# Introduction to algorithms for recourse models

Maarten H. van der Vlerk
University of Groningen

Tucson (AZ), October 9, 2004

## Outline

- Introduction
  - recourse model
  - difficulties
- Special cases:
  large-scale LP, SR, CR, multi-stage
  - motivation
  - properties
  - algorithms
  - example
- Concluding remarks
  - references

## Recourse model (two stage)

Motivation:
LP (. . . ) with random parameters

$$\begin{array}{lll} \min_x & cx & \\ \text{s.t.} & Ax = b & m_1 \text{ det. constr.} \\ & T(\omega)x = h(\omega) & m_2 \text{ random constr.} \\ & x \geq 0 & x \in \mathbb{R}^{n_1} \end{array}$$

with $\omega$ random

Decision problem: choose $x$ before $\omega$ known

Extend model: allow recourse actions to correct observed infeasibilities $h(\omega) - T(\omega)x$

Minimize costs: for $x$ and $\omega$ fixed

$$Q(x, \xi(\omega)) := \begin{array}{ll} \min_y & q(\omega)y \\ \text{s.t.} & Wy = h(\omega) - T(\omega)x \\ & y \in \mathbb{R}^{n_2}_+ \end{array}$$

with $r$-vector $\xi(\omega)$, known distribution
$\xi(\omega) := \big(q(\omega), T_1(\omega), \ldots, T_{m_2}(\omega), h(\omega)\big)$

All in LP context

For every *first-stage* decision $x$

$$\mathcal{Q}(x) := \mathbb{E}_\xi Q(x, \xi(\omega))$$

gives associated expected recourse costs

Two-stage linear recourse problem (DE)

$$\min \{cx + \mathcal{Q}(x) : x \in X\} \qquad \text{(RP)}$$

where $X := \{x \in \mathbb{R}^{n_1}_+ : Ax = b\}$

Choose $x \leftarrow$ min total expected costs

- direct costs $cx$
- expected future costs $\mathcal{Q}(x)$

Extensions:

- integer variables [Shabbir Ahmed]
- random recourse: $W(\omega)$
- non-linear
- risk: include e.g. var $Q(x, \xi(\omega))$ in objective [keynote lecture Rockafellar]
- chance constraints [Rene Henrion]
- multi-stage (later)

How to solve recourse problem RP

$$\min\{cx + \mathcal{Q}(x) : x \in X\}$$

Observe

- $X$ is convex set
- $\mathcal{Q}$ is convex function (!)

$\longrightarrow$ RP is CONVEX opt. problem (CP)

Theory for CP: KKT conditions, algorithms
$\longrightarrow$ well-solved ...

Why do we need special algorithms for RP ?

Consider e.g. descent algorithm for CP:
In each iteration $t$, need

- function evaluation $cx^t + \mathcal{Q}(x^t)$
- first-order information "$\nabla$" $\mathcal{Q}(x^t)$
- Hessian, ...

That is:
calculate $\mathcal{Q}(x)$ and "$\nabla$" $\mathcal{Q}(x)$ MANY TIMES

However

$$\begin{aligned}\mathcal{Q}(x) &= \mathbb{E}_\xi Q(x, \xi(\omega)) \\ &= \mathbb{E}_\xi \min_{y \geq 0} \left\{q(\omega)y : Wy = h(\omega) - T(\omega)x\right\}\end{aligned}$$

One function evaluation $\mathcal{Q}(x)$:

- solve 2nd-stage LP for all realizations $\xi$
- take expectation

If $\xi$ discrete with

$$p^{i_1 i_2 \ldots i_r} := \Pr\{\xi_1 = \xi_1^{i_1}, \xi_2 = \xi_2^{i_2}, \ldots, \xi_r = \xi_r^{i_r}\}$$

then

$$\mathcal{Q}(x) = \sum_{i_1}\sum_{i_2}\ldots\sum_{i_r} p^{i_1 i_2 \ldots i_r} Q(x, (\xi_1^{i_1}, \xi_2^{i_2}, \ldots, \xi_r^{i_r}))$$

If $\xi$ continuous with joint pdf $f(s)$, $s \in \mathbb{R}^r$

$$\mathcal{Q}(x) = \int\int\ldots\int Q(x, s) f(s)\, ds_1\, ds_2\, \ldots\, ds_r$$

Remark: no closed form for $Q(x, s)$ in general

HOPELESS in general ...
(e.g. continuous, $r \geq 4$)

Remark: Similar in CC: evaluation $\Pr\{T(\omega)x \geq h(\omega)\}$

Consider SPECIAL CASES

- (small) discrete distribution $\longrightarrow \mathbb{E}_\xi$ easy
- special structure 2nd-stage LP
  $\longrightarrow Q(x, \xi(\omega))$ easy

Use structural / mathematical properties to

- evaluate $\mathcal{Q}(x)$ and "$\nabla$" $\mathcal{Q}(x)$, many times
- optimize

Borrow from det. LP, CP, ...

Finite discrete distribution

Assume $\Xi = \{\xi^1, \xi^2, \ldots, \xi^S\}$ with
$\Pr\{\xi = \xi^s\} = p^s$, $s = 1, \ldots, S$

Notation: $\xi^s = (q^s, T_1^s, \ldots, T_{m_2}^s, h^s)$

Then $\mathbb{E}_\xi$ easy (?): $\mathcal{Q}(x) = \sum_{s=1}^{S} p^s Q(x, \xi^s)$

In this case

$$\begin{aligned}&\min\left\{cx + \mathcal{Q}(x) : x \in X\right\} \\ &= \min\left\{cx + \sum_{s=1}^{S} p^s Q(x, \xi^s) : Ax = b,\ x \geq 0\right\}\end{aligned}$$

$$\begin{array}{llllllll} = \min & cx & + & p^1 q^1 y^1 & + & \ldots & + & p^S q^S y^S \\ & Ax & & & & & & = b \\ & T^1 x & + & W y^1 & & & & = h^1 \\ & \vdots & & & & \ddots & & \\ & T^S x & & & & & + & W y^S = h^S \\ & x \geq 0 & & y^1 \geq 0 & & \ldots & & y^S \geq 0 \end{array}$$

RP is large-scale LP problem

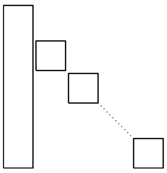
Idea:
approximate continuous distr. by discrete
$\longrightarrow$ solve every RP by solving large-scale LP

