Two Algorithms for Risk-averse Reformulation of Multi-stage Stochastic Programming Problems

Martin Šmíd[1], Václav Kozmík[2]

Abstract. Many real-life applications lead to risk-averse multi-stage stochastic problems; therefore effective solution of these problems is of great importance. Many tools can be used to their solution (GAMS, Coin-OR, APML or, for smaller problems, Excel); it is, however, mostly up to researcher to reformulate the problem into its deterministic equivalent. Moreover, such solutions are usually one-time, not easy to modify for different applications.

We overcome these problems by providing a front-end software package, written in C++, which enables to enter problem definitions in a way close to their mathematical definition. Creating of a deterministic equivalent (and its solution) is up to the computer.

In particular, our code is able to solve linear multi-stage with Multi-period Mean-CVaR or Nested Mean-CVaR criteria. In the present paper, we describe the algorithms, transforming these problems into their deterministic equivalents.

Keywords: Multi-stage stochastic programming, deterministic equivalent, multi-period CVaR, nested CVaR, optimization algorithm.

JEL classification: C44
AMS classification: 90C15

1 Introduction

Generally, adding risk measures into decision problems destroys their favourable properties such as convexity and linearity. The CVaR risk measure, however, is an exception because it may be reformulated as a convex minimization problem [2], which can be linearised if the underlying distribution is discrete with finite number of atoms. It follows that the favourable properties of CVaR are inherited by the Mean-CVaR risk criterion (see e.g. [4]). Consequently, Mean-CVaR decision problems with convex or linear payoff and constraint functions may be reformulated as convex ones, linear ones, respectively. In the present paper, we do the same in the dynamic case. In particular, we show how to preserve convexity/linearity of multi-stage stochastic programming problems with the Multi-period Mean-CVaR criterion or the Nested Mean-CVaR criterion (see [3] for more about these and other dynamic risk measures).

2 Multistage Stochastic Programming Problem

Let $T \geq 0$. We define T+1-stage Stochastic Programming Problem as

$$\min_{x_k \in \mathcal{X}_k 0 \leq k \leq T} \quad \rho\left(f_0(\overline{\xi}_0, \overline{x}_0), \dots, f_T(\overline{\xi}_T, \overline{x}_T)\right), \tag{1}$$

$$\mathcal{X}_k = \mathcal{X}_k(\overline{\xi}_k, \tilde{x}_{k-1}) = \left\{ x_k \in \mathbb{R}^{d_k}, x_k \in \mathcal{F}_k : x_k \in R_k(\overline{\xi}_k), g_k(\overline{\xi}_k, \overline{x}_k) \lesseqgtr_k 0 \text{ a.s.} \right\}.$$

Here, $\xi = (\xi_0, \dots, \xi_T)$ is a random process taking values in $\mathbb{R}^{n_0} \times \dots \times \mathbb{R}^{n_T}$ with ξ_0 deterministic, $\mathcal{F}_0, \dots, \mathcal{F}_T$ is its induced filtration. Further, for each $0 \leq k \leq T$, d_k is a deterministic constant. For each $0 < k \leq T$, we define $\overline{x}_k = (\tilde{x}_{k-1}, x_k)$, where $\tilde{x}_{k-1}$ is a sub-vector of $\overline{x}_{k-1}$, and we put $\overline{x}_0 = x_0$; symbol $\overline{\xi}_k$ is defined analogously. For each $0 \leq k \leq T$ and each feasible $\overline{x}_k$, $f_k(\bullet, \overline{x}_k)$ is a measurable function such that $\mathbb{E}|f_k(\overline{\xi}_k, \overline{x}_k)| < \infty$, and g_k is a r_k-vector of functions, each of which is non-constant in x_k and measurable in $\overline{\xi}_k$; here, $r_k \in \{0, 1, \dots\}$ is $\overline{\xi}_k$-measurable. Further, for each $0 \leq k \leq T$,, $\lesseqgtr_k$ is a $\overline{\xi}_k$-measurable r_k-vector of symbols from $\{=, \leq, \geq\}$ and R_k is a Cartesian product of d_k closed $\overline{\xi}_k$-measurable intervals. Finally, ρ is a real (risk) functional on the space of integrable random variables which can be the expected sum

$$\rho = \rho^{\mathbb{E}}, \qquad \rho(Z_0, \dots, Z_T) = \sum_{k=0}^{T} \mathbb{E} Z_k,$$

