Approximations of Stochastic Programs, Scenario Tree Reduction and Construction

W. Römisch

Humboldt-Universität Berlin
Institut für Mathematik
10099 Berlin, Germany

www.math.hu-berlin.de/~romisch

(J. Dupačová, N. Gröwe-Kuska, H. Heitsch)

GAMS Workshop, Heidelberg, Sept. 1-3, 2003

Supported by the DFG (SPP and FZT 86) and the BMBF

Introduction

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

Typical financial and production planning model:

$$\min\{\mathbb{E}[\sum_{t=1}^{T} c_t(\xi_t, x_t)] \quad : \quad x_t \in X_t \,,\, x_t \text{ nonanticipative}, \\ A_{tt}(\xi_t)x_t + A_{t,t-1}(\xi_t)x_{t-1} \geq g_t(\xi_t)\}$$

Alternative for the minimization of expected costs:

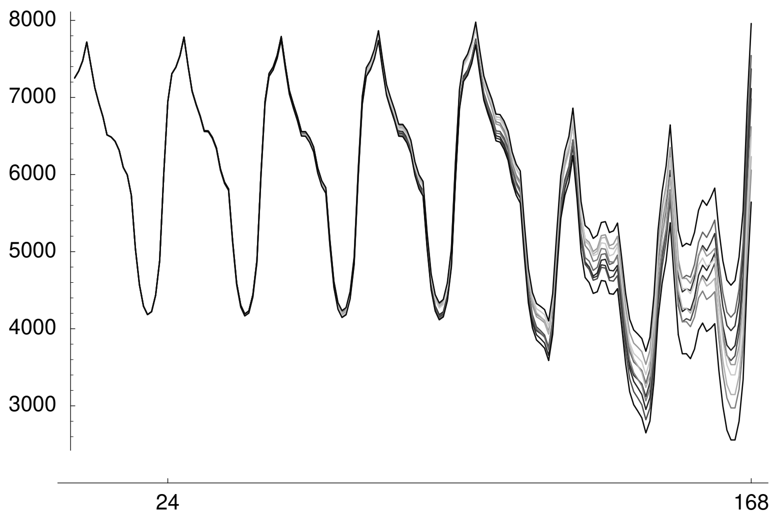
Minimizing some risk measure $\mathbb{F}$ of the stochastic cost process $\{c_t(\xi_t, x_t)\}_{t=1}^T$ (risk management).

First step of its numerical solution:
Approximation of $\{\xi_t\}_{t=1}^T$ by finitely many scenarios with certain probabilities. Nonanticipativity leads to a scenario tree structure of the approximation.

Scenario tree approximations

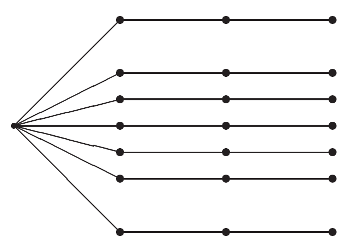
(a) Simulation of (sufficiently many) scenarios of the stochastic data process ξ;
(b) construction of scenario trees from simulation scenarios or probability distribution information;
(c) (optional) follow-up treatment of the scenario tree.

(a) Methods:
- Calibrating statistical models to historical data.
- direct use of “comparable” historical data as scenarios.



Scenarios for the weekly electrical load

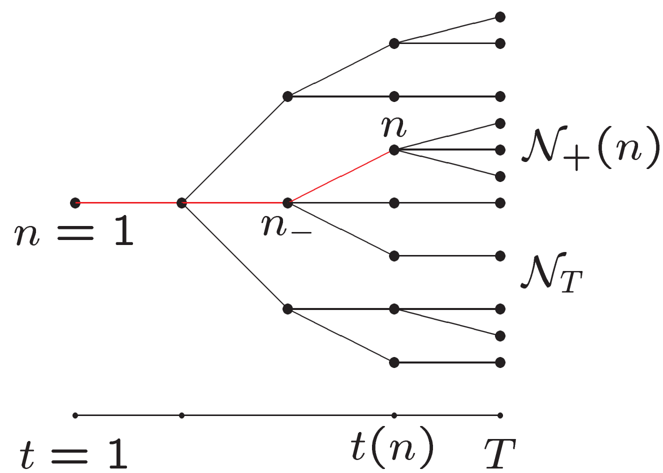
(b) Constructions based on simulation scenarios:
<u>Given:</u> S data scenarios with fixed starting point ξ_1, i.e., the scenarios form a fan.



