Stability-based generation of scenario trees for multistage stochastic programs

H. Heitsch and W. Römisch
Humboldt-University Berlin
Institute of Mathematics
10099 Berlin, Germany

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

ISMP 2006, Rio de Janeiro, July 30–August 5, 2006
Multistage stochastic programs

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

Multistage stochastic program:

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

where \( X_t \) are nonempty and polyhedral sets, \( A_{t,0} \) are fixed recourse matrices and \( b_t(\cdot), h_t(\cdot) \) and \( A_{t,1}(\cdot) \) are affine functions depending on \( \xi_t \), where \( \xi \) varies in a polyhedral set \( \Xi \).

If the process \( \{\xi_t\}_{t=1}^T \) has a finite number of scenarios, they exhibit a scenario tree structure.
To have the multistage stochastic program well defined, we assume
\[ x_t \in L_{r'}(\Omega, \mathcal{F}, IP; IR^{mt}) \quad \text{and} \quad \xi_t \in L_r(\Omega, \mathcal{F}, IP; IR^d), \]
where \( r \geq 1 \) and

\[ r' := \begin{cases} \frac{r}{r-1}, & \text{if costs are random} \\ r, & \text{if only right-hand sides are random} \\ \infty, & \text{if all technology matrices are random and } r = T. \end{cases} \]

Then nonanticipativity may be expressed as

\[ x \in \mathcal{N}_{r'}(\xi) \]

\[ \mathcal{N}_{r'}(\xi) = \{ x \in \times_{t=1}^T L_{r'}(\Omega, \mathcal{F}, IP; IR^{mt}) : x_t = IE[x_t|\mathcal{F}_t(\xi)], \forall t \}, \]
i.e., as a subspace constraint, by using the conditional expectations

\[ IE[\cdot|\mathcal{F}_t(\xi)]. \]

For \( T = 2 \) we have \( \mathcal{N}_{r'}(\xi) = IR^{m_1} \times L_{r'}(\Omega, \mathcal{F}, P; IR^{m_2}). \)

\[ \rightarrow \text{infinite-dimensional optimization problem} \]
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 \( \pi^n \), \( \pi^n := \sum_{\nu \in \mathcal{N}_+(n)} \pi^{\nu} \) probability of node \( n \), \( \xi^n \) realization of \( \xi_{t(n)} \).
Tree representation of the optimization model

\[
\min \left\{ \sum_{n \in \mathcal{N}} \pi^n \langle b_t(n)(\xi^n), x^n \rangle \left| x^n \in X_t(n), n \in \mathcal{N}, A_{1,0}x^1 = h_1(\xi^1) \right. \right. \\
A_{t(n),0}x^n + A_{t(n),1}x^n = h_t(n)(\xi^n), n \in \mathcal{N} \right. \right. \}
\]

How to solve the optimization model ?

- Standard software (e.g., CPLEX)
- Decomposition methods for (very) large scale models

Questions:

- Under which conditions and in which sense do multistage models behave stable with respect to perturbations of \( \xi \)?
- Can such stability results be used to generate (multivariate) scenario trees?
Dynamic programming

Theorem: (Evstigneev 76, Rockafellar/Wets 76)
Under weak assumptions the multistage stochastic program is equivalent to the (first-stage) convex minimization problem

$$\min \left\{ \int_{\Xi} f(x_1, \xi) P(d\xi) : x_1 \in X_1 \right\},$$

where $f$ is an integrand on $\mathbb{R}^{m_1} \times \Xi$ given by

$$f(x_1, \xi) := \langle b_1(\xi_1), x_1 \rangle + \Phi_2(x_1, \xi^2),$$
$$\Phi_t(x_1, \ldots, x_{t-1}, \xi^t) := \inf \left\{ \langle b_t(\xi_t), x_t \rangle + \mathbb{E} \left[ \Phi_{t+1}(x_1, \ldots, x_t, \xi^{t+1}) \right| F_t \right\} :$$
$$x_t \in X_t, A_{t,0}x_t + A_{t,1}(\xi_t)x_{t-1} = h_t(\xi_t) \right\},$$