[1]Institute of Information Theory and Automation of the Czech Academy of Sciences, Pod Vodárenskou věží 4, Praha 8, Czech Republic, smid@utia.cas.cz

[2]Institute of Information Theory and Automation of the Czech Academy of Sciences, Pod Vodárenskou věží 4, Praha 8, Czech Republic

the Multi-period Mean-CVaR

$$\rho = \rho^m_{\lambda,\alpha} \qquad \rho^m_{\lambda,\alpha}(Z_0,\dots,Z_T) = Z_0 + \sum_{k=1}^{T} \rho_{\lambda,\alpha}(Z_k|\mathcal{F}_{k-1}), \qquad \rho_{\lambda,\alpha}(Z|\mathcal{F}) = [(1-\lambda)\mathbb{E}(Z|\mathcal{F}) + \lambda \mathrm{CVaR}(Z|\mathcal{F})],$$

where $0 \le \lambda, \alpha \le 1$, or the Nested Mean-CVaR

$$\rho = \rho^n_{\lambda,\alpha} \qquad \rho^n_{\lambda,\alpha}(Z_0,\dots,Z_T) = \rho_{\lambda,\alpha}(\rho_{\lambda,\alpha}(\dots\rho_{\lambda,\alpha}(\sum_{k=0}^{T} Z_k|\mathcal{F}_{T-1})\dots\mathcal{F}_1)|\mathcal{F}_0),$$

where $0 \le \lambda, \alpha \le 1$.

3 Reformulation of the Risk-Averse Problems

If $\rho = \rho^{\mathbb{E}}$ and $f_0,\dots,f_T$ are convex or linear, then so is ρ. Specially, if ξ is discrete with a finite number of atoms and all the f's and g's are linear, then the deterministic equivalent of (1) is a linear programming problem.

If $\rho = \rho^m$ then, by Proposition 1 (i) (see Appendix), problem (1) may be reformulated as

$$\min{}_{x_{0,\dots,T}\in\mathcal{X}_{0,\dots,T},,u_{0,\dots,T-1}\in\mathcal{F}_{0,\dots,T-1}} \quad f_0(\overline{\xi}_0,\overline{x}_0) \quad +\mathbb{E}\left(\sum_{k=1}^{T} h_k(\overline{\xi}_k,\overline{x}_k,u_{k-1})\right) \tag{2}$$

where $h_k(\overline{\xi}_k,\overline{x}_k,u_{k-1}) = u_{k-1} + \left[f_k(\overline{\xi}_k,\overline{x}_k) - u_{k-1}\right]_{\mu,\nu}$; here, $[x]_{a,b} = \begin{cases} xa & x \le 0 \\ xb & x \ge 0 \end{cases}$, $\mu = 1-\lambda$ and $\nu = 1-\lambda+\frac{\lambda}{\alpha}$. If $f_1,\dots,f_T$ are convex, so are $h_t,\dots,h_T$ (note that $[x]_{\mu,\nu}$ is convex). Moreover, if the distribution of ξ is discrete finite, we apply Proposition 1 (ii) to reformulate (2) as

$$\min{}_{x_{0,\dots,T}\in\mathcal{X}_{0,\dots,T},,u_{0,\dots,T-1}\in\mathcal{F}_{0,\dots,T-1},\theta_{1,\dots,T}\in\Theta_{1,\dots,T}} f_0(\overline{\xi}_0,\overline{x}_0) + \mathbb{E}\left(\sum_{k=1}^{T}\theta + \sum_{k=0}^{T-1} u_k\right),$$

$$\Theta_k = \Theta_k(\overline{\xi}_k,\overline{x}_k,u_{k-1}) = \{\theta_k \in \mathbb{R}, \theta_k \in \mathcal{F}_k : \theta_k \ge [f_k(\overline{\xi}_k,\overline{x}_k) - u_{k-1}]_{\mu,\nu}\}, \quad 1 \le k \le T. \tag{3}$$

Specially, if f_k and g_k are linear, $k = 0,\dots,T$, then the deterministic equivalent of (3) is a linear programming problem.