In general too large: this LP has

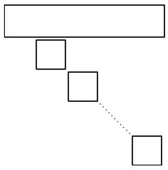
- $n_1 + Sn_2$ variables
- $m_1 + Sm_2$ constraints

Example: $\xi = (\xi_1, \xi_2, \ldots, \xi_9)$ independent,
10 realizations each $\longrightarrow S = 10^9$

Observe: large-scale LP has special structure



dual
block
angular



- Exploit structure $\longrightarrow$ decomposition (later)
- Solve by interior-point method
  (no details; see e.g. tutorial Birge (Berlin '01) on `http://stoprog.org`)

Easy 2$^{\text{nd}}$-stage problem: Simple Recourse

Motivation: algorithmic

- How to use special structure / properties?
- Introduce solution approaches in easy setting

Simple Recourse (SR):
linear penalty costs for shortage & surplus
w.r.t. each individual random constraint

Simplify:

- assume $q$ and $T$ fixed
- only rhs random: $h(\omega) =: h \in \mathbb{R}^{m_2}$

$$\begin{aligned} Q(x,h) &= \min_{y \geq 0} \; q^+ y^+ + q^- y^- \\ &\quad \text{s.t. } y^+ - y^- = h - Tx \\ &= \sum_{i=1}^{m_2} \min_{y_i \geq 0} \; q_i^+ y_i^+ + q_i^- y_i^- \\ &\qquad \text{s.t. } y_i^+ - y_i^- = h_i - T_i x \end{aligned}$$

separates: min problem for each constraint

Assuming $q^+ + q^- \geq 0 \quad \Leftrightarrow Q(x,h) > -\infty$
(*sufficiently expensive recourse*)

$$Q(x,h) = \sum_{i=1}^{m_2} v_i(z_i, h_i), \quad z := Tx \text{ tender var.}$$

where

$$v_i(z_i, h_i) := q_i^+ (h_i - z_i)^+ + q_i^- (h_i - z_i)^-$$

and $(s)^+ := \max\{0, s\}$, $(s)^- := \max\{0, -s\}$

Indeed, $Q(x,h)$ EASY: separable, closed form

Hence $\mathcal{Q}(x) = \Psi(z)$, with $z := Tx$

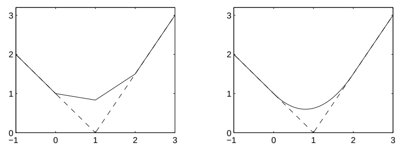
$$\Psi(z) = \sum_{i=1}^{m_2} \Psi_i(z_i)$$

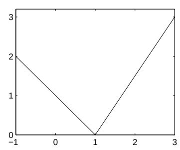
where

$$\Psi_i(z_i) = q_i^+ \mathbb{E}_{h_i} (h_i - z_i)^+ + q_i^- \mathbb{E}_{h_i} (h_i - z_i)^-$$

is separable: each term $\Psi_i$

- depends only on $z_i$
- expectation wrt marginal distr. $h_i$: EASY





$q_i^+ = 1 \; q_i^- = 1.5$
L: $v_i(z_i, 1)$
TL: $\Psi_i(z_i)$, $h_i \sim U\{0,1,2\}$
TR: $\Psi_i(z_i)$, $h_i \sim U(0,2)$

Straightforward analysis: properties of $\Psi_i$
$\longrightarrow$ properties $\Psi$ and $\mathcal{Q}$

$$\Psi_i(z_i) = q_i^+ (\mu_i - z_i) + (q_i^+ + q_i^-) \int_{-\infty}^{z_i} F_i(t)dt$$

with $F_i$ cdf of $h_i \in H_i$ and $\mu_i = \mathbb{E}_{h_i} h_i$ (!)

Properties $\Psi_i$ immediate from $F_i$

- convex
- finite iff $\mathbb{E}_{h_i}|h_i| < \infty$
- subdifferential $-q_i^+ + (q_i^+ + q_i^-) \left[\Pr\{h_i < z_i\}, F_i(z_i)\right]$
- piecewise linear if $h_i$ discrete
- differentiable if $h_i$ continuous
- asymptotes $v_i(z_i, \mu_i)$
- $z_i \notin \operatorname{conv} H_i$: equal to asymptotes

SR problem

$$\min \left\{ cx + \sum_{i=1}^{m_2} \Psi_i(z_i) : x \in X,\ z = Tx \right\}$$

Algorithms for SR: use knowledge of $\mathcal{Q}$ ($\Psi$)

- closed form
- properties

(A) Specific distribution $\longrightarrow$ formula for $\Psi$

Example: suppose each $h_i \sim \mathcal{E}(\lambda_i)$

$$\Psi_i(z_i) = \begin{cases} q_i^+(1/\lambda_i - z_i), & z_i \leq 0 \\ q_i^+(1/\lambda_i - z_i) + \\ \quad (q_i^+ + q_i^-)\left(\frac{e^{-\lambda_i z_i} - 1}{\lambda_i} + z_i\right), & z_i \geq 0 \end{cases}$$

- convex
- differentiable

$\longrightarrow$ solve SR by favorite CP algorithm

(B) Finite discrete distribution [Wets]

Properties / formula:
$\longrightarrow$ SR equivalent to SMALL LP

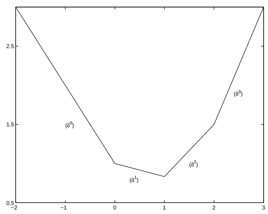
$h_i \sim F_i$, realizations $h_i^1 < h_i^2 < \ldots < h_i^{S_i}$
w.p. $p_i^s$, $\mathbb{E}_{h_i} h_i = \mu_i$

Then $\Psi_i$ is piecewise linear convex, with slope

$$(\Psi_i)'_+(z_i) = -q_i^+ + \left(q_i^+ + q_i^-\right) F_i(h_i^k) =: \delta_i^k$$

for $z_i \in \left[h_i^k, h_i^{k+1}\right)$, $k = 0, 1, \ldots, S_i$
(with $h_i^0 := -\infty$ and $h_i^{S_i+1} := +\infty$)

