Stochastic Programming:
From statistical data to optimal decisions

W. Römisch
Humboldt-University Berlin
Department of Mathematics

(K. Emich, H. Heitsch, A. Möller)

6th International Conference on **Computational Management Science**,
May 1-3, 2009, Geneva

Introduction

Practical optimization models often contain parameters of stochastic nature (e.g. statistical data available). In many cases it is not appropriate to replace them by some statistical estimate. Alternatives consist in modeling the random elements by a finite number of scenarios with given probabilities and incorporating them into the optimization model. Such stochastic programming models have the advantages:

- Solutions are robust with respect to changes of the data.
- The risk of decisions can be measured and managed.
- Simulation studies show that solutions of stochastic programs may be advantageous compared to deterministic ones.

Modeling

Assumptions: Information on the underlying probability distribution is available (e.g., statistical data) and the distribution does not depend on decisions.

Modeling questions: Are recourse actions available if stochasticity influences decisions ? Is the decision process based on recursive observations ?

- No recourse actions available: **Chance constraints**.
- Recourse actions available, but no recursive observations: **Two-stage stochastic programs** (possibly multi-period).
- Recursive observation and decision process: **Multi-stage stochastic programs**.

Integer variables should be incorporated if they are model-important.

Chance constraints

Let us consider the (linear) chance constrained model

$$\min\{\langle c,x\rangle : x\in X, P(\{\xi\in\Xi : T(\xi)x\geq h(\xi)\})\geq p\},$$

where $c\in\mathbb{R}^m$, X and Ξ are polyhedra in $\mathbb{R}^m$ and $\mathbb{R}^s$, respectively, $p\in(0,1)$, P is a probability measure on Ξ, i.e., $P\in\mathcal{P}(\Xi)$, and the right-hand side $h(\xi)\in\mathbb{R}^d$ and the (d,m)-matrix $T(\xi)$ are affine functions of ξ.

Challenges:
Although the sets $H(x)=\{\xi\in\Xi : T(\xi)x\geq h(\xi)\}$ are (convex) polyhedral subsets of Ξ, the function

$$x\to P(H(x))$$

is, in general, non-concave and non-differentiable on $\mathbb{R}^m$, hence, the optimization model is nonconvex. Concavity results are available for probability distributions satisfying certain concavity properties (e.g., normal distributions) (Prekopa 95, Henrion-Strugarek 08).

Two-stage stochastic programs

$$\min \left\{ \langle c, x \rangle + \int_\Xi \Phi(\xi; q(\xi), h(\xi) - T(\xi)x) P(d\xi) : x \in X \right\},$$

where

$$\Phi(\xi; u, t) := \inf\{\langle u, y \rangle : y \in Y, W(\xi)y = t\}$$

$P := \mathbb{P}\xi^{-1} \in \mathcal{P}_2(\Xi)$ is the probability distribution of the random vector ξ, $c \in \mathbb{R}^m$, $X \subseteq \mathbb{R}^m$ is a bounded polyhedron, $q(\xi) \in \mathbb{R}^{\overline{m}}$, $Y \in \mathbb{R}^{\overline{m}}$ is a polyhedral cone, $W(\xi)$ a $r \times \overline{m}$-matrix, $h(\xi) \in \mathbb{R}^r$ and $T(\xi)$ a $r \times m$-matrix. We assume that $q(\xi)$, $h(\xi)$, $W(\xi)$ and $T(\xi)$ are affine functions of ξ.

Theory and Algorithms: The function $\Phi : \Xi \times X \to \overline{\mathbb{R}}$ is well understood for fixed recourse (i.e., $W(\xi) \equiv W$) (Walkup-Wets 69). Convexity, optimality and duality results, decomposition methods, Monte-Carlo type methods, scenario reduction and stability analysis are well developed.

References: Ruszczyński-Shapiro 03, Kall-Mayer 05.