Finally, if $\rho = \rho^n_{\lambda,\alpha}$ and ξ is discrete with a finite number of atoms such that, without loss of generality, the support of ξ_k is $\{1,2,\dots,m_k\}$ for any k, and that $\mathbb{P}[\xi_{k+1}\in\bullet|\xi_0,\dots,\xi_k] = \mathbb{P}[\xi_{k+1}\in\bullet|\tilde{\xi}_k]$ (the latter implying $\rho_{\lambda,\alpha}(h(\overline{\xi}_k)|\mathcal{F}_{k-1}) = \rho_{\lambda,\alpha}(h(\overline{\xi}_k)|\tilde{\xi}_{k-1})$ for any h by the definition of Mean-CVaR), then, by the translational invariance of CVaR and by Proposition 1 (iii),

$$\begin{aligned}\min_{x_\bullet\in\mathcal{X}_\bullet} \rho(f_0,\dots,f_k) &= \min_{x_\bullet\in\mathcal{X}_\bullet}\left[f_0 + \rho_{\lambda,\alpha}(f_1 + \rho_{\lambda,\alpha}(f_2 + \dots + \rho_{\lambda,\alpha}(f_T|\tilde{\xi}_{T-1})\dots|\tilde{\xi}_1))\right] \\ &= \min_{x_0\in\mathcal{X}_0}\{f_0 + \rho_{\lambda,\alpha}(\min_{x_1\in\mathcal{X}_1}\{f_1 + \rho_{\lambda,\alpha}(\min_{x_2\in\mathcal{X}_2}\{f_2 + \dots + \rho_{\lambda,\alpha}(\min_{x_T\in\mathcal{X}_T}\{f_T\}|\tilde{\xi}_{T-1})\dots|\tilde{\xi}_2)\}|\tilde{\xi}_1)\})\} \\ &= \min_{x_0\in\mathcal{X}_0} q_0(\xi_0,x_0)\end{aligned}$$

where, for any $0 \le k \le T-1$,

$$q_k(\overline{\xi}_k,\overline{x}_k) = f_k(\overline{\xi}_k,\overline{x}_k) + \rho_{\lambda,\alpha}\left(Q_k(\overline{\xi}_{k+1},\tilde{x}_k)|\tilde{\xi}_k\right), \qquad Q_k(\overline{\xi}_{k+1},\tilde{x}_k) = \min_{x_{k+1}\in\mathcal{X}_k(\overline{\xi}_{k+1},\tilde{x}_k)} q_{k+1}(\overline{\xi}_{k+1},\overline{x}_{k+1}),$$

and $q_T(\overline{\xi}_T,\overline{x}_T) = f_T(\overline{\xi}_T,\overline{x}_T)$. Moreover, if, for some $1 \le k \le T-1$,

$$Q_k(\overline{\xi}_{k+1},\tilde{x}_k) = \min_{y\in\mathcal{Y}_k^{\xi_{k+1}}(\tilde{\xi}_k,\tilde{x}_k)} \zeta_k^{\xi_{k+1}}(y;\tilde{\xi}_k,\tilde{x}_k) \tag{4}$$

for some functions $\zeta^1,\zeta^2,\dots,\zeta^{m_{k+1}}$ and parametric sets $\mathcal{Y}^1,\dots,\mathcal{Y}^{m_{k+1}}$, we get, by Proposition 1 (iii) and (iv), that

$$q_k(\overline{\xi}_k,\overline{x}_k) = f_k(\overline{\xi}_k,\overline{x}_k) + \min_{u\in\mathbb{R},y^i\in\mathcal{Y}_k^i(\tilde{\xi}_k,\tilde{x}_k),\theta^i\ge[\zeta^i(y^i;\tilde{\xi}_k,\tilde{x}_k)-u]_{\mu,\nu},i=1,\dots,m_{k+1}}\left\{u + \sum_{i=1}^{m_{k+1}}\theta^i\pi^i_{k+1}(\tilde{\xi}_k)\right\} \tag{5}$$

where $\pi^i_{k+1}(\tilde{\xi}_k) = \mathbb{P}[\xi_{k+1} = i|\tilde{\xi}_k]$, $1 \leq i \leq m_{k+1}$, and, consequently,