Observe: $-q_i^+ = \delta_i^0 < \delta_i^1 < \ldots < \delta_i^{S_i} = q_i^-$




$q_i^+ = 1$, $q_i^- = 1.5$
$h_i \sim \mathcal{U}\{0, 1, 2\}$

Hence (LP representation of PL function)

$$\begin{aligned} \Psi_i(z_i) = \min_u \quad & \delta_i u_i \equiv \sum_{k=0}^{S_i} \delta_i^k u_i^k \\ \text{s.t.} \quad & \sum_{k=0}^{S_i} u_i^k = z_i - \mu_i \\ u_i \in U_i & \begin{cases} u_i^0 \leq h_i^1 - \mu_i \\ 0 \leq u_i^k \leq h_i^{k+1} - h_i^k, \quad k = 1, \ldots, S_i \end{cases} \end{aligned}$$

$\longrightarrow$ SR equivalent to SMALL LP

$$\begin{aligned} \min_{x,u} \quad & cx + \sum_{i=1}^{m_2} \delta_i u_i \\ \text{s.t.} \quad & Ax = b \\ & T_i x - \sum_{k=0}^{S_i} u_i^k = \mu_i, \quad i = 1, \ldots, m_2 \\ & u_i \in U_i, \qquad\qquad i = 1, \ldots, m_2 \end{aligned}$$

- $n_1 + \sum_{i=1}^{m_2} (S_i + 1)$ variables
- $m_1 + m_2$ constraints
- $\sum_{i=1}^{m_2} S_i$ simple bounds

Indeed, small compared to "naive" LP

- $n_1 + 2m_2 \prod_{i=1}^{m_2} S_i$ variables
- $m_1 + m_2 \prod_{i=1}^{m_2} S_i$ constraints

assuming independence

Example: $h \in \mathbb{R}^9$ (independent),
10 realizations each

$$(n_1 + 99) \times (m_1 + 9) + 99 \text{ simple bounds}$$

versus

$$(n_1 + 18 \cdot 10^9) \times (m_1 + 9 \cdot 10^9)$$

Conclusion: exploit structure !

Other example: 2nd-stage network flow
[Wallace]

(C) SR continuous distribution (only rhs $h$)

Idea:

- construct discrete approximation of distribution
- solve as small LP

$\longrightarrow$ approximate solution

Same approach for more general recourse
Introduce concepts / details in simple setting

Developing idea:

Would like "small" discrete approximation
$\longrightarrow$ "good" approximation of

- true optimal solution
- true optimal value

i.e. not necessarily of original distribution

Construct lower and upper bounds,
satisfied if gap is small enough . . .

Use properties of $\mathcal{Q} \leftarrow \Psi_i(z_i) := \mathbb{E}_{h_i} v_i(z_i, h_i)$

$$v_i(z_i, h_i) := q_i^+ (h_i - z_i)^+ + q_i^- (h_i - z_i)^-$$

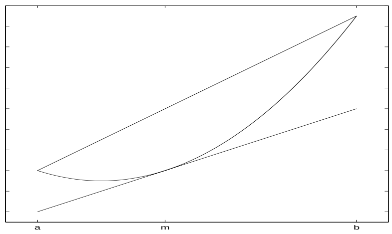
is finite and convex in $h_i$ for fixed $z_i$

For any convex function $\varphi : \mathbb{R} \mapsto \mathbb{R}$
and random variable $\omega$ with

- support in $[a, b]$
- mean value $\mu$

$$\begin{array}{lcll} \varphi(\omega) & \geq & \varphi(\mu) + u(\omega - \mu) & \left(\text{subgradient inequality}\right) \\ \varphi(\omega) & \leq & \text{chord}(a, b), \quad \omega \in [a, b] & (\text{E-M ineq.}) \end{array}$$

with $u$ a subgradient of $\varphi$ at $\mu$




Hence

$$\mathbb{E}_\omega \varphi(\omega) \geq \varphi(\mu) \qquad \text{(Jensen ineq.)}$$

and

$$\mathbb{E}_\omega \varphi(\omega) \leq \frac{b - \mu}{b - a} \varphi(a) + \frac{\mu - a}{b - a} \varphi(b) \qquad \text{(E-M ineq.)}$$

Apply to $\Psi_i(z_i) = \mathbb{E}_h v_i(z_i, h_i)$ assuming

- support $h_i$ in $H_i := [a_i, b_i]$ (truncate)
- mean value $\bar{h}_i$

For any $z_i$, to obtain

- lower bound for $\Psi_i(z_i)$: replace $h_i$ by $\eta_i$

  $\Pr\left\{\eta_i = \bar{h}_i\right\} = 1$
- upper bound for $\Psi_i(z_i)$: replace $h_i$ by $\upsilon_i$

  $\Pr\{\upsilon_i = a_i\} = \frac{b_i - \bar{h}_i}{b_i - a_i} \qquad \Pr\{\upsilon_i = b_i\} = \frac{\bar{h}_i - a_i}{b_i - a_i}$

Indeed, "small" discrete approximations
However, bad approximation of $\Psi_i$ . . .

Improve approximations:

- partition support $H_i = \bigcup_{j=1}^{n_i} H_i^j$

  $H_i^j := (s_i^{j-1}, s_i^j]$, $a_i = s_i^0 < s_i^1 < \ldots < s_i^{n_i} = b_i$
- apply approach to conditional distributions on each $H_i^j$

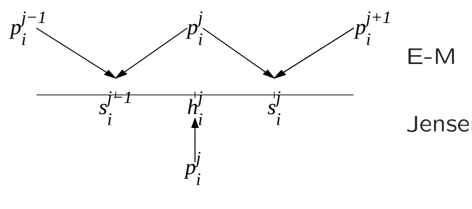
Result:

LB (Jensen): $n_i$-point discrete distribution

- on conditional means $\bar{h}_i^j \in H_i^j$
- $\Pr\left\{\eta_i = \bar{h}_i^j\right\} = \Pr\left\{h_i \in H_i^j\right\} =: p_i^j$

UB (E-M): $n_i + 1$-point discrete distribution

- on endpoints $a_i = s_i^0, s_i^1, \ldots, s_i^{n_i} = b_i$
- $\Pr\left\{\upsilon_i = s_i^j\right\} = \frac{\bar{h}_i^j - s_i^{j-1}}{s_i^j - s_i^{j-1}} p_i^j + \frac{s_i^{j+1} - \bar{h}_i^{j+1}}{s_i^{j+1} - s_i^j} p_i^{j+1}$



Conceptual algorithm SR, continuous distr.