for $t = 2, \ldots, T$, where $\Phi_{T+1}(x_1, \ldots, x_T, \xi^{T+1}) := 0.$

$\rightarrow$ The integrand $f$ depends on the probability measure $\mathbb{P}$ and, thus, also on the probability distribution $P = \mathbb{P} \circ \xi^{-1}$ of $\xi$ in a nonlinear way! Hence, earlier approaches to stability fail!
Quantitative Stability

Let us introduce some notations. Let $F$ denote the objective function defined on $L_r(\Omega, F, IP; IR^s) \times L_{r'}(\Omega, F, IP; IR^m) \rightarrow IR$ by $F(\xi, x) := IE[\sum_{t=1}^{T} \langle b_t(\xi_t), x_t \rangle]$, let

$$\mathcal{X}_t(x_{t-1}; \xi_t) := \{x_t \in X_t | A_{t,0}x_t + A_{t,1}(\xi_t)x_{t-1} = h_t(\xi_t)\}$$

denote the $t$-th feasibility set for every $t = 2, \ldots, T$ and

$$\mathcal{X}(\xi) := \{x \in L_{r'}(\Omega, F, IP; IR^m)|x_1 \in X_1, x_t \in \mathcal{X}_t(x_{t-1}; \xi_t)\}$$

the set of feasible elements with input $\xi$. Then the multistage stochastic program may be rewritten as

$$\min\{F(\xi, x) : x \in \mathcal{X}(\xi) \cap \mathcal{N}_{r'}(\xi)\}.$$ 

Let $v(\xi)$ denote its optimal value and, for any $\alpha \geq 0$,

$$l_\alpha(F(\xi, \cdot)) := \{x \in \mathcal{X}(\xi) \cap \mathcal{N}_{r'}(\xi) : F(\xi, x) \leq v(\xi) + \alpha\}$$

$$S(\xi) := l_0(F(\xi, \cdot))$$

denote the $\alpha$-level set and the solution set of the stochastic program with input $\xi$. 
The following conditions are imposed:

(A1) \( \xi \in L_r(\Omega, \mathcal{F}, IP; IR^s) \) for some \( r \geq 1 \).

(A2) There exists a \( \delta > 0 \) such that for any \( \tilde{\xi} \in L_r(\Omega, \mathcal{F}, IP; IR^s) \) with \( \|\tilde{\xi} - \xi\|_r \leq \delta \), any \( t = 2, \ldots, T \) and any \( x_1 \in X_1, x_t \in X_t(x_{t-1}; \tilde{\xi}_t), \tau = 2, \ldots, t - 1 \), the set \( X_t(x_{t-1}; \tilde{\xi}_t) \) is nonempty (relatively complete recourse locally around \( \xi \)).

(A3) The optimal values \( v(\tilde{\xi}) \) are finite for all \( \tilde{\xi} \in L_r(\Omega, \mathcal{F}, IP; IR^s) \) with \( \|\tilde{\xi} - \xi\|_r \leq \delta \) and the objective function \( F \) is level-bounded locally uniformly at \( \xi \), i.e., for some \( \alpha > 0 \) there exists a \( \delta > 0 \) and a bounded subset \( B \) of \( L_r'(\Omega, \mathcal{F}, IP; IR^m) \) such that \( l_\alpha(F(\tilde{\xi}, \cdot)) \) is nonempty and contained in \( B \) for all \( \tilde{\xi} \in L_r(\Omega, \mathcal{F}, IP; IR^s) \) with \( \|\tilde{\xi} - \xi\|_r \leq \delta \).