$$Q_{k-1}(\bar{\xi}_k, \tilde{x}_k) = \min_{(x_k,u,y^1,\theta^1,\dots,y^{m_{k+1}},\theta^{m_{k+1}}) \in \mathcal{Y}^{\xi_k}_{k-1}(\tilde{\xi}_{k-1},\tilde{x}_{k-1})} \left\{ f_k(\bar{\xi}_k, \overline{x}_k) + u + \sum_{i=1}^{m_{k+1}} \theta^i \pi^i_{k+1}(\tilde{\xi}_k) \right\}$$

where

$$\mathcal{Z}^{\xi_k}_{k-1}(\tilde{\xi}_{k-1}, \tilde{x}_{k-1}) = \mathcal{X}_k(\bar{\xi}_k, \tilde{x}_{k-1}) \times \mathbb{R} \times \mathcal{Y}^1(\tilde{\xi}_k, \tilde{x}_k) \times \{\theta^1 \geq [\zeta^1(y^1; \tilde{\xi}_k, \tilde{x}_k) - u]_{\mu,\nu}\} \times \cdots \times \{\theta^{m_{k+1}} \geq [\zeta^{m_{k+1}}(y^{m_{k+1}}; \tilde{\xi}_k, \tilde{x}_k) - u]_{\mu,\nu}\}$$

which fulfills (4) with $k-1$ in place of k. As (4) holds for $k = T-1$, we get, by induction, that (5) holds for any $0 \leq k \leq T-1$. Consequently,

$$\min_{x_\bullet \in \mathcal{X}_\bullet} \rho(f_0, \dots, f_k) = \min_{x_0 \in \mathcal{X}_0, u \in \mathbb{R}, y^i \in \mathcal{Y}^i_1(\tilde{\xi}_0, \tilde{x}_0), \theta^1 \geq [\zeta^i_1(y^i; \tilde{\xi}_0, \tilde{x}_0) - u]_{\mu,\nu}, i=1,\dots,m_1} \left\{ f_0(\xi_0, x_0) + u + \sum_{i=1}^{m_1} \theta_i \pi_i \right\} \quad (6)$$

Specially, if $f_0, \dots, f_T$are linear in x and $g_0, \dots, g_T$ affine in x then, by induction, (6) is a linear programming problem.

4 Conclusion

In this paper, two algorithms linearizing risk averse multistage stochastic programming problems were described. These algorithms are implemented in the MS++, which is a C++ software package developed by the authors. Even though the algorithms use standard techniques, which are routinely applied to solve particular problems (see e.g. [1]), they have been neither rigorously described nor generally implemented yet to the best knowledge of the authors. The package is freely available on `https://github.com/cyberklezmer/mspp`.

Acknowledgements

This work was supported by grant No. GA 16-01298S of the Czech Science Foundation. The support is kindly acknowledged.

References

[1] Václav Kozmík and David P Morton. Evaluating policies in risk-averse multi-stage stochastic programming. *Mathematical Programming*, 2014.

[2] R Tyrrell Rockafellar and Stanislav Uryasev. Conditional value-at-risk for general loss distributions. *Journal of banking & finance*, 26(7):1443–1471, 2002.

[3] Alexander Shapiro and Alois Pichler. Time and dynamic consistency of risk averse stochastic programs. *Published electronically in: Optimization Online*, 2016.

[4] Martin Šmíd, František Zapletal, and Jana Hančlová. Which carbon derivatives are applicable in practice? a case study of a european steel company. *Kybernetika*, 53(6):1071–1085, 2017.

Appendix

Proposition 1. *(i) For any integrable random variable Z,*

$$\rho_{\lambda,\alpha}(Z|\mathcal{F}) = \min_{u \in \mathbb{R}} \left[u + \mathbb{E}([Z-u]_{\mu,\nu} |\mathcal{F}) \right]$$

where $[x]_{a,b} = \begin{cases} xa & x \leq 0 \\ xb & x \geq 0 \end{cases}$, $\mu = 1 - \lambda$, $\nu = 1 - \lambda + \frac{\lambda}{\alpha}$.