Cluster-analysis-based methods: “Bundling” scenarios in a cluster and definition of succesors and predecessor, e.g., using distances of probability distributions.

(c) Tree reduction using probability metrics.

A scenario tree is based on a finite set $\mathcal{N} \subset \mathbb{N}$ of nodes.



Scenario tree with $T = 5$, $|\mathcal{N}| = 23$ and 11 leaves

$n = 1$ stands for the period root node,
n_- is the unique predecessor of node n,
path(n):=$\{1, \ldots, n_-, n\}$, $t(n) := |\text{path}(n)|$,
$\mathcal{N}_t := \{n : t(n) = t\}$, nodes $n \in \mathcal{N}_T$ are the leaves,
Scenario: path(n) for some $n \in \mathcal{N}_T$,
$\mathcal{N}_+(n)$ is the set of successors to node n,
$\{\pi_n\}_{n \in \mathcal{N}_T}$ are the scenario probabilities and
$\pi_n := \sum_{n_+ \in \mathcal{N}_+(n)} \pi_{n_+}$, $n \in \mathcal{N}$.
$\{\xi^n\}_{n \in \mathcal{N}_t}$ are the realizations of ξ_t,
$\{x^n\}_{n \in \mathcal{N}_t}$ the realizations of x_t.

Scenario tree formulation of the model:

$$\min \quad \Big\{ \sum_{n \in \mathcal{N}} \pi_n c_{t(n)}(\xi^n, x^n) : x^n \in X_{t(n)},$$
$$A_{t(n),t(n)}(\xi^n) x^n + A_{t(n),t(n_-)}(\xi^n) x^{n_-} \geq g_{t(n)}(\xi^n)\,,\ n \in \mathcal{N} \Big\}$$

Specially structured programs !

Solving stochastic programs

First idea: Use of standard software for solving the stochastic program in scenario tree form !
But: Models are huge even for small trees and, in addition, special structures are often not exploited !

$\Rightarrow$ Decomposition is the only feasible alternative in many (practical) situations.

Direct or primal decomposition approaches:
- starting point: Benders decomposition based on both *feasibility* and *objective* cuts;
- variants: *regularization*, *nesting*, *stochastic cuts*.

Dual decomposition approaches:
(i) Scenario decomposition by Lagrangian dualization of nonanticipativity constraints (solving the dual by bundle subgradient methods, augmented Lagrangian decomposition, variable or operator splitting methods);
(ii) nodal decomposition by dualizing dynamic constraints;
(iii) geographical decomposition by Lagrangian relaxation of coupling constraints.

Presently, nested Benders decomposition, stochastic decomposition and scenario decomposition (based on augmented Lagrangians and on operator splitting) are mostly used.

Distances of probability distributions

Let P denote the probability distribution of the stochastic data process $\{\xi_t\}_{t=1}^T$, where ξ_t has dimension r, i.e., P is a probability measure on $\Xi \subseteq \mathbb{R}^{rT} = \mathbb{R}^s$. Let c be a non-negative symmetric continuous function on $\Xi \times \Xi$, which plays the role of a *(normalized) global continuity modulus of the stochastic costs* f_0 as a function of the first-stage decision x and of ξ. This means: If the deterministic equivalent of the stochastic program takes the form

$$\min\{\int_\Xi f_0(x,\xi)P(d\xi) : x \in X\},$$

we choose c such that the property

$$|f_0(x,\xi) - f_0(x,\tilde{\xi})| \leq L(\|x\|)c(\xi,\tilde{\xi}),\ \forall \xi,\tilde{\xi} \in \Xi, x \in X,$$

holds with some constant $L(\|x\|)$.