Norm in \( L_r \): \( \|\xi\|_r := \left( \sum_{t=1}^{T} IE[\|\xi_t\|_r^r] \right)^{1/r} \)
Let (A1), (A2) and (A3) be satisfied, $r > 1$ and $X_1$ be bounded. Then there exist positive constants $L$ and $\delta$ such that

$$|v(\xi) - v(\tilde{\xi})| \leq L(\|\xi - \tilde{\xi}\|_r + D_f(\xi, \tilde{\xi}))$$

holds for all $\tilde{\xi} \in L_r(\Omega, \mathcal{F}, IP; IR^s)$ with $\|\tilde{\xi} - \xi\|_r \leq \delta$.

Assume that technology matrices are non-random and the solution $x^*$ of the original problem is unique.
If $(\xi^{(n)})$ is a sequence in $\times_{t=1}^T L_r(\Omega, \mathcal{F}_t(\xi), IP; IR^s)$ such that

$$\|\xi^{(n)} - \xi\|_r \text{ and } D_f(\xi^{(n)}, \xi)$$

converge to 0 and if $(x^{(n)})$ is a sequence of solutions of the approximate problems, then the sequence $(x^{(n)})$ converges to $x^*$ with respect to the weak topology in $L_{r'}$.

Here, $D_f(\xi, \tilde{\xi})$ denotes the filtration distance of $\xi$ and $\tilde{\xi}$ defined by

$$D_f(\xi, \tilde{\xi}) = \inf_{x \in S(\xi)} \sum_{t=2}^{T-1} \max\{\|x_t - IE[x_t|\mathcal{F}_t(\tilde{\xi})]\|_{r'}, \|\tilde{x}_t - IE[\tilde{x}_t|\mathcal{F}_t(\xi)]\|_{r'}\}. $$
Remark:
The continuity property of infima in the Theorem can be supplemented by a quantitative stability property of the set $S_1(\xi)$ of first stage solutions. Namely, there exists a constant $\hat{L} > 0$ such that

$$\sup_{x \in S_1(\tilde{\xi})} d(x, S_1(\xi)) \leq \Psi_{\xi}^{-1}(\hat{L}(\|\xi - \tilde{\xi}\|_r + D_f(\xi, \tilde{\xi}))),$$

where $\Psi_{\xi}(\tau) := \inf \{ IE[f(x_1, \xi)] - v(\xi) : d(x_1, S_1(\xi)) \geq \tau, x_1 \in X_1 \}$ with $\Psi_{\xi}^{-1}(\alpha) := \sup\{\tau \in IR_+ : \Psi_{\xi}(\tau) \leq \alpha\}$ ($\alpha \in IR_+$) is the growth function of the original problem near its solution set $S_1(\xi)$.

Remark:
Simple examples show that the filtration distance is indispensable for the stability result to hold.
Generation of scenario trees

(i) In most practical situations scenarios $\xi^i$ with known probabilities $p_i$, $i = 1, \ldots, N$, can be generated, e.g., simulation scenarios from (parametric or nonparametric) statistical models of $\xi$ or (nearly) optimal quantizations of the probability distribution of $\xi$.

(ii) Construction of a scenario tree out of the scenarios $\xi^i$ with probabilities $p_i$, $i = 1, \ldots, N$.,
Approaches for (ii):

(1) Bound-based approximation methods,
   (Frauendorfer 96, Kuhn 05, Edirisinghe 99, Casey/Sen 05).

(2) Monte Carlo-based schemes (inside or outside decomposition methods) (e.g. Shapiro 03, 06, Higle/Rayco/Sen 01, Chiralaksanakul/Morton 04).

(3) the use of Quasi Monte Carlo integration quadratures
   (Pennanen 05, 06).

(4) EVPI-based sampling schemes (inside decomposition schemes)
   (Corvera Poire 95, Dempster 04).

(5) Moment-matching principle (Høyland/Wallace 01, Høyland/Kaut/Wallace 03).

(6) (Nearly) best approximations based on probability metrics
   (Pflug 01, Hochreiter/Pflug 02, Mirkov/Pflug 06; Gröwe-Kuska/Heitsch/Römisch 01, 03, Heitsch/Römisch 05).