(ii) If $Z|\mathcal{F} \sim (z_i, \pi_i)^n_{i=1}$, then

$$\rho_{\lambda,\alpha}(Z|\mathcal{F}) = \min_{u \in \mathbb{R}, \theta_i \geq \mu(z_i - u), \theta \geq \nu(z_i - u), i=1,\dots,n} \left[u + \sum_{i=1}^n \theta_i \pi_i \right].$$

(iii) If, in addition, $\mathcal{F}$ is finite and $z_i = \min_{x_i \in \mathcal{X}_i} \zeta_i(x_i)$, $i = 1, \dots, n$, for some sets $\mathcal{X}_1, \mathcal{X}_2 \dots$ and functions $\zeta_1, \zeta_2, \dots$, all possibly dependent on some $\mathcal{F}$-measurable parameter, then

$$\rho_{\lambda,\alpha}(Z|\mathcal{F}) = \min_{x_i \in \mathcal{X}_i, i=1,\dots,n} \{\rho_{\lambda,\alpha}(\zeta_i(x_i))\}$$

(iv) Moreover,

$$\rho_{\lambda,\alpha}(Z|\mathcal{F}) = \min_{u\in\mathbb{R},x_i\in\mathcal{X}_i,i=1,\dots,n} \left\{ u + \sum_{i=1}^{n} [\zeta_i(x_i) - u]_{\mu,\nu}\, \pi_i \right\}$$
$$= \min_{u\in\mathbb{R},x_i\in\mathcal{X}_i,\theta_i\geq\mu(\zeta_i(x_i)-u),\theta_i\geq\nu(\zeta_i(x_i)-u),i=1,\dots,n} \left[u + \sum_{i=1}^{n} \theta_i \pi_i \right].$$

Proof. Ad (i) and (ii). Clearly, $cx + d\,[x]_+ = [x]_{c.c+d}$ for any $c, d \geq 0$. Using that and [2] and we get

$$\begin{aligned}
\varrho(Z) &= (1-\lambda)\mathbb{E}(Z) + \lambda \mathrm{CVaR}_\alpha(Z) \\
&= (1-\lambda)\mathbb{E}Z + \lambda \min_u \left(u + \frac{1}{\alpha}\mathbb{E}\,[Z-u]_+ \right) \\
&= \min_u \left[(1-\lambda)\mathbb{E}Z + \lambda u + \frac{\lambda}{\alpha}\mathbb{E}\,[Z-u]_+ \right] \\
&= \min_u \left[(1-\lambda)\mathbb{E}Z - (1-\lambda)u + u + \frac{\lambda}{\alpha}\mathbb{E}\,[Z-u]_+ \right] \\
&= \min_u \left[(1-\lambda)\mathbb{E}(Z-u) + u + \frac{\lambda}{\alpha}\mathbb{E}\,[Z-u]_+ \right] \\
&= \min_u \left[u + \mathbb{E}\left((1-\lambda)(Z-u) + \frac{\lambda}{\alpha}\,[Z-u]_+ \right) \right] \\
&= \min_u \left[u + \mathbb{E}\,[Z-u]_{\mu,\nu} \right] = \min_{u,\theta_i=[Z_i-u]_{\mu,\nu}} \left[u + \sum_{\mathrm{i}} \pi_i\theta_i \right] \\
&= \min_{u,\theta_i\geq\mu(Z_i-u),\theta_i\geq\nu(Z_i-u_i)} \left[u + \sum_{\mathrm{i}} \pi_i\theta_i \right] = \min_{u,\theta\in\sigma(Z),\theta\geq\mu(Z-u),\theta\geq\nu(Z-u)} [u + \mathbb{E}(\theta)]
\end{aligned}$$

Ad. (iii) and (iv). Denote $x = (x_1, \dots, x_n)$, $\mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_n$ and $\hat{x} = (\arg\min_{x_1\in\mathcal{X}_1} \zeta_1(x), \dots, \arg\min_{x_n\in\mathcal{X}_n} \zeta_n(x))$. For any $x \in \mathcal{X}$, introduce a random variable $Y(x) \sim (\zeta_i(x_i), \pi_i)$. We have $Z \leq Y(x)$ for any $x \in \mathcal{X}$ so, by monotonicity of Mean-CVaR, $\rho(Z) \leq \rho(Y(x))$ so $\rho(Z) \leq \inf_x \rho(Y(x))$. As $Y(\hat{x}) = Z$, the infimum is attained, so

$$\rho(Z) = \min_{x\in\mathcal{X}} (\rho(Y(x)).$$

The first equality in (iv) now follows by (i), the second one may be got analogously to the proof of (ii). □