Example:
Linear multiperiod two-stage model with fixed recourse in each period: $c(\xi,\tilde{\xi}) := \max\{1, \|\xi\|^{T-1}, \|\tilde{\xi}\|^{T-1}\}\|\xi - \tilde{\xi}\|$ (Römisch/Wets 03 forthcoming).

We consider the Fortet-Mourier metric of two measures P and Q on Ξ

$$\zeta_c(P,Q) = \sup\{|\int_\Xi f(\xi)(P-Q)(d\xi)| : f \in \mathcal{F}_c\},$$

where the class $\mathcal{F}_c$ is defined by

$$\mathcal{F}_c := \{f : \Xi \to \mathbb{R} : f(\xi) - f(\tilde{\xi}) \leq c(\xi,\tilde{\xi}), \forall \xi,\tilde{\xi} \in \Xi\},$$

its dual, the Kantorovich-Rubinstein functional

$$\hat{\mu}_c(P,Q) = \inf\{\int_{\Xi\times\Xi} c(\xi,\tilde{\xi})\eta(d\xi,d\tilde{\xi}) : \pi_1\eta - \pi_2\eta = P - Q\}.$$

and its upper bound, the Kantorovich functional or transportation metric

$$\mu_c(P,Q) := \inf\{\int_{\Xi\times\Xi} c(\xi,\tilde{\xi})\eta(d\xi,d\tilde{\xi}) : \pi_1\eta = P,\ \pi_2\eta = Q\}.$$

Theorem: (Stability)
Under weak conditions on the stochastic program, the optimal values are Lipschitz continuous and the solution sets are upper semicontinuous w.r.t. ζ_c ($\hat{\mu}_c$, μ_c).
(Rachev/Römisch 02, Römisch 03)

ζ_c and $\hat{\mu}_c$, μ_c are defined on sets of probability measures $\mathcal{P}_c(\mathbb{R}^s)$ satisfying a certain moment condition w.r.t. c.

Example:
If $c(\xi,\tilde{\xi}) := \max\{1, \|\xi\|^{p-1}, \|\tilde{\xi}\|^{p-1}\}\|\xi - \tilde{\xi}\|$, then
$\mathcal{P}_c(\Xi) = \{Q \in \mathcal{P}(\Xi) : \int_\Xi \|\xi\|^p Q(d\xi) < \infty\} \quad (p \geq 1)$.

(References: Rachev 91, Rachev/Rüschendorf 98)

Approach:
Select a probability metric d such that the stochastic program is stable w.r.t. d.
Given P and a tolerance $\varepsilon > 0$, determine a scenario tree such that its probability distribution P_{tr} has the property

$$d(P, P_{tr}) \leq \varepsilon\,.$$

Probability distances of discrete distributions

P: scenarios ξ_i with probabilities p_i, $i = 1, \ldots, N$,
Q: scenarios $\tilde{\xi}_j$ with probabilities q_j, $j = 1, \ldots, M$.

Then it holds

$$\mu_c(P, Q) = \sup\{\sum_{i=1}^{N} p_i u_i + \sum_{j=1}^{M} q_j v_j : u_i + v_j \leq c(\xi_i, \tilde{\xi}_j)\ \forall i, j\}$$

$$= \inf\{\sum_{i,j} \eta_{ij} c(\xi_i, \tilde{\xi}_j) : \eta_{ij} \geq 0, \sum_j \eta_{ij} = p_i, \sum_i \eta_{ij} = q_j\}$$

(*linear transportation problem*)

$$\hat{\mu}_c(P, Q) = \inf\{\sum_{i,j} \eta_{ij} c(\xi_i, \tilde{\xi}_j) : \eta_{ij} \geq 0, \sum_i \eta_{ij} - \sum_j \eta_{ij} = q_j - p_i\}$$

$$= \zeta_c(P, Q)$$

(can be reformulated as a *minimum cost flow problem*)