Survey: Dupačová/Consigli/Wallace 00
Constructing scenario trees

Let $\xi$ be the original stochastic process on some probability space $(\Omega, \mathcal{F}, IP)$ with parameter set $\{1, \ldots, T\}$ and state space $IR^d$. We aim at generating a scenario tree $\xi_{tr}$ such that

$$\|\xi - \xi_{tr}\|_r \quad \text{and} \quad D_f(\xi, \xi_{tr})$$

and, thus,

$$|v(\xi) - v(\xi_{tr})|$$

are small.

To determine such a scenario tree, we start with a discrete approximation $\xi_f$ consisting of scenarios $\xi^i = (\xi^i_1, \ldots, \xi^i_T)$ with probabilities $p_i$, $i = 1, \ldots, N$. $\xi_f$ is a fan of individual scenarios.
The fan $\xi_f$ is chosen such that it is adapted to the filtration $(\mathcal{F}_t(\xi))_{t=1}^T$ and

$$\|\xi - \xi_f\|_r \leq \varepsilon_{\text{appr}}.$$ 

Algorithms are developed that generate a scenario tree $\xi_{tr}$ by deleting and bundling scenarios of $\xi_f$ (that are similar at $t$) such that it is also adapted to the filtration $(\mathcal{F}_t(\xi))_{t=1}^T$ and satisfies

$$(1) \quad \|\xi_f - \xi_{tr}\|_r \leq \varepsilon_r$$

$$(2) \quad \inf_{x \in S(\xi_f)} \sum_{t=2}^{T-1} \|x_t - \mathbb{IE}[x_t|\mathcal{F}_t(\xi_{tr})]\|_{r'} \leq \varepsilon_f.$$ 

Since it holds

$$D_f(\xi, \xi_{tr}) \leq \varepsilon_{\text{appr}} + \inf_{x \in S(\xi_f)} \sum_{t=2}^{T-1} \|x_t - \mathbb{IE}[x_t|\mathcal{F}_t(\xi_{tr})]\|_{r'},$$

if $\xi_f$ is sufficiently close to $\xi$, we obtain in case $\varepsilon_{\text{appr}} + \varepsilon_r \leq \delta$ that

$$|v(\xi) - v(\xi_{tr})| \leq L(2\varepsilon_{\text{appr}} + \varepsilon_r + \varepsilon_f).$$
(1) Forward tree generation

Let scenarios $\xi^i$ with probabilities $p_i$, $i = 1, \ldots, N$, fixed root $\xi_1^* \in \mathbb{R}^d$, $r \geq 1$, and tolerances $\varepsilon_r$, $\varepsilon_t$, $t = 2, \ldots, T$, be given such that $\sum_{t=2}^{T} \varepsilon_t \leq \varepsilon_r$.

**Step 1:** Set $\hat{\xi}^1 := \xi_f$ and $C_1 = \{ I = \{1, \ldots, N\} \}$.

**Step t:** Let $C_{t-1} = \{C_{t-1}^1, \ldots, C_{t-1}^{K_{t-1}}\}$. Determine disjoint index sets $I^k_t$ and $J^k_t$ of remaining and deleted scenarios such that $I^k_t \cup J^k_t = C^k_{t-1}$, a mapping $\alpha_t : I \rightarrow I$

$$\alpha_t(j) = \begin{cases} i^k_t(j), & j \in J^k_t, k = 1, \ldots, K_{t-1}, \\ j, & \text{otherwise}, \end{cases}$$

where $i^k_t(j) \in I^k_t$ such that

$$i^k_t(j) \in \arg \min_{i \in I^k_t} |\hat{\xi}^{t-1,i}_t - \hat{\xi}^{t-1,j}_t|_t,$$
a stochastic process $\hat{\xi}^t$

$$\hat{\xi}^t, i = \begin{cases} 
\xi^\alpha_t(i), & \tau \leq t, \\
\xi^i, & \text{otherwise,} 
\end{cases}$$

such that

$$\|\hat{\xi}^t - \hat{\xi}^{t-1}\|_{r,t} \leq \varepsilon_t.$$ 

Set $I_t := \bigcup_{k=1}^{K_t-1} I^k_t$ and $C_t := \{\alpha_t^{-1}(i) : i \in I^k_t, k = 1, \ldots, K_t-1\}$.