Iteration $t$:
current partitions $P_i^t$, $i = 1, \ldots, m_2$
$\longrightarrow \eta^t$ (Jensen) and $\upsilon^t$ (E-M)

- solve SR with rhs $\eta^t$ (small LP: fast!)
  $\longrightarrow$ solution $x^t$, $\mathrm{LB}^t$ on opt. value SR
- evaluate $cx^t + \mathcal{Q}^t(x^t)$ with rhs $\upsilon^t$
  $\longrightarrow$ $\mathrm{UB}^t$ on opt. value SR
- if $\mathrm{UB}^t - \mathrm{LB}^t \leq \varepsilon^*$ STOP; $x^* = x^t$
- refine partitions $\longrightarrow P_i^{t+1}$; REPEAT

Converges to optimal value / solution (...)
[Stability: see plenary lecture Römisch]

How to refine partition?
Several options:

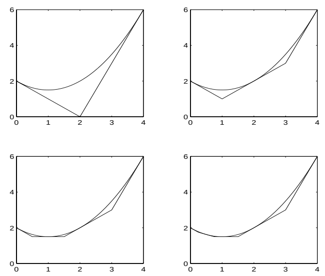
(1) Brute force:
Split every cell in current partition
e.g. equal length or on conditional mean
$\longrightarrow$ # realizations $\eta_{t+1}$ and $\upsilon_{t+1}$ doubled
i.e. exponential growth # realizations !

(2) Adapt to problem:
Split cells which contribute most to gap
$\mathrm{UB}^t(x^t) - \mathrm{LB}^t(x^t)$
e.g. split one cell in $H_{i,t}$, $i = 1, \ldots, m_2$
$\longrightarrow$ linear growth # realizations

(3) Specific for SR with random rhs $h$
(uses $\Psi_i(z_i)$ = asymptotes for $z_i \notin \mathrm{conv}\, H_i$)
For $i = 1, \ldots, m_2$

- split cell $\subset \mathrm{conv}\, H_i$ containing $z_i^t = T_i x^t$ at $z_i^t$
- if no such cell exists $\mathrm{LB}_i^t(x^t) = \Psi_i(T_i x^t)$
  $\longrightarrow$ no update needed

Remark: (3) implies $\mathrm{LB}^{t+1}(x^t) = \mathcal{Q}(x^t)$
$\longrightarrow$ no need for UB for SR [Kall & Stoyan]



Function $\Psi_i$
$h_i \sim \mathcal{U}(0,4)$
$q_i^+ = 1$, $q_i^- = 3$

"General" two-stage recourse problems

Recall two-stage recourse problem is

$$\min \left\{ cx + \mathcal{Q}(x) : x \in X \right\}$$

where $X := \{x \geq 0 : Ax = b\}$,

$$\mathcal{Q}(x) := \mathbb{E}_\xi Q(x, \xi(\omega))$$

and

$$Q(x, \xi(\omega)) := \min_{y \geq 0} \left\{ q(\omega) y : Wy = h(\omega) - T(\omega) x \right\}$$

Second-stage LP problem specified by
recourse structure $\big(q(\omega), W\big)$
e.g. SR: $W = (I_m \ -I_m)$ and $q(\omega) = (q^+ \ q^-)$

Next: algorithms for "general" recourse
$\longrightarrow$ conditions on $\big(q(\omega), W\big)$ such that
problem is "nice" for algorithms

(1) Simplify presentation: $q(\omega) = q$ fixed

(2) Avoid $Q(x, \xi(\omega)) = \pm\infty$

$Q(x, \xi(\omega)) = -\infty$: infinite reward for recourse
$\longrightarrow$ always exclude:
Dual feasible set $\{\lambda \in \mathbb{R}^{m_2} : \lambda W \leq q\} \neq \emptyset$
Called *sufficiently expensive recourse*

$Q(x, \xi(\omega)) = +\infty$: no recourse action possible
Interpretation OK; inconvenient in algorithms
$\longrightarrow$ exclude (for now):
$\mathrm{pos}\, W := \{t : \exists y \geq 0, Wy = t\} = \mathbb{R}^{m_2}$
Called *complete recourse*

(3) $\mathcal{Q}(x)$ finite:

Given SER & CR $\longrightarrow Q(x, \xi(\omega))$ finite,
depends on distribution $\xi(\omega)$.
Since $q$ is fixed, sufficient condition is

$$\mathbb{E}_\xi |\xi(\omega)| < \infty \text{ (componentwise)}$$

Algorithms tailored to properties RP
$\longrightarrow$ study properties of $Q(x, \xi(\omega))$ and $\mathcal{Q}(x)$

SER & CR $\longrightarrow Q(x, \xi(\omega))$ finite.
Hence, by LP duality

$$Q(x, \xi(\omega)) := \min_{y \geq 0} \left\{ qy : Wy = h(\omega) - T(\omega)x \right\} = \max_{\lambda \in \mathbb{R}^{m_2}} \left\{ \lambda \left( h(\omega) - T(\omega)x \right) : \lambda W \leq q \right\}$$

Dual set is polyhedral, non-empty, bounded
$\longrightarrow$ finitely many extreme points $\lambda^1, \lambda^2, \ldots, \lambda^K$

$$Q(x, \xi(\omega)) = \max_{k=1,\ldots,K} \lambda^k \left( h(\omega) - T(\omega)x \right)$$

$Q(x, \xi)$ is pointwise maximum of
finitely many linear functions
$\longrightarrow$ for fixed $\xi$,
$Q(x, \xi)$ is polyhedral convex function of $x$

Hence $\mathcal{Q}(x)$ is convex

- polyhedral if $\xi$ finite discrete
- differentiable if $\xi$ continuous (...)

Algorithms for CR model, $\xi$ discrete

Assume $\xi$ discrete: for $s = 1, \ldots, S$
$\xi^s = (h^s, T^s)$ with $p^s := \Pr\{\xi = \xi^s\}$

Equivalent large-scale LP

$$\min cx + \sum_{s=1}^{S} p^s q y^s : \quad Ax = b, \quad T^s x + W y^s = h^s \quad \forall s, \quad x \geq 0, \quad y^s \geq 0 \quad \forall s$$

Special structure: L-shaped, ...
$\longrightarrow$ decomposition, algorithms

$\mathcal{Q}(x)$ polyhedral: pointwise max of
finitely many linear functions:
supporting hyperplanes of epi $\mathcal{Q}$

$$\text{RP} = \min_{x \in X, \theta \in \mathbb{R}} \left\{ cx + \theta : \theta \geq E^j x + e^j, \ j \in J \right\}$$

Algorithmic perspective:

- How to construct *optimality cuts* $\theta \geq E^j x + e^j$ ?
- $|J|$ too large $\longrightarrow$ don't add all ...