Special case: Scenario reduction

$$\{\tilde{\xi}_1, \ldots, \tilde{\xi}_M\} \subseteq \{\xi_1, \ldots, \xi_N\} \quad \text{w.l.o.g.} \quad M = N\,.$$

Scenario Reduction

We consider the transportation metric μ_c on $\mathcal{P}(\Xi)$ where $c : \Xi \times \Xi \to I\!R_+$ is adapted to the stochastic program as described above.

Let $P = \sum_{i=1}^{N} p_i \delta_{\xi_i}$ and $Q = \sum_{j=1}^{N} q_j \delta_{\xi_j}$.

Theorem: (optimal reduction of a scenario set J)

$$D_J := \min_{q_j \geq 0,\, \sum_j q_j = 1} \mu_c\Big(\sum_{i=1}^{N} p_i \delta_{\xi_i}, \sum_{j \notin J} q_j \delta_{\xi_j}\Big) = \sum_{i \in J} p_i \min_{j \notin J} c(\xi_i, \xi_j)$$

The minimum is attained at $\bar{q}_j = p_j + \sum_{i \in J_j} p_i$, $\forall j \notin J$, where $J_j := \{i \in J : j = j(i)\}$ and $j(i) \in \arg\min_{j \notin J} c(\xi_i, \xi_j)$, $\forall i \in J$.
(optimal redistribution)

Optimal reduction of a scenario set with fixed cardinality:

$$\min\{D_J = \sum_{i \in J} p_i \min_{j \notin J} c(\xi_i, \xi_j) \;:\; J \subset \{1, ..., N\}\,,\, \# J = N - n\}$$

(combinatorial optimization problem of set-covering type)

Theory: Dupačová/Gröwe-Kuska/Römisch 03
Fast heuristics (forward, backward): Heitsch/Römisch 03
Implementation: GAMS/SCENRED (Gröwe-Kuska)

Fast heuristics

Algorithm 1: (Simultaneous backward reduction)

Step [0]: Sorting of $\{c(\xi_j, \xi_k) : \forall j\}, \forall k$,

$J^{[0]} := \emptyset$.

Step [i]: $l_i \in \arg\min_{l \notin J^{[i-1]}} \sum_{k \in J^{[i-1]} \cup \{l\}} p_k \min_{j \notin J^{[i-1]} \cup \{l\}} c(\xi_k, \xi_j)$.

$J^{[i]} := J^{[i-1]} \cup \{l_i\}$.

Step [N-n+1]: Optimal redistribution.

Algorithm 2: (Fast forward selection)

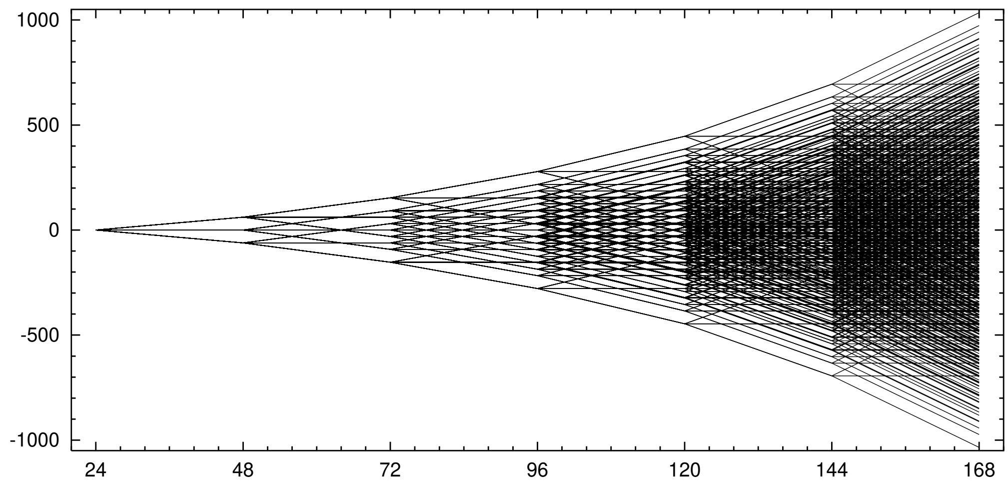
Step [0]: Compute $c(\xi_k, \xi_u)$, $k, u = 1, \ldots, N$,

$J^{[0]} := \{1, \ldots, N\}$.