Mixed-integer two-stage stochastic programs

$$\min \left\{ \langle c, x \rangle + \int_{\Xi} \Phi(q(\xi), h(\xi) - T(\xi)x) P(d\xi) : x \in X \right\},$$

where Φ is given by

$$\Phi(u,t) := \inf \left\{ \langle u_1, y \rangle + \langle u_2, \bar{y} \rangle : Wy + \bar{W}\bar{y} \leq t, y \in \mathbb{Z}^{\hat{m}}, \bar{y} \in \mathbb{R}^{\bar{m}} \right\}$$

for all pairs $(u,t) \in \mathbb{R}^{\hat{m}+\bar{m}} \times \mathbb{R}^r$, and $c \in \mathbb{R}^m$, X is a closed subset of $\mathbb{R}^m$, Ξ a polyhedron in $\mathbb{R}^s$, W and $\bar{W}$ are $(r, \hat{m})$- and $(r, \bar{m})$-matrices, respectively, $q(\xi) \in \mathbb{R}^{\hat{m}+\bar{m}}$, $h(\xi) \in \mathbb{R}^r$, and the (r, m)-matrix $T(\xi)$ are affine functions of ξ, and $P \in \mathcal{P}_2(\Xi)$.

Theory and Algorithms: The function Φ is well understood (Blair-Jeroslow 77, Bank et al 82), nonconvex optimization models, structural analysis (Schultz 93), decomposition methods (surveys: Schultz 03, Sen 05), sampling methods, stability analysis, scenario reduction.

Multistage stochastic programs

Let $\{\xi_t\}_{t=1}^T$ be a discrete-time stochastic data process defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and with ξ_1 deterministic. The stochastic decision x_t at period t is assumed to be measurable with respect to $\mathcal{F}_t(\xi) := \sigma(\xi_1, \ldots, \xi_t)$ (nonanticipativity).

Multistage stochastic programming model:

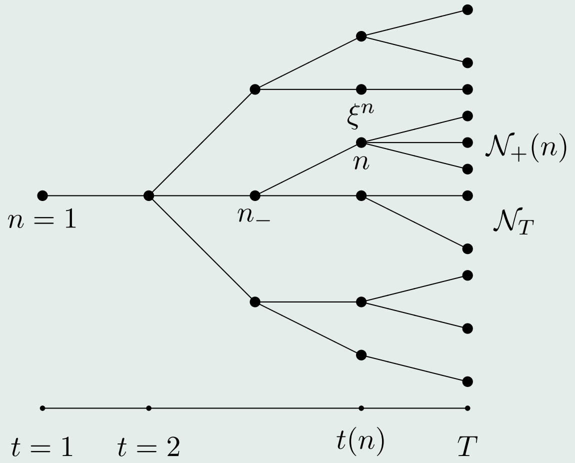
$$\min\left\{\mathbb{E}\left[\sum_{t=1}^{T}\langle b_t(\xi_t), x_t\rangle\right] \middle| \begin{array}{l} x_t \in X_t, x_t \text{ is } \mathcal{F}_t(\xi)\text{-measurable}, t = 1, .., T \\ A_{t,0}x_t + A_{t,1}(\xi_t)x_{t-1} = h_t(\xi_t), t = 2, .., T \end{array}\right\}$$

where X_t, $t = 1, \ldots, T$, are polyhedral, the vectors $b_t(\cdot)$, $h_t(\cdot)$ and $A_{t,1}(\cdot)$ are affine functions of ξ_t, where ξ varies in a polyhedral set Ξ.

If the process $\{\xi_t\}_{t=1}^T$ has a finite number of scenarios, they exhibit a scenario tree structure. If the measurability constraint is missing, the model is two-stage.

Data process approximation by scenario trees

The process $\{\xi_t\}_{t=1}^T$ is approximated by a process forming a scenario tree being based on a finite set $\mathcal{N} \subset \mathbb{N}$ of nodes.



Scenario tree with $T = 5$, $N = 22$ and 11 leaves

$n = 1$ root node, n_- unique predecessor of node n, $\text{path}(n) = \{1, \ldots, n_-, n\}$, $t(n) := |\text{path}(n)|$, $\mathcal{N}_+(n)$ set of successors to n, $\mathcal{N}_T := \{n \in \mathcal{N} : \mathcal{N}_+(n) = \emptyset\}$ set of leaves, $\text{path}(n)$, $n \in \mathcal{N}_T$, scenario with (given) probability π^n, $\pi^n := \sum_{\nu \in \mathcal{N}_+(n)} \pi^\nu$ probability of node n, ξ^n realization of $\xi_{t(n)}$.