Optimality cuts: outer linearization of $\mathcal{Q}$

Recall

$$Q(\bar{x}, \xi^s) = \max_{k=1,\ldots,K} \lambda^k \left( h^s - T^s \bar{x} \right)$$

$\longrightarrow$ opt. dual solution $\lambda^{k(\bar{x},s)}$:
$-\lambda^{k(\bar{x},s)} T^s$ subgradient of $Q(\cdot, \xi^s)$ at $\bar{x}$

Hence

$$\mathcal{Q}(\bar{x}) = \sum_{s=1}^{S} p^s \lambda^{k(\bar{x},s)} (h^s - T^s \bar{x}) = E(\bar{x})\, \bar{x} + e(\bar{x})$$

where

$$E(\bar{x}) := -\sum_{s=1}^{S} p^s \lambda^{k(\bar{x},s)} T^s \quad \in \partial \mathcal{Q}(\bar{x})$$

$$e(\bar{x}) := \sum_{s=1}^{S} p^s \lambda^{k(\bar{x},s)} h^s$$

so that

$$\theta \geq E(\bar{x})\, x + e(\bar{x})$$

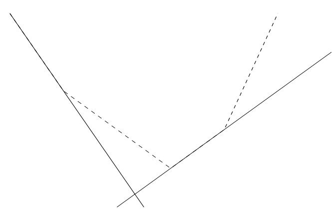
is an optimality cut at $\bar{x}$

That is: Compute $\mathcal{Q}(\bar{x})$ by solving
DUAL of 2nd-stage problems, $s = 1, \ldots, S$
$\longrightarrow \mathcal{Q}(\bar{x})$ and optimality cut, sharp at $\bar{x}$

Remark: dual feasible set is $\{\lambda : \lambda W \leq q\}$
$\longrightarrow$ hot starts, bunching, ...

Opt. cuts $\longrightarrow$ outer linearization of $\mathcal{Q}$
Hence

- subset of all cuts $\longrightarrow$ lower bound of $\mathcal{Q}$
- add cut at $\bar{x}$ $\longrightarrow$ approximation sharp at $\bar{x}$



L-shaped algorithm for RP [Van Slyke & Wets]:
add cuts iteratively

Iteration $\nu$:

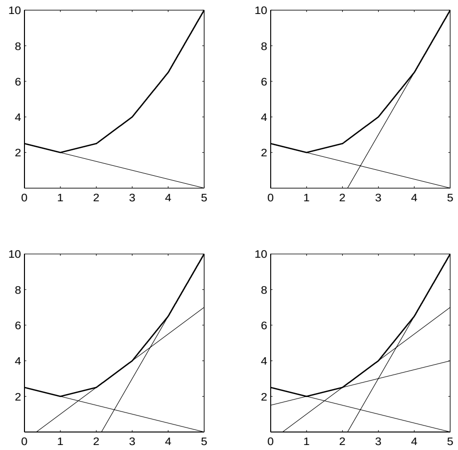
(i) Solve MASTER LP problem

$$\min_{\substack{x\in X\\ \theta\in\mathbb{R}}} \left\{cx+\theta : \theta \geq E^j x + e^j,\ j=1,\dots,\nu-1\right\}$$

$\longrightarrow$ current optimal solution $(x^\nu, \theta^\nu)$

(ii) Solve $S$ dual 2nd-stage problems
with $x^\nu$ and $\xi^s$, $s=1,\dots,S$
$\longrightarrow$ $\mathcal{Q}(x^\nu) = E^\nu x^\nu + e^\nu$ and opt. cut at $x^\nu$

(iii) If $\theta^\nu \geq \mathcal{Q}(x^\nu)$ STOP: $x^\nu$ is optimal (!)
Otherwise: iteration $\nu+1$




L-shaped algorithm applied to SR problem

$$\min\{x/2+\mathcal{Q}(x) : 0 \leq x \leq 5\}$$

with $q^+ = 1$, $q^- = 3$,
$T(\omega)=1$ fixed, and $h(\omega) \sim U\{1,2,3,4\}$

What if recourse is not complete,
i.e. $Q(x,\xi) = +\infty$ is possible?

Define

$$\begin{aligned} K &:= \left\{x\in\mathbb{R}^{n_1} : Q(x,\xi^s) < +\infty,\ s=1,\dots,S\right\} \\ &= \left\{x\in\mathbb{R}^{n_1} : \mathcal{Q}(x) < +\infty\right\}\end{aligned}$$

$K$ is a polyhedral convex set (...)
Very many facets in general

Remark: If $X \subset K$ don't bother...
Known as *relatively complete recourse*
(hard to verify)

Obviously, if $\bar{x} \notin K$ then $\bar{x}$ not optimal
$\longrightarrow$ algorithmic approach: iterate,
cut off such $\bar{x}$ by a linear *feasibility cut*

In L-shaped algorithm: iteration $\nu$

(i) Solve master problem

$$\begin{aligned} \min_{x\in X, \theta\in\mathbb{R}} \quad & cx+\theta \\ \text{s.t.} \quad & D^l x \geq d^l, \qquad l=1,\dots,m \\ & \theta \geq E^j x + e^j,\quad j=1,\dots,n \end{aligned}$$

$\longrightarrow$ optimal solution $(x^\nu, \theta^\nu)$

(ii') Solve auxiliary problems (CR):
for $s=1,\dots,S$, with $e := (1\,1 \dots\, 1)$

$$\begin{aligned} w(x^\nu,\xi^s) := \ & \min\ e(y^+ + y^-) \\ & \text{s.t}\ \ Wy + y^+ - y^- = h^s - T^s x^\nu \\ & \qquad y \geq 0,\ y^+ \geq 0,\ y^- \geq 0 \\ = \ & \max\{\sigma(h^s - T^s x^\nu) : \sigma W \leq 0, |\sigma| \leq e\} \end{aligned}$$