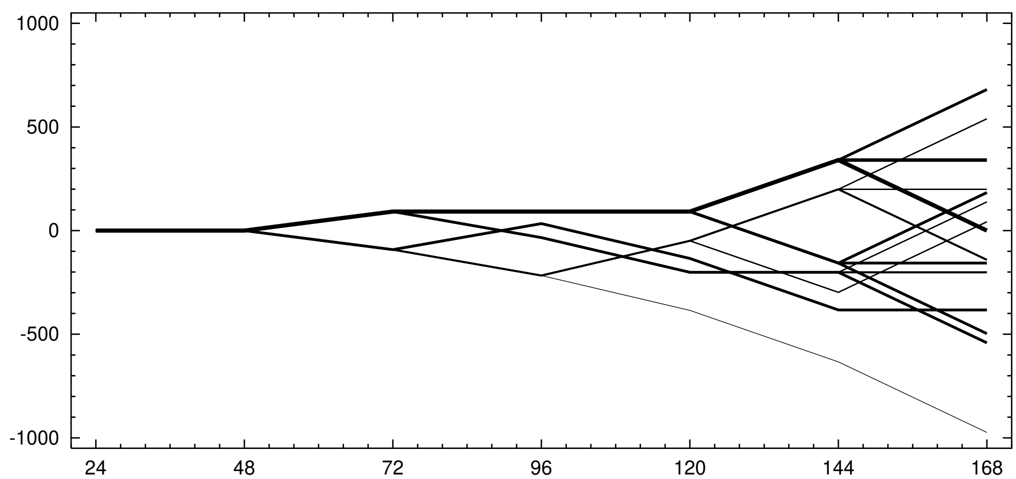
Step [i]: $u_i \in \arg\min_{u \in J^{[i-1]}} \sum_{k \in J^{[i-1]} \setminus \{u\}} p_k \min_{j \notin J^{[i-1]} \setminus \{u\}} c(\xi_k, \xi_j)$,

$J^{[i]} := J^{[i-1]} \setminus \{u_i\}$.

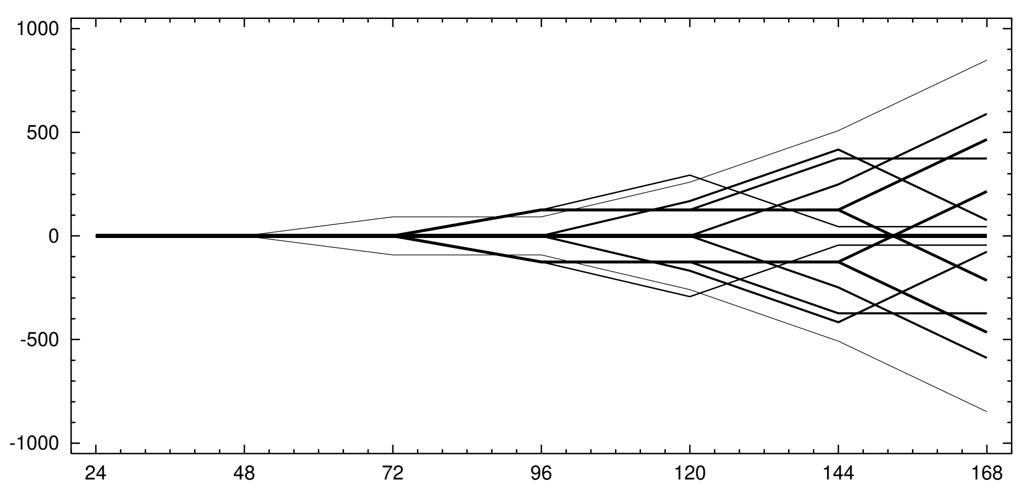
Step [n+1]: Optimal redistribution.