Tree representation of the optimization model

$$\min\left\{\sum_{n\in\mathcal{N}}\pi^n\langle b_{t(n)}(\xi^n),x^n\rangle \,\middle|\, \begin{array}{l} x^n\in X_{t(n)},\, n\in\mathcal{N} \\ A_{t(n),0}x^n + A_{t(n),1}x^{n-} = h_{t(n)}(\xi^n),\, n\in\mathcal{N}\end{array}\right\}$$

The node-based optimization model may be solved by

- standard software (e.g., X-PRESS, CPLEX)

- decomposition methods for large scale models (Ruszczyński 03).

Mean-risk objective vs expectation:

The expectation objective may be replaced by convex (multi-period) risk functionals. If the risk functional is polyhedral, the linearity structure is maintained.

Scenario (tree) reduction and generation

Theoretical basis: Stability estimates

Scenario reduction: Developed for (mixed-integer) two-stage stochastic programs.

Scenario tree generation:

(i) Development of a stochastic model for the data process ξ (parametric [e.g. time series model], nonparametric [e.g. resampling from statistical data]) and generation of simulation scenarios;

(ii) Construction of a scenario tree out of the simulation scenarios by recursive scenario reduction and bundling over time such that the optimal expected revenue stays within a prescribed tolerance.

Implementation: GAMS-SCENRED

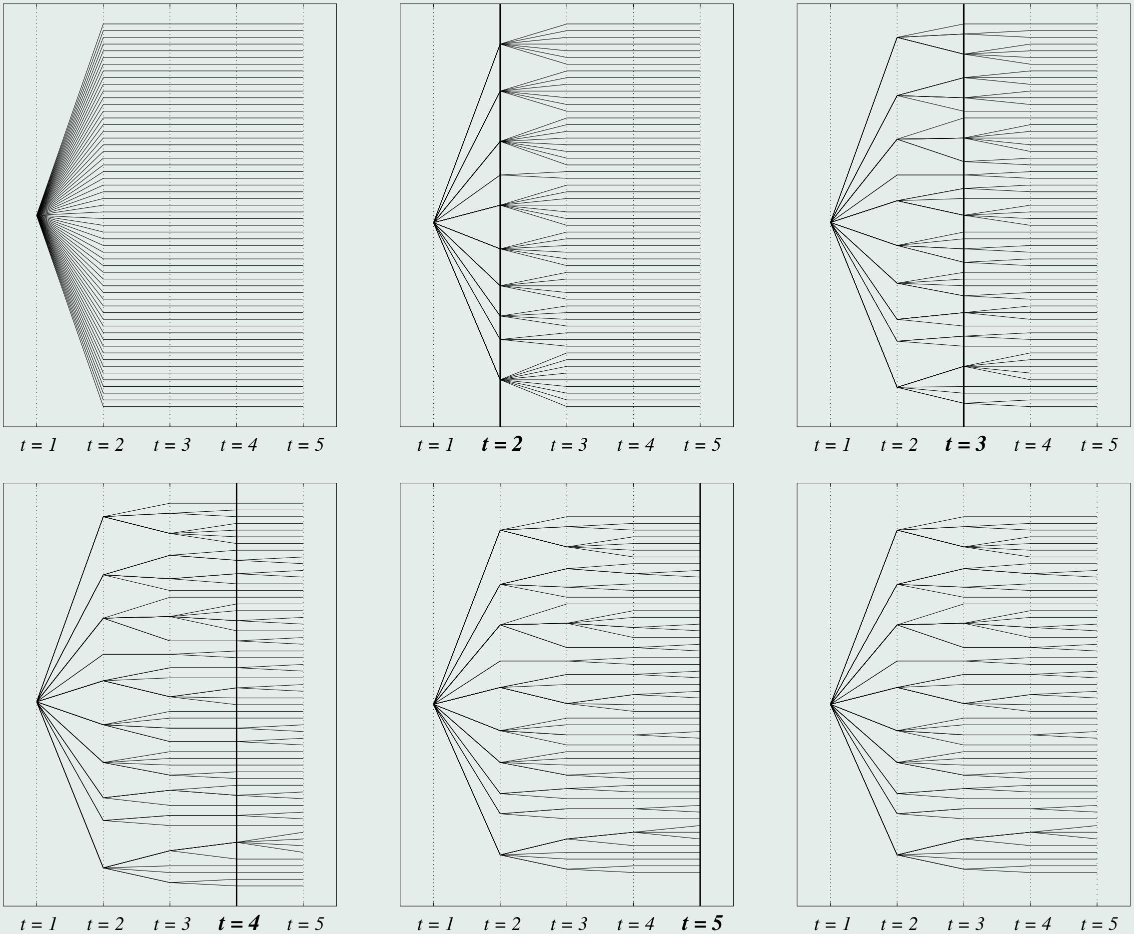


Illustration of the forward tree construction for an example including T=5 time periods starting with a scenario fan containing N=58 scenarios

<Start Animation>

Example: Airline network revenue management

Airline revenue management deals with strategies for controlling the booking process within a network of flights. Often statistical data is available for the (passenger) demand. The objective consists in maximizing the expected revenue. The booking process is controlled by seat protection levels or by (so-called) bid prices.

