]$$

$$= \langle \bar{x}, \bar{y} \rangle_*$$

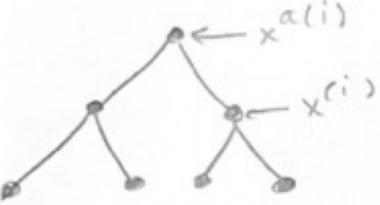
$$= \langle x, \bar{y} \rangle_* \text{ by symmetry}$$

So it is a projection

tree formulation (Nested Cutting Plane)

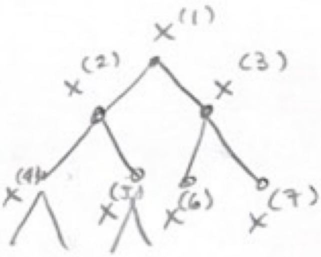
if node i is on level t: $f^{(i)}(x^{(i)}) = c_t^T x^{(i)}$

$p^{(i)}$: probability of hitting node i



constraints must be written such that they only relate $x^{(i)}$ to $x^{a(i)}$ but not previous nodes. (look backward only one step)

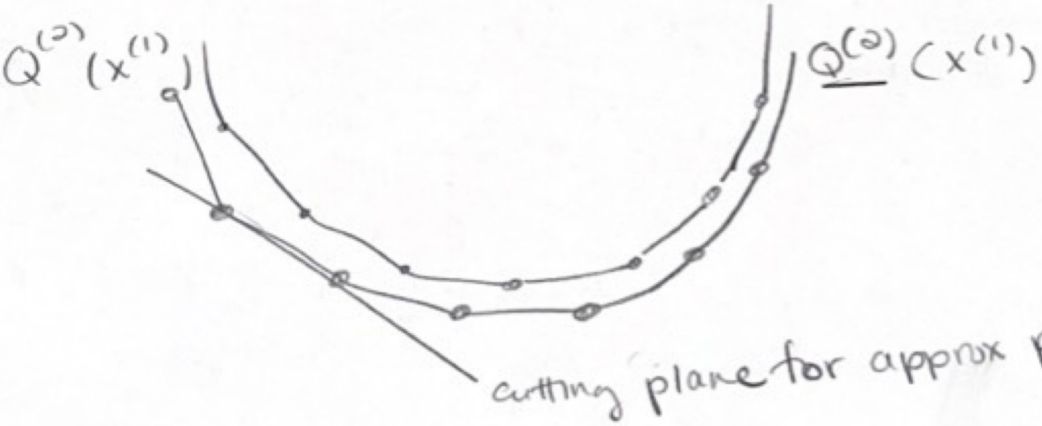
$$\min_{i \in N} p^{(i)} f^{(i)}(x^{(i)}) \quad \text{s.t.} \quad T^{(i)} x^{a(i)} + W^{(i)} x^{(i)} = h^{(i)}, \quad i \in N \setminus \{1\}$$
$$W^{(1)} x^{(1)} = h^{(1)}$$



Then we can view the tree problem as a collection of Nested Two stage problems

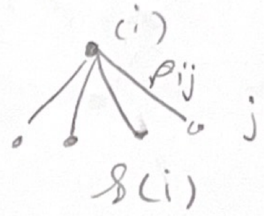
Approx by $\underline{Q}^{(4)}(x^{(2)})$, $\underline{Q}^{(5)}(x^{(2)})$ → cutting plane approx

and we know $Q^{(4)}(x^{(2)}) \geq \underline{Q}^{(5)}(x^{(2)})$



feasibility cuts too complex in this setting

- x decisions move down the tree
- cuts (information on shape of cost to go function) move up the tree to improve the approximation
- improved approximations are used to find new decisions
- repeat



$$\min f^{(i)}(x^{(i)}) + \sum_{j \in S(i)} [p_{ij} v_j]$$

$v_j \geq$ all cuts for node j $\quad j \in S(i)$

$T^{(i)} x^{a(i)} + W^{(i)} x^{(i)} = h^{(i)} \quad \leftarrow \pi^{(i)}$

$x^{(i)}$ has local constraints

- Stochastic Dual Dynamic Programming $\rightarrow$ For AR process: we can get stagewise independence with state variables
  - assume independence between stages $\rightarrow$ can share cuts between nodes at same level t

- Römisch – Scenario tree / Energy systems pruning using probability metrics

- Powell – forward pass on random path in tree, many passes $\rightarrow$ approx converges