Original load scenario tree



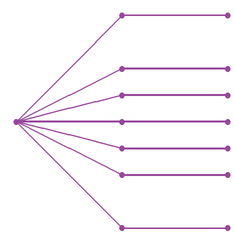
Reduced load scenario tree / backward



Reduced load scenario tree / forward

Constructing scenario trees from data scenarios

Let a fan of data scenarios $\xi^i = (\xi_1^i, \ldots, \xi_T^i)$ with probabilities π^i, $i = 1, \ldots, N$, be given, i.e., all scenarios coincide at the starting point $t = 1$, i.e., $\xi_1^1 = \ldots = \xi_1^N =: \xi_1^*$. Hence, it has the form

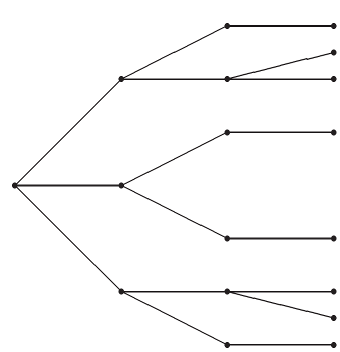


$t = 1$ may be regarded as the root node of a scenario tree consisting of N branches (leaves, scenarios).

Representation of scenario trees:
One may use scenario partition matrices M (indicating which scenario concides with which at some time t) or node partition matrices N (indicating which nodes are predecessors and successors of a node n, respectively).

Example:

$$M = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 2 & 2 \\ 1 & 1 & 2 & 3 \\ 1 & 4 & 4 & 4 \\ 1 & 4 & 5 & 5 \\ 1 & 6 & 6 & 6 \\ 1 & 6 & 7 & 7 \\ 1 & 6 & 7 & 8 \end{pmatrix}$$



$$N = \begin{pmatrix} 1 & 2 & 5 & 11 \\ 1 & 2 & 6 & 12 \\ 1 & 2 & 6 & 13 \\ 1 & 3 & 7 & 14 \\ 1 & 3 & 8 & 15 \\ 1 & 4 & 9 & 16 \\ 1 & 4 & 9 & 17 \\ 1 & 4 & 10 & 18 \end{pmatrix}$$

Let P denote the probability distribution of $\xi = \{\xi_t\}_{t=1}^T$ given by the fan of scenarios. P plays the role of the original distribution. Let the function c be adapted to the underlying stochastic program containing P.

We describe an **algorithm** that produces, for each $\varepsilon > 0$, a scenario tree with root node ξ_1^* and less nodes than that of P, such that for its probability distribution P_ε we obtain for the transportation metric μ_c:

$$\mu_c(P, P_\varepsilon) < \varepsilon.$$

Algorithm: (backward variant)
Let $\varepsilon_t > 0$, $t = 1, \ldots, T$, be given such that $\sum_{t=1}^T \varepsilon_t \leq \varepsilon$, set $t := T$, $I_{T+1} := \{1, \ldots, N\}$, $\pi_{T+1}^i := \pi^i$ and $P_{T+1} := P$.