$\longrightarrow$ optimal dual solution $\sigma^{\nu s}$

$x^\nu \notin K$ iff $w(x^\nu,\xi^s) > 0$ for some $s$

In that case

$$\sigma^{\nu s}(h^s - T^s x) \begin{cases} \leq 0 & \forall\ x : \mathcal{Q}(x) < +\infty \\ > 0, & x = x^\nu \end{cases}$$

$\longrightarrow$ feasibility cut $D^{m+1}x \geq d^{m+1}$
Return to step (i) with $m \leftarrow m+1$

Otherwise proceed with (ii):
generate optimality cut

## Variants

Multi-cut L-shaped [Birge & Louveaux]

Generate opt. cuts for each $Q(x, \xi^s)$ instead of $\mathcal{Q}(x)$ (aggregated)

- more detailed information to master $\longrightarrow$ less iterations expected
- larger master problem
- Numerical evidence: faster if $S \not\gg m_1$ ($1^{\text{st}}$-stage constr.)

Regularized decomposition [Ruszczyński]

Multi-cut

NLP: Regularizing term $\|x - a^\nu\|^2$, with $a^\nu$ incumbent solution, in objective to overcome

- inefficient initial iterations
- degeneracy during later iterations $\longrightarrow$ keep at most $n_1 + S$ constraints

(Implementation: QDECOM)

- Numerical evidence: outperforms L-shaped and Multi-cut

## L-shaped: *time* decomposition large-scale LP

$$\min cx + \sum_{s=1}^{S} p^s q y^s : \quad \begin{aligned} &Ax = b \\ &T^s x + W y^s = h^s \quad \forall s \\ &x \geq 0, \qquad y^s \geq 0 \quad \forall s \end{aligned}$$

Indeed: $1^{\text{st}}$-stage variables $x$ fixed
$\longrightarrow$ $\mathcal{Q}(x)$: decomposed in $S$ small LPs
$\longrightarrow$ LB model of $\mathcal{Q}(x)$ by cutting planes

Similar approach for MIP with $x \in \mathbb{Z}^n$, $y \in \mathbb{R}^m$ known as Benders Decomposition

Alternative: decomposition by *scenarios* $\xi^s$:
Introduce copies $x^s$ of $x$, $s = 1, \ldots, S$

$$\min \sum_{s=1}^{S} p^s (c x^s + q y^s) : \quad \left.\begin{aligned} &A x^s = b \\ &T^s x^s + W y^s = h^s \\ &x^s \geq 0, \qquad y^s \geq 0 \end{aligned}\right\} \forall s$$
$$x^1 = x^2 = \ldots = x^S$$

Relax *non-anticipativity constraints*
$\longrightarrow$ decomposition in $S$ *scenario problems*

More details: multi-stage problems (later)

## Algorithms for CR with continuous $\xi$

(A) Conceptual algorithm:

- approximate using discrete distributions
- solve by e.g. L-shaped
- stop if approximation good enough, else: refine discrete approximations, repeat

Iteration $t$:

Current partition $P^t$ of $\Xi \subset \mathbb{R}^r$ (rectangular cells)

- Jensen $\longrightarrow$ $\eta^t$ on conditional means
  E-M $\longrightarrow$ $v^t$ on $2^r$ vertices of cells
- Solve RP with $\eta^t$ $\longrightarrow$ $x^t$ and LB$^t$
  Evaluate $x^t$ using $v^t$ $\longrightarrow$ UB$^t$
- If UB$^t$ − LB$^t \leq \varepsilon^*$ STOP: solution $x^t$
  else: update partition $P^t$; iteration $t+1$

How to update partition ?

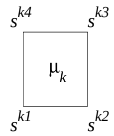
- E-M: evaluate $Q(x^t, \cdot)$ at $2^r$ vertices of each cell
- "Bad" choice never undone: extra work in iterations $t+1, \ldots$

$\longrightarrow$ Invest (CPU) time in "good" update . . .

## Current partition $P^t$: cells $\Xi_1, \ldots, \Xi_K$

Which cell(s) to partition ?
Choose $\Xi_k$ with largest gap UB$_k$ − LB$_k$

$$\begin{aligned} LB_k &:= p_k Q(x^t, \mu_k) \\ UB_k &:= \sum_{i=1}^{2^r} p_{ki} Q(x^t, s^{ki}) \end{aligned}$$




or all $\Xi_k$ with UB$_k$ − LB$_k > \varepsilon^t$

(# cells $\leftrightarrow$ # iterations: solve by L-shaped)

How to partition a cell $\Xi_k$ ?

- Choice of coordinate $\xi_j$, $j = 1, \ldots, r$
- Where ?
  e.g. equal size, at $\mu_{kj} := \mathbb{E}_\xi \xi_j | \xi \in \Xi_k$

Many heuristic choices possible . . .

The following works well in practice:

- Try each partition at $\mu_{kj}$, $j = 1, \ldots, r$
- Apply Jensen & E-M $\longrightarrow$ gap UB$_{kj}$ − LB$_{kj}$
- Keep partition which minimizes gap

Lot of work !

Each iteration of "discrete approximation" expensive

- update discrete approximations
- solve RP$^t$ by L-shaped

$\longrightarrow$ exact LB$^t$ and UB$^t$ (also for $t$ small)

Hopefully few iterations needed ...

Alternative: use "cheap iterations"

- sample from distribution $\xi$
- fast approximation $\mathcal{Q}(x^t)$
  $\longrightarrow$ opt. cut, LB in expectation
  (better cuts for $t$ large)

Willing to use many such iterations ...