Aims:

- Stochastic programming model for airline network revenue management;
- Approximate representation of the multivariate booking demand processes by scenario trees generated from resampled historical demand scenarios;
- Lagrangian decomposition of the node-based stochastic integer program; algorithm design and numerical experience.

Notation

Input data

π^n: probability of node n;

stochastic (as scenario tree):

$d^n_{i,j,k}$: passenger demand;

$\gamma^n_{i,j,k}$: cancelation rates;

deterministic:

$f^b_{i,j,k,t(n)}$: fares;

$f^c_{i,j,k,t(n)}$: refunds;

$\mathcal{C}_{l,m}$: capacity;

Variables

$b^n_{i,j,k}$: bookings;

$c^n_{i,j,k}$: cancelations;

$B^n_{i,j,k}$: cumulative bookings;

$C^n_{i,j,k}$: cumulative cancelations;

$P^n_{i,j,k}$: protection level;

$z^{P,n}_{i,j,k}$, $z^{d,n}_{i,j,k}$: slack variables;

$\tilde{z}^n_{i,j,k}$: auxiliary integer variables;

Indices

$t = 0, \ldots, T$: data collection points;

$i = 1, \ldots, I$: origin-destination-itin.;

$j = 1, \ldots, J$: fare classes;

$k = 1, \ldots, K$: points of sale;

$l = 1, \ldots, L$: legs;

$\mathcal{I}_l$: index set of itineraries;

$m = 1, \ldots, M(l)$: compartments;

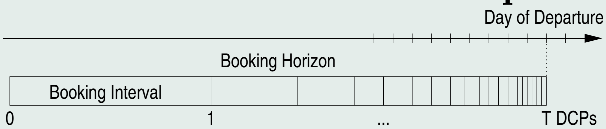
$\mathcal{J}_m(l)$: index set of fare classes;

$n = 0, \ldots, N$: nodes;

$t(n)$: time of node n;

n_-: preceding node of node n;

Time horizon and data collection points (dcp):



Airline network revenue management model (node representation)

Objective

$$\max_{(P^n_{i,j,k})} \left\{ \sum_{n=0}^{N} \pi^n \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \left[f^b_{i,j,k,t(n)} b^n_{i,j,k} - f^c_{i,j,k,t(n)} c^n_{i,j,k} \right] \right\}$$

Constraints

Cumulative bookings

$$B^0_{i,j,k} := \bar{B}^0_{i,j,k}; \qquad C^0_{i,j,k} := \bar{C}^0_{i,j,k}; \qquad B^n_{i,j,k} := B^{n-}_{i,j,k} + b^n_{i,j,k}$$