**Step T+1:** Let $C_T = \{C_T^1, \ldots, C_T^{K_T}\}$. Construct a stochastic process $\xi_{tr}$ having $K_T$ scenarios $\xi^{k}_{tr}$ such that $\xi^{k}_{tr,t} := \xi^\alpha_t(i)$ with probabilities $\pi^i_T = \sum_{j \in C^k_T} p_j$ if $i \in C^k_T$, $k = 1, \ldots, K_T$, $t = 2, \ldots, T$.

**Proposition:** $\|\xi_f - \xi_{tr}\|_r \leq \sum_{t=2}^{T} \varepsilon_t \leq \varepsilon_r.$
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>
(2) Bounding approximate filtration distances

Aim: \[ \Delta(\xi_f, \xi_{tr}) := \inf_{x \in S(\xi_f)} \sum_{t=2}^{T-1} \| x_t - IE[x_t | F_t(\xi_{tr})] \|_{r'} \leq \varepsilon_f \]

Two possibilities:

(i) Estimates in terms of some solutions with input \( \xi_f \), which would require to solve a two-stage model.

(ii) Estimates in terms of the input \( \xi_f \).

Proposition:
Let (A2) and (A3) be satisfied and \( X_1 \) be bounded. Assume that the technology matrices \( A_{t,1} \) are non-random, \( 1 \leq r' < \infty \) and \( \xi_f \) is sufficiently close to \( \xi \). Then there exists a constant \( \hat{L} \geq 0 \) such that

\[ \Delta(\xi_f, \xi_{tr}) \leq \hat{L} \left( \sum_{i \in I_2} \sum_{j \in I_2,i} p_j | \xi^j - \xi^i | r' \right)^{\frac{1}{r'}} \]

Condition: \[ \sum_{i \in I_2} \sum_{j \in I_2,i} p_j | \xi^j - \xi^i | r' \leq \varepsilon_f^{r'} \]
Numerical experience

We consider the electricity portfolio management of a municipal power company. Data was available on the electrical load demand and on electricity prices at the market place EEX.

A multivariate statistical model is developed for the yearly demand-price process $\xi$ that allowed to generate yearly demand-price scenarios $\xi_i$, with probabilities $p_i = \frac{1}{N}, i = 1, \ldots, N$.

These scenarios are assumed to form the process $\xi_t$. Branching in $\xi_{\text{tr}}$ was allowed at most monthly. The tolerances $\varepsilon_t$ at branching points were chosen such that

$$\varepsilon_t = \frac{\varepsilon}{T}[1 + \overline{q}(\frac{1}{2} - \frac{t}{T})], \quad t = 2, \ldots, T,$$

where the parameter $\overline{q} \in [0, 1]$ affects the branching structure of the constructed trees. For the test runs we used $\overline{q} = 0.6$.

The test runs were performed on a PC with a 3 GHz Intel Pentium CPU and 1 GByte main memory.
a) Forward tree construction with relative filtration tolerance $\varepsilon_{\text{rel},f} = 0.35$

b) Forward tree construction with relative filtration tolerance $\varepsilon_{\text{rel},f} = 0.45$

Yearly demand-price scenario trees with relative tolerance $\varepsilon_{\text{rel},r} = 0.25$
a) Forward tree construction with relative filtration tolerance $\varepsilon_{rel,f} = 0.6$

b) Forward tree construction with relative filtration tolerance $\varepsilon_{rel,f} = 0.7$