Stochastic Decomposition [Higle & Sen]

Assume CR, $q$ fixed, $\mathcal{Q}(x) \geq 0$ (known LB)

Iteration $t$ of SD:

- solve master LP $\longrightarrow$ opt. solution $(x^t, \theta^t)$
- sample $\xi^t = (T^t, h^t)$
- update optimality cuts:

(i) Generate new cut at $x^t$:

Calculate $Q(x^t, \xi^t)$, i.e. solve 1 dual problem

$$\max \left\{ \lambda \left( h^t - T^t x \right) : \lambda W \leq q \right\} \longrightarrow \lambda_t^t$$

Set $\Lambda^t = \Lambda^{t-1} \cup \{\lambda_t^t\}$ (known dual vertices)

Calculate lower bound for $Q(x^t, \xi^s)$, $s = 1, \dots, t-1$

$$\max \left\{ \lambda \left( h^s - T^s x \right) : \lambda \in \Lambda^t \right\} \longrightarrow \lambda_s^t \quad \text{FAST!}$$

New cut at $x^t$:

$$\theta \geq \frac{1}{t} \sum_{s=1}^{t} \lambda_s^t \left( h^s - T^s x \right) =: E_t^t x + e_t^t$$

(ii) Update old cuts:

- expected to be loose (based on $\Lambda^{t-1}$)
- possibly invalid (sample)

For $s = 1, \dots, t-1$

$$\theta \geq \frac{t-1}{t} \left( E_s^{t-1} x + e_s^{t-1} \right) =: E_s^t x + e_s^t$$

Old cuts $\longrightarrow$ 0 as # iterations $t$ increases

Problem: $\{x^t\} \not\to x^*$

Solution: track incumbent solutions $\{\bar{x}_s^t\}$

- new incumbent only if
  $x^t$ "really better" than $\bar{x}_s^t$
- update cut at incumbent $\bar{x}_s^t$ using (i)

Identifiable subsequence of $\{\bar{x}_s^t\} \to x^*$

Stopping criterion: if for some time

- change in objective small AND
- no new dual vertices AND
- incumbent not changed

Other sampling methods

- Importance sampling [Dantzig, Infanger]
- External sampling [Shapiro et al]

Multi-stage recourse

Usual $H$-stage LP structure:

$$\begin{array}{llllll}
\min & c^1 x^1 + & c^2 x^2 + & \dots & c^{H-1} x^{H-1} + & c^H x^H \\
& W^1 x^1 & & & & = h^1 \\
& T^2 x^1 + W^2 x^2 & & & & = h^2 \\
& & T^3 x^2 + W^3 x^3 & & & = h^2 \\
& \vdots & & \ddots & & \\
& & & T^{H-1} x^{H-2} + W^{H-1} x^{H-1} & & = h^{H-1} \\
& & & & T^H x^{H-1} + W^H x^H & = h^H \\
& x^1 \geq 0 & x^2 \geq 0 & \dots & & x^H \geq 0
\end{array}$$

Stage $t$:

- decision vector $x^t$
- parameters: $W^t$ and $\xi^t = (c^t, T^t, h^t)$
- separated constraints: $x^t \geq 0$ (...)

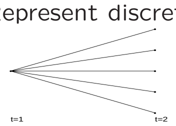
Constraint matrix has STAIRCASE structure: coupling between neighboring stages only

Multi-stage recourse: information structure

Recall: two-stage recourse

Decision – Observation – Decision

Represent discrete distribution by "tree":

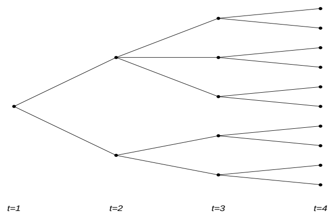



Decisions: solve (N)LP in each node of tree

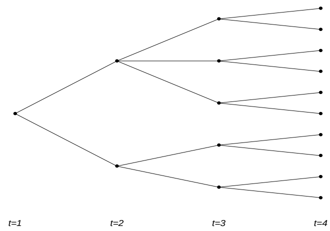
Multi-stage recourse:
Uncertainty revealed at different moments

D – O – D – . . . – O – D

Assume discrete distributions at $t = 2, \ldots, H$
Represented by *scenario tree*




Again: decision $x_k^t$ in every node $(t, k)$,
$t = 1, \ldots, H,\ k = 1, \ldots, K^t$




At every node $(t, k)$ $(1 < t < H)$

- observation $\xi_k^t = (c_k^t, T_k^t, h_k^t)$
- solve NLP

$$\begin{array}{ll} \min & c_k^t x_k^t + \mathcal{Q}_k^t(x_k^t) \\ \text{s.t.} & W^t x_k^t = h_k^t - T_k^t x_{a(k)}^{t-1} \\ & x_k^t \geq 0 \end{array}$$

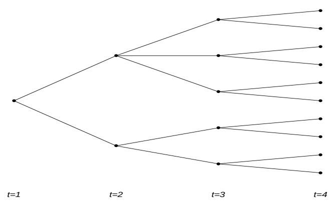
with $\mathcal{Q}_k^t(x_k^t)$ expected future costs
and $(t-1, a(k))$ unique *ancestor* node of $(t, k)$
$\longrightarrow$ connections between nodes !

Observations (cf. two-stage):

- MSRP is a large-scale LP (HUGE)
- each $\mathcal{Q}_k^t$ is a polyhedral function

Solvability: size of scenario tree

- Scenario generation [Georg Pflug]




Consider a node $(t, k)$ $(1 < t < H)$
"myopic view"

- for ancestor node $(t-1, a(k))$
  $(t, k)$ is a second-stage problem
- for every *child* node $(t+1, j),\ j \in D^t(k)$
  $(t, k)$ is the first-stage problem

$\longrightarrow$ MSRP consists of
nested two-stage problems

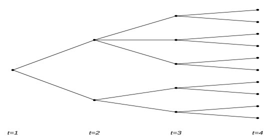
Algorithm: Nested L-shaped/Benders [Birge]
Master problem LP$(t, k)$ at every node

$$\begin{array}{lll} \min\limits_{x_k^t \geq 0,\ \theta_k^t} & c_k^t x_k^t + \theta_k^t & \\ \text{s.t.} & W^t x_k^t = h_k^t - T_k^t x_{a(k)}^{t-1} & \\ & E_{kj}^t x_k^t + \theta_k^t \geq e_{kj}^t, & j = 1, \ldots, n_k^t \\ & \text{(feasibility cuts)} & \end{array}$$

send information through tree ($\longleftrightarrow$)

Iteration of Nested L-shaped / Benders

FORWARD:
for $t = 1$ to $H$
  for $k = 1$ to $K^t$
    solve master LP$(t, k)$
    $\longrightarrow$ $x_k^t, \theta_k^t$ and multipliers $\lambda_k^t$: recourse constr.
      $\sigma_k^t$: opt. cuts
  end
end