Cumulative cancelations

$$C^n_{i,j,k} = \lfloor \gamma^n_{i,j,k} B^n_{i,j,k} + 0.5 \rfloor$$

Cancelations

$$c^n_{i,j,k} = C^n_{i,j,k} - C^{n-}_{i,j,k}$$

Passenger demands and protection levels

$$b^n_{i,j,k} \le d^n_{i,j,k}; \quad b^n_{i,j,k} \le P^{n-}_{i,j,k} - B^{n-}_{i,j,k} + C^n_{i,j,k} \quad \text{(disjunctive constraints)}$$

Leg capacity limits

$$\sum_{i \in \mathcal{I}_l} \sum_{j \in \mathcal{J}_m(l)} \sum_{k=1}^{K} P^n_{i,j,k} \le \mathcal{C}_{l,m} \qquad (n \in \mathcal{N}_{T-1})$$

Integrality and nonnegativity constraints

$$B^n_{i,j,k}, C^n_{i,j,k}, P^n_{i,j,k} \in \mathbb{Z}; \quad b^n_{i,j,k} \ge 0; \quad c^n_{i,j,k} \ge 0$$

Airline network revenue management model (final)

Objective

$$\max_{(P^n_{i,j,k})} \left\{ \sum_{n=0}^{N} \pi^n \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \left[f^b_{i,j,k,t(n)} b^n_{i,j,k} - f^c_{i,j,k,t(n)} c^n_{i,j,k} \right] \right\}$$

Constraints

Cumulative bookings

$$B^0_{i,j,k} := \bar{B}^0_{i,j,k}; \qquad C^0_{i,j,k} := \bar{C}^0_{i,j,k}; \qquad B^n_{i,j,k} := B^{n-}_{i,j,k} + b^n_{i,j,k}$$

Cumulative cancelations

$$C^n_{i,j,k} = \lfloor \gamma^n_{i,j,k} B^n_{i,j,k} + 0.5 \rfloor$$

Cancelations

$$c^n_{i,j,k} = C^n_{i,j,k} - C^{n-}_{i,j,k}$$

Passenger demands

$$b^n_{i,j,k} + z^{b,n}_{i,j,k} = d^n_{i,j,k}$$

Protection levels

$$B^n_{i,j,k} - C^n_{i,j,k} + z^{P,n}_{i,j,k} = P^{n-}_{i,j,k}$$

Number of bookings (disjunctive constraints) ($\kappa > 0$, adequately large)

$$0 \le z^{b,n}_{i,j,k} \le (1 - \tilde{z}^n_{i,j,k}) d^n_{i,j,k} \quad 0 \le z^{P,n}_{i,j,k} \le \tilde{z}^n_{i,j,k} \kappa \quad \tilde{z}^n_{i,j,k} \in \{0, 1\}$$

Leg capacity limits

$$\sum_{i \in \mathcal{I}_l} \sum_{j \in \mathcal{J}_m(l)} \sum_{k=1}^{K} P^n_{i,j,k} \le \mathcal{C}_{l,m} \qquad (n \in \mathcal{N}_{T-1})$$

Integrality and nonnegativity constraints

$$B^n_{i,j,k}, C^n_{i,j,k}, P^n_{i,j,k} \in \mathbb{Z}; \quad b^n_{i,j,k} \ge 0; \quad c^n_{i,j,k} \ge 0$$

Comments:

- large scale structured integer linear program
- solvable by a standard solver (e.g. CPLEX) in reasonable time for smaller networks when neglecting integer constraints
- **Dimensions:** (S number of scenarios)
  - $4IJKN$ continuous variables,
  - $IJK(N+1-S)+2IJKN$ integer variables,
  - $IJKN$ binary variables
  - $7IJK(N-1)+\sum_{n\in\mathcal{N}_{T-1}}\sum_{l=1}^{L} M(l)$ constraints
- Protection levels $(P_{i,j,k}^{n})_{n\in\mathcal{N}}$ have the same tree structure as the input data
- The (deterministic) protection levels of the first stage may be taken as a basis for the computer reservation system
- At the next dcp a new scenario tree has to be generated and the problem is resolved etc.