Yearly demand-price scenario trees with relative tolerance $\varepsilon_{rel,r} = 0.5$
<table>
<thead>
<tr>
<th>$\varepsilon_{rel,r}$</th>
<th>$\varepsilon_{rel,f}$</th>
<th>Scenarios</th>
<th>Nodes</th>
<th>Stages</th>
<th>Time (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.10</td>
<td>0.20</td>
<td>98</td>
<td>774 988</td>
<td>6</td>
<td>25.01</td>
</tr>
<tr>
<td>0.30</td>
<td>99</td>
<td>774 424</td>
<td>6</td>
<td>25.05</td>
<td></td>
</tr>
<tr>
<td>0.15</td>
<td>0.25</td>
<td>94</td>
<td>719 714</td>
<td>12</td>
<td>24.97</td>
</tr>
<tr>
<td>0.35</td>
<td>94</td>
<td>723 495</td>
<td>10</td>
<td>24.99</td>
<td></td>
</tr>
<tr>
<td>0.20</td>
<td>0.30</td>
<td>90</td>
<td>670 321</td>
<td>9</td>
<td>24.94</td>
</tr>
<tr>
<td>0.40</td>
<td>90</td>
<td>670 478</td>
<td>10</td>
<td>24.94</td>
<td></td>
</tr>
<tr>
<td>0.25</td>
<td>0.35</td>
<td>85</td>
<td>619 296</td>
<td>9</td>
<td>24.95</td>
</tr>
<tr>
<td>0.45</td>
<td>87</td>
<td>620 340</td>
<td>10</td>
<td>24.93</td>
<td></td>
</tr>
<tr>
<td>0.30</td>
<td>0.40</td>
<td>80</td>
<td>547 824</td>
<td>11</td>
<td>24.86</td>
</tr>
<tr>
<td>0.50</td>
<td>83</td>
<td>567 250</td>
<td>11</td>
<td>24.91</td>
<td></td>
</tr>
<tr>
<td>0.35</td>
<td>0.45</td>
<td>72</td>
<td>482 163</td>
<td>11</td>
<td>24.94</td>
</tr>
<tr>
<td>0.55</td>
<td>76</td>
<td>498 732</td>
<td>11</td>
<td>24.90</td>
<td></td>
</tr>
<tr>
<td>0.40</td>
<td>0.50</td>
<td>67</td>
<td>426 794</td>
<td>8</td>
<td>24.92</td>
</tr>
<tr>
<td>0.60</td>
<td>71</td>
<td>444 060</td>
<td>11</td>
<td>24.90</td>
<td></td>
</tr>
<tr>
<td>0.45</td>
<td>0.55</td>
<td>60</td>
<td>368 380</td>
<td>7</td>
<td>24.97</td>
</tr>
<tr>
<td>0.65</td>
<td>65</td>
<td>383 556</td>
<td>11</td>
<td>24.87</td>
<td></td>
</tr>
<tr>
<td>0.50</td>
<td>0.60</td>
<td>50</td>
<td>309 225</td>
<td>6</td>
<td>24.99</td>
</tr>
<tr>
<td>0.70</td>
<td>60</td>
<td>319 380</td>
<td>11</td>
<td>24.88</td>
<td></td>
</tr>
<tr>
<td>0.55</td>
<td>0.65</td>
<td>44</td>
<td>247 303</td>
<td>6</td>
<td>25.00</td>
</tr>
<tr>
<td>0.75</td>
<td>51</td>
<td>265 336</td>
<td>10</td>
<td>24.91</td>
<td></td>
</tr>
<tr>
<td>0.60</td>
<td>0.70</td>
<td>37</td>
<td>188 263</td>
<td>6</td>
<td>25.17</td>
</tr>
<tr>
<td>0.80</td>
<td>45</td>
<td>203 321</td>
<td>9</td>
<td>24.98</td>
<td></td>
</tr>
</tbody>
</table>

Numerical results for yearly demand-price scenario trees