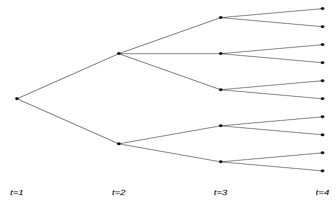



BACK:
for $t = H - 1$ to 1
  for $k = 1$ to $K^t$
    $E_k^t = E_k^t(\lambda_j^{t+1},\ j \in D^{t+1}(k))$
    $e_k^t = e_k^t(\lambda_j^{t+1}, \sigma_j^{t+1},\ j \in D^{t+1}(k))$
    $\bar{\theta}_k^t = e_k^t - E_k^t x_k^t$ (current approx. $\mathcal{Q}_k^t(x_k^t)$)
    if $\bar{\theta}_k^t > \theta_k^t$ then
      add opt. cut $(E_k^t, e_k^t)$ to LP$(t, k)$
    else
      if $t = 1$ STOP: $x_1^1$ optimal
    end
  end
end

Variants:

- feasibility cuts for non-CR
- various strategies Forward/Back

Alternative approach: scenario decomposition




For each *scenario*

$$\xi_s := (\xi_s^1, \xi_s^2, \ldots, \xi_s^H), \quad s = 1, \ldots, S := K^H$$

define the *scenario problem*

$$\min f_s(x)$$

with $x := (x^1, x^2, \ldots, x^H)$ and

$$f_s(x) := \begin{cases} \sum_{t=1}^{H} c_s^t x^t, & x \in X(\xi_s) \\ +\infty, & \text{otherwise} \end{cases}$$

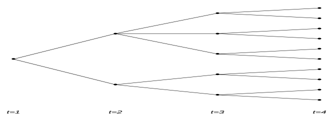
Observe: LP problem (relatively small)

Then MSRP is

$$\min_x \sum_{s=1}^{S} p_s f_s(x_s) : \text{non-anticipativity constr.}$$

NAC: For $t = 1, \ldots, H-1$

$$x_s^t = x_\sigma^t \text{ for all } s, \sigma\text{: } \xi_s^{[1,t]} = \xi_\sigma^{[1,t]}$$




Various representations of NAC, e.g.

- $Gx = 0$ with $G$ sparse
- for all $s$ and $t$: $x_s^{[1,t]} = \mathbb{E}\{x^{[1,t]} | \xi_s^{[1,t]}\}$

Lagrangian Methods
(Progressive Hedging [Rockafellar & Wets])

Relax non-anticipativity constraints
Iterate:

- solve Lagrangian scenario subproblems (including regularizing term)
- update multipliers

Advantage: problem structure intact

Concluding remarks

SP algorithms for recourse problems

- are necessary
- use structure / properties
- allow solving realistic problems

This tutorial

- outline of main algorithmic concepts
- many advanced ideas not covered

Challenges: many, including

- multi-stage
- mixed-integer

Textbooks:

J.R. Birge and F.V. Louveaux. *Introduction to Stochastic Programming*. Springer, 1997.

Yu. Ermoliev and R.J-B. Wets. *Numerical Techniques for Stochastic Optimization*. Springer, Berlin, 1988.

P. Kall and S.W. Wallace. *Stochastic Programming*. Wiley, 1994. Also available as PDF file, e.g. via `http://stoprog.org`.

A. Prékopa. *Stochastic Programming*. Kluwer, 1995.

A. Ruszczynski and A. Shapiro, editors. *Stochastic Programming*, vol. 10 of *Handbooks in OR and MS*. North-Holland, 2003.

References:

J.F. Benders. Partitioning procedures for solving mixed-variables programming problems. *Numerische Mathematik*, 4:238–252, 1962.

J.R. Birge. Decomposition and partitioning methods for multistage stochastic linear programs. *Oper. Res.*, 33(5):989–1007, 1985.

J.R. Birge. Stochastic programming computation and applications, INFORMS Journal on Computing 9(2): 111-133, 1997. (PDF via `http://stoprog.org`)

J.R. Birge and F.V. Louveaux. A multicut algorithm for two-stage stochastic linear programs. *European Journal of Operations Research*, 34:384–392, 1988.

G.B. Dantzig and G. Infanger. Large-scale stochastic linear programs—importance sampling and Benders decomposition. In C. Brezinski and U. Kulisch, editors, *Computational and applied mathematics, I (Dublin, 1991)*, pages 111–120. North-Holland, Amsterdam, 1992.

K. Frauendorfer. *Stochastic two-stage programming*, volume 392 of *Lecture Notes in Economics and Mathematical Systems*. Springer-Verlag, Berlin, 1992.

J.L. Higle and S. Sen. Stochastic decomposition: An algorithm for two-stage linear programs with recourse. *Math. Oper. Res.*, 16:650–669, 1991.

G. Infanger. Monte Carlo (importance) sampling within a Benders decomposition algorithm for stochastic linear programs. *Ann. Oper. Res.*, 39(1-4):69–95 (1993), 1992.

P. Kall and D. Stoyan. Solving stochastic programming problems with recourse including error bounds. *Math. Operationsforsch. Statist. Ser. Optim.*, 13(3):431–447, 1982.

J. T. Linderoth, A. Shapiro, and S. J. Wright. The empirical behavior of sampling methods for stochastic programming. Optimization Online, `http://www.optimization-online.org`, 2002.

J. Mayer. *Stochastic Linear Programming Algorithms: A Comparison Based on a Model Management System.* Gordon and Breach, Amsterdam, 1998.

M.V.F. Pereira and L.M.V.G Pinto. Stochastic optimization of a multireservoir hydroelectric system—a decomposition approach. *Water Resources Research*, 21(6):779–792, 1985.

A. Ruszczyński. A regularized decomposition method for minimizing a sum of polyhedral functions. *Math. Program.*, 35:309–333, 1986.

A. Shapiro. Statistical inference of multistage stochastic programming problems. Optimization Online, `http://www.optimization-online.org`, 2002.

R. Van Slyke and R.J-B. Wets. L-shaped linear programs with applications to control and stochastic programming. *SIAM Journal on Applied Mathematics*, 17:638–663, 1969.

S.W. Wallace. Solving stochastic programs with network recourse. *Networks*, 16(3):295–317, 1986.

R.J-B. Wets. Solving stochastic programs with simple recourse. *Stochastics*, 10:219–242, 1984.

Stochastic Programming Bibliography, `http://mally.eco.rug.nl/spbib.html`