Lagrangian decomposition

Idea: Dualization of leg capacity limits

Lagrangian function Λ:

$$
\begin{aligned}
\Lambda(\lambda, P) \;:=\; & \sum_{n=0}^{N} \pi^n \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} \left(f_{i,j,k}^{b,n} b_{i,j,k}^n - f_{i,j,k}^{c,n} c_{i,j,k}^n \right) \\
& + \sum_{n \in \mathcal{N}_{T-1}} \pi^n \sum_{l=1}^{L} \sum_{m=1}^{M(l)} \lambda_{l,m}^n \left(\sum_{i \in \mathcal{I}_l} \sum_{j \in \mathcal{J}_m(l)} \sum_{k=1}^{K} \mathcal{C}_{l,m} - P_{i,j,k}^n \right) \\
= \; & \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} \left(\sum_{n=0}^{N} \pi^n \left(f_{i,j,k}^{b,n} b_{i,j,k}^n - f_{i,j,k}^{c,n} c_{i,j,k}^n \right) \right. \\
& \left. - \sum_{n \in \mathcal{N}_{T-1}} \pi^n \sum_{l \in \mathcal{L}_i} \sum_{m=1}^{M(l)} \delta_{j,l,m} \lambda_{l,m}^n P_{i,j,k}^n \right) + \sum_{n \in \mathcal{N}_{T-1}} \pi^n \sum_{l=1}^{L} \sum_{m=1}^{M(l)} \lambda_{l,m}^n \mathcal{C}_{l,m} \\
= \; & \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} \Lambda_{i,j,k}(\lambda, P_{i,j,k}) + \sum_{n \in \mathcal{N}_{T-1}} \pi^n \sum_{l=1}^{L} \sum_{m=1}^{M(l)} \lambda_{l,m}^n \mathcal{C}_{l,m}
\end{aligned}
$$

where $\mathcal{L}_i = \{l : i \in \mathcal{I}_l\}$ and $\delta_{j,l,m} = \begin{cases} 1 & j \in \mathcal{J}_m(l) \\ 0 & \text{otherwise} \end{cases}$

Dual function D:

$$\begin{aligned} D(\lambda) &= \sup_P \Lambda(\lambda, P) \\ &= \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} \sup_{P_{i,j,k}} \Lambda_{i,j,k}(\lambda, P_{i,j,k}) + \sum_{n\in\mathcal{N}_{T-1}} \pi^n \sum_{l=1}^{L}\sum_{m=1}^{M(l)} \lambda^n_{l,m}\mathcal{C}_{l,m} \end{aligned}$$

The function D is convex nondifferentiable and decomposable.

Dual problem:

$$\inf_\lambda D(\lambda)$$

The relative duality gap is small (theory by Bertsekas 82).

Subgradients:

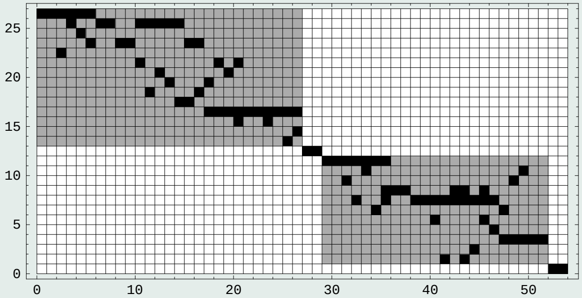
$$[\partial D(\lambda)]^n_{l,m} = \pi^n \left(\mathcal{C}_{l,m} - \sum_{i\in\mathcal{I}_l} \sum_{j\in\mathcal{J}_m(l)} \sum_{k=1}^{K} P^n_{i,j,k} \right)$$

The Lagrange multipliers $\lambda^n_{l,m}$, $n \in \mathcal{N}_t$, may be interpreted as bid prices at t for leg l and compartment m. However, they are presently only available for $n \in \mathcal{N}_{T-1}$.