For $t = T, \ldots, 2$:
Step t: Determine an index set $I_t \subseteq I_{t+1}$ such that

$$\mu_{c_t}(P_t, P_{t+1}) < \varepsilon_t \,,$$

where $\{\xi^i\}_{i \in I_t}$ is the support of P_t and c_t is defined by $c_t(\xi, \tilde{\xi}) := c((\xi_1, \ldots, \xi_t, 0, \ldots, 0), (\tilde{\xi}_1, \ldots, \tilde{\xi}_t, 0, \ldots, 0))$;

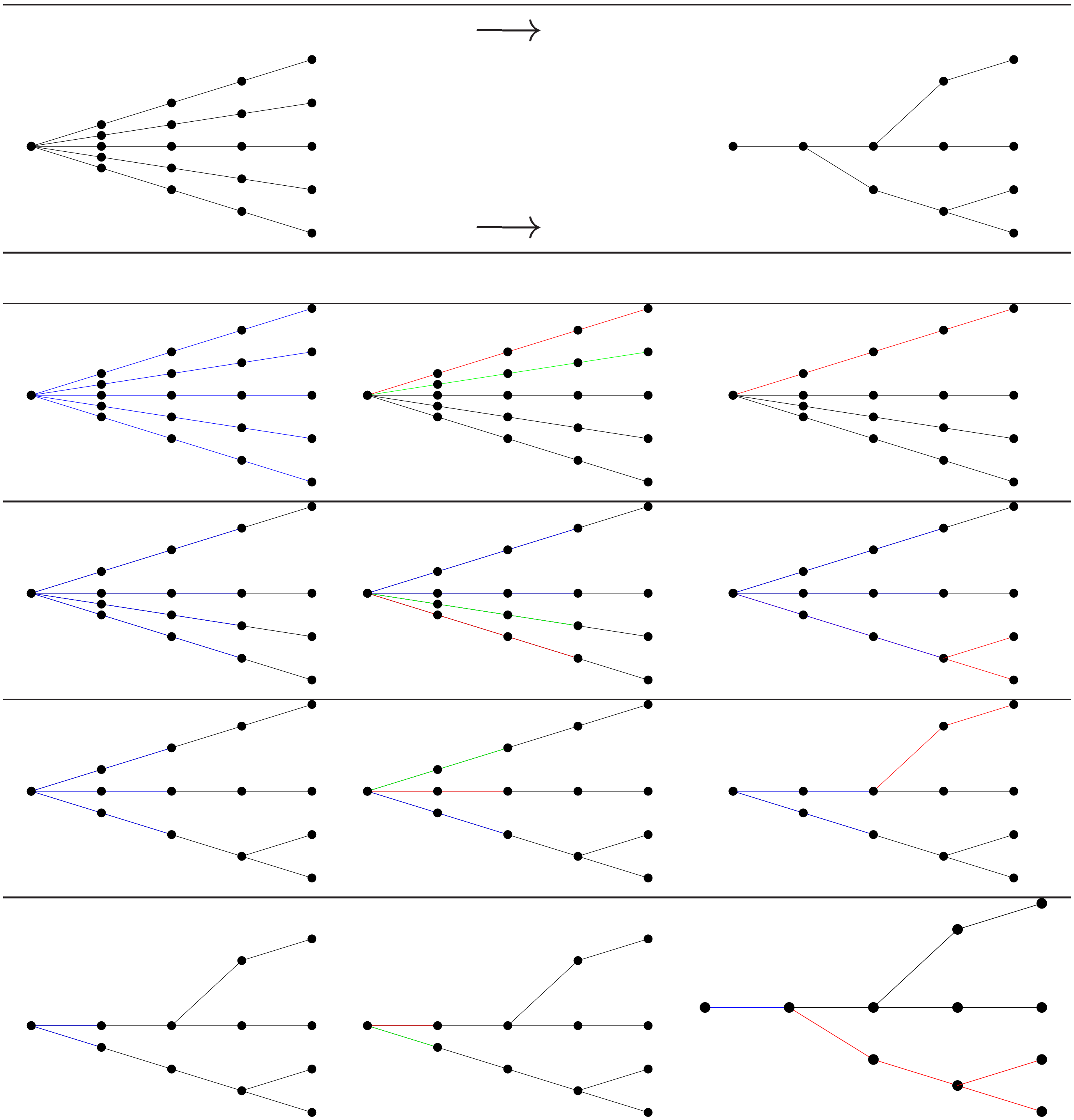
(scenario reduction w.r.t. {1,. . . ,t})