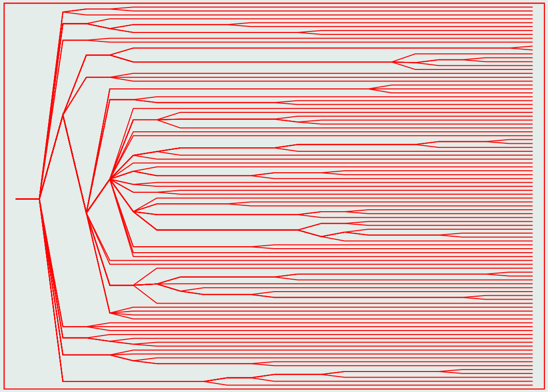
Dual solution algorithm

- Solution of the dual problem by a bundle subgradient method (e.g. proximal bundle method by Kiwiel or Helmberg)
- Solution of the subproblems by dynamic programming on scenario trees.
- Primal-proximal heuristic to determine a good primal feasible solution (e.g. by Daniilidis and Lemaréchal).

A realistic mid-size airline network example



ODI-Leg-Matrix



Scenario tree

RM problem dimensions

#ODIs	54
#ODI-Fareclass-POS	489
#Legs	27
#Leg-Compartments	54
#DCPs	23

Tree and Size

#Scenarios	98
#Nodes	1.441
#Variables	3.473.367
#Constraints	2.774.445
#Coupling Constr.	5.238

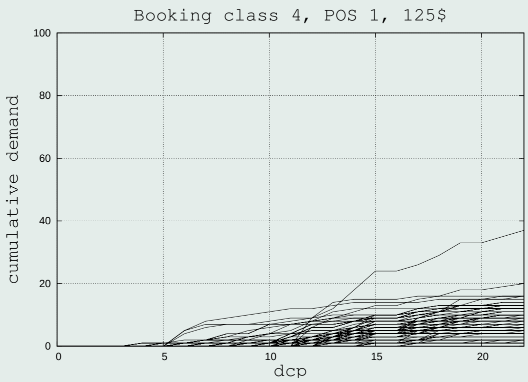
Numerical results

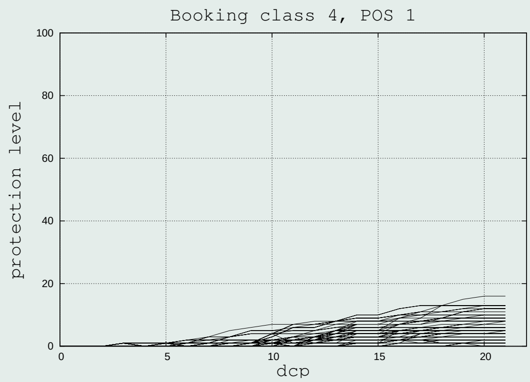
Bundle methods

Dual value	179349.78
Dimension	5238
max bundle size	10
#Iterations	46
#DP	22494
time	09:05:55.36
time in DP	1:23.39