Step 1: Determine a probability measure P_ε such that its marginal distributions $P_\varepsilon \Pi_t^{-1}$ are $\delta_{\xi_1^*}$ for $t = 1$ and

$$P_\varepsilon \Pi_t^{-1} = \sum_{i \in I_t} \pi_t^i \delta_{\xi_t^i} \qquad \text{and} \qquad \pi_t^i := \pi_{t+1}^i + \sum_{j \in J_{t,i}} \pi_{t+1}^j \,,$$

where $J_{t,i} := \{j \in I_{t+1} \setminus I_t : i_t(j) = i\}$, $i_t(j) \in \arg\min_{i \in I_t} c_t(\xi^j, \xi^i)\}$ are the index sets according to the redistribution rule.

Scenario tree construction algorithm



Blue refers to computing of c-distances of scenarios, green to deleting and adding its weight to the red scenario.

Application:

ξ is the multivariate data process having the components

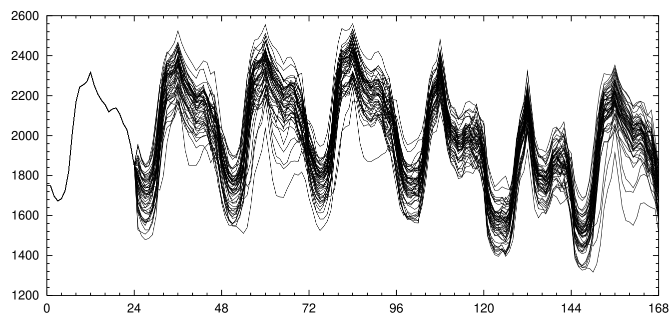
a) electrical load,

b) electricity prices for baseload contracts (at EEX),

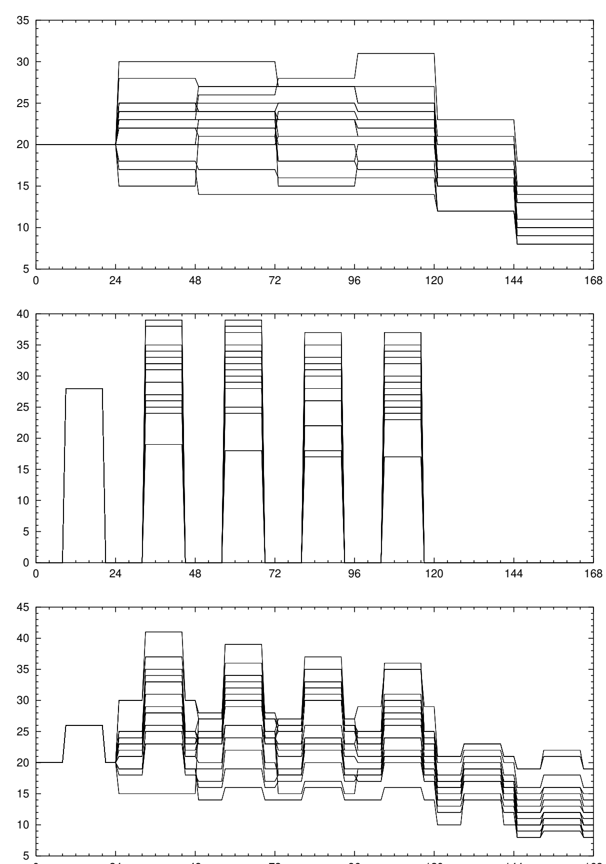
c) electricity prices for peakload contracts (at EEX),

d) electricity prices for individual hours (at EEX).

Data scenarios obtained from a stochastic model calibrated to the historical load data of the German power utility VEAG and historical price data of the European Energy Exchange (EEX) at Frankfurt/Leipzig.



a) Scenario tree for the electrical load



b), c), d) Scenario trees for prices of baseload contracts, peakload contracts and individual hours, respectively