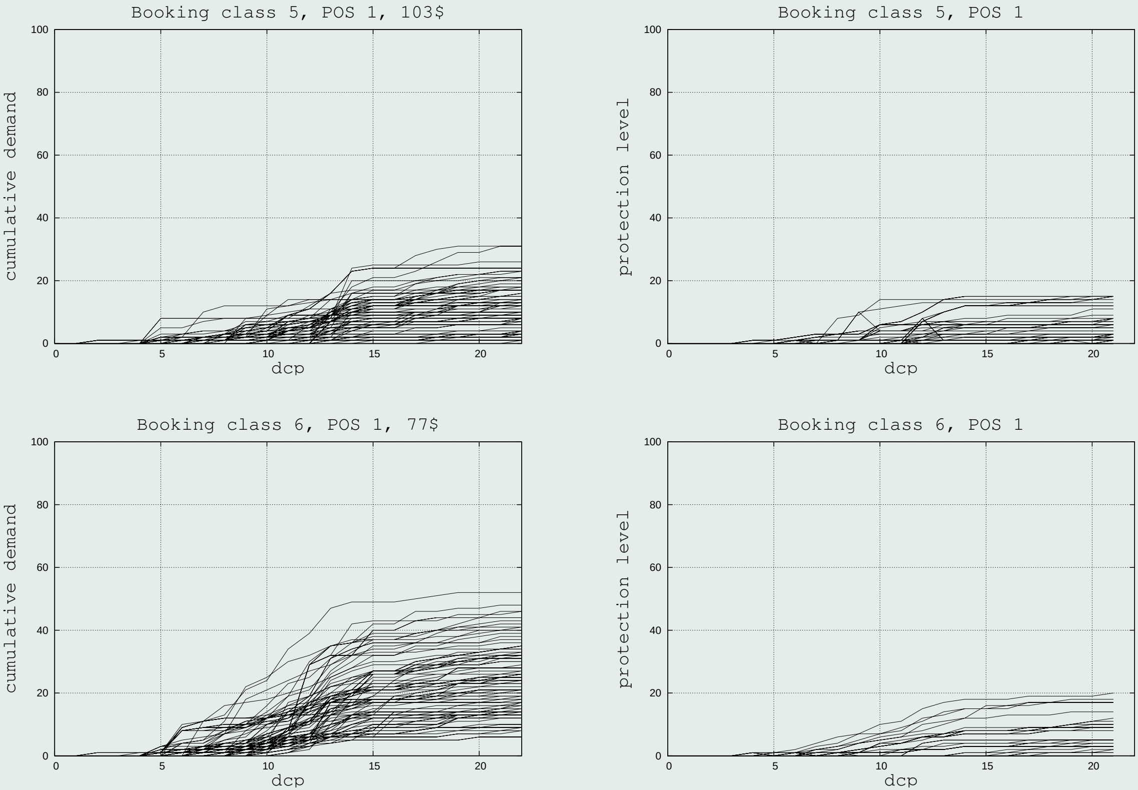
Lagrange heuristic

Primal value	179134.76
Duality gap	0.001
time	5:33.87





Cumulative demand and protection level of booking class 3 in the economy compartment of ODI 9



Cumulative demands and protection levels of booking class 4 and 5 in the economy compartment of ODI 9

Conclusions and future work

We presented an approach to airline network revenue management using a scenario tree-based dynamic stochastic optimization model. The approach

- starts from a finite number of demand scenarios and their probabilities,
- requires no assumptions on the demand distributions except their decision-independence.

Stochastic programming approaches lead to solutions that are more robust with respect to perturbations of input data. However, the models have higher complexity.

Future work:

- Implementation refinements of the decomposition scheme

(URL: `www.math.hu-berlin.de/~romisch`, Email: `romisch@math.hu-berlin.de`)

References

A. Daniilidis, C. Lemaréchal: On a primal-proximal heuristic in discrete optimization, *Mathematical Programming* 104 (2005), 105–128.

H. Heitsch, W. Römisch, C. Strugarek: Stability of multistage stochastic programs, *SIAM Journal Optimization* 17 (2006), 511–525.

H. Heitsch, W. Römisch: A note on scenario reduction for two-stage stochastic programs, *Operations Research Letters* 35 (2007), 731–738.

H. Heitsch, W. Römisch: Scenario tree modeling for multistage stochastic programs, *Mathematical Programming* 118 (2009), 371–406.

H. Heitsch, W. Römisch: Scenario tree reduction for multistage stochastic programs, *Computational Management Science* 6 (2009), 117–133.

A. Möller, W. Römisch, K. Weber: Airline network revenue management by multistage stochastic programming, *Computational Management Science* 5 (2008), 355–377.

W. Römisch: Stability of stochastic programming problems, in *Stochastic Programming* (Ruszczyński, A., Shapiro, A., eds.), Elsevier, Amsterdam, 2003, 483–554.

R. Schultz: Stochastic programming with integer variables, *Mathematical Programming* 97 (2003), 285–309.

S. Sen: Algorithms for stochastic mixed-integer programming models, in *Discrete Optimization* (K. Aardal, G. L. Nemhauser, R. Weissmantel eds.), Handbooks in Operations Research and Management Science Vol. 12, Elsevier, Amsterdam 2005, 515–558.