SOLVING STOCHASTIC PROGRAMMING PROBLEMS BY PROGRESSIVE HEDGING WITH RISK MEASURES IN THE OBJECTIVE

Terry Rockafellar
University of Washington, Seattle
University of Florida, Gainesville

Workshop on Variational and Stochastic Analysis
CMM, University of Chile, Santiago
15–16 March 2017

Stochastic Structure with Emerging Information

Pattern of "decisions" and "observations" in N stages:

$$x_1,\ \xi_1,\ x_2,\ \xi_2, \dots, x_N,\ \xi_N \quad \text{with} \quad x_k \in \boldsymbol{R}^{n_k},\ \xi_k \in \Xi_k$$

$$x = (x_1, \dots, x_N) \in \boldsymbol{R}^n = \boldsymbol{R}^{n_1} \times \cdots \times \boldsymbol{R}^{n_N}$$

$$\xi = (\xi_1, \dots, \xi_N) \in \Xi \ \subset \Xi_1 \times \cdots \Xi_N$$

Interpretation: each $\xi \in \Xi$ is an information **scenario**

Nonanticipativity of decisions

x_k can respond to $\xi_1, \dots, \xi_{k-1}$ but not to $\xi_k, \dots, \xi_N$:

$$x(\xi) = (x_1, x_2(\xi_1), x_3(\xi_1, \xi_2), \dots, x_N(\xi_1, \xi_2, \dots, \xi_{N-1}))$$

Simplifying assumptions for this talk:

- the scenario space Ξ has only finitely many elements ξ
- each scenario $\xi \in \Xi$ has known probability $p(\xi) > 0$

$\longrightarrow$ Ξ is a probability space

Response Function Framework (Rock. & Wets, 1976)

$$\Xi \subset \Xi_1 \times \cdots \Xi_N, \qquad \boldsymbol{R}^n = \boldsymbol{R}^{n_1} \times \cdots \times \boldsymbol{R}^{n_N}$$

$\mathcal{L}$ = **all** functions from **scenario** space Ξ to **decision** space $\boldsymbol{R}^n$

$$x(\cdot) : \xi = (\xi_1, \ldots, \xi_N) \mapsto x(\xi) = (x_1(\xi), \ldots, x_N(\xi))$$

Nonanticipativity subspace: with $\xi = (\xi_1, \ldots, \xi_{k-1}, \xi_k, \ldots, \xi_N)$

$$\mathcal{N} = \left\{ x(\cdot) \in \mathcal{L} \,\middle|\, x_k(\xi) \text{ depends only on } \xi_1, \ldots, \xi_{k-1} \right\}$$

$\longrightarrow$ $x(\cdot)$ is nonanticipative $\iff$ $x(\cdot) \in \mathcal{N}$

Expectation inner product: for $x(\cdot), w(\cdot) \in \mathcal{L}$

$$\langle x(\cdot), w(\cdot) \rangle = E_\xi[x(\xi) \cdot w(\xi)] = \sum_{\xi \in \Xi} p(\xi) \sum_{k=1}^N x_k(\xi) \cdot w_k(\xi)$$

Complementary subspace: $\mathcal{M} = \mathcal{N}^\perp$ ("martingale" space)

$$\mathcal{M} = \left\{ w(\cdot) \in \mathcal{L} \,\middle|\, E_{\xi_k, \ldots, \xi_N}[w_k(\xi_1, \ldots, \xi_{k-1}, \xi_k \ldots, \xi_N)] = 0 \right\}$$

Multistage Stochastic Programming in this Setting

Problem ingredients: for each scenario $\xi \in \Xi$, let

$$C(\xi) = \text{nonempty closed convex set in } \boldsymbol{R}^n$$
$$g(x,\xi) = \text{continuous convex function of } x \in C(\xi)$$

Scenario constraint on responses: $x(\cdot) \in \mathcal{C}$, where

$$\mathcal{C} = \{ x(\cdot) \in \mathcal{L} \mid x(\xi) \in C(\xi) \subset \boldsymbol{R}^n \text{ for all } \xi \in \Xi \}$$

Risk-neutral objective function: $\mathcal{G} : \mathcal{C} \to \boldsymbol{R}$, where

$$\mathcal{G}(x(\cdot)) = E_\xi[g(x(\xi),\xi)] = \sum_{\xi \in \Xi} p(\xi) g(x(\xi),\xi)$$

Note: $\mathcal{C} \subset \mathcal{L}$ is closed convex, $\mathcal{G} : \mathcal{C} \to \boldsymbol{R}$ is continuous

Stochastic programming problem

minimize $\mathcal{G}(x(\cdot))$ over all functions $x(\cdot) \in \mathcal{C} \cap \mathcal{N}$

$\mathcal{N}$ = nonanticipativity subspace of $\mathcal{L}$

Risk-averse objective function as an alternative:

$\mathcal{G}(x(\cdot)) = \mathrm{CVaR}_\alpha(G(x(\cdot)))$ for the r.v. $G(x(\cdot)) : \xi \to g(x(\xi),\xi)$

$\mathrm{CVaR}_\alpha(\text{r.v. } X)$ = conditional expectation in upper α-tail of X

Progressive Hedging Approach (Rock.& Wets 1991)

General form of the procedure:

- Introduce information cost "multipliers" $w(\cdot) \in \mathcal{M} = \mathcal{N}^{\perp}$
- In iterations $\nu = 1, 2, \ldots$

solve "hindsight" problems for the separate scenarios ξ in which the cost $g(x, \xi)$ is modified to

$$g^{\nu}(x, \xi) = g(x, \xi) + w^{\nu}(\xi) \cdot x + \frac{r}{2} ||x - x^{\nu}(\xi)||^2$$

with respect to the current $x^{\nu}(\cdot) \in \mathcal{N}$ and $w^{\nu}(\cdot) \in \mathcal{M}$

- This yields $\hat{x}^{\nu}(\xi)$ for each ξ, but the response function $\hat{x}^{\nu}(\cdot)$

won't be nonanticipative. Restore nonanticpativity by projection onto $\mathcal{N}$ and generate an update for the information costs in $\mathcal{M}$

- the risk-neutral case of $E_{\xi}[g^{\nu}(x(\xi), \xi)] = \sum_{\xi \in \Xi} p(\xi) g(x(\xi), \xi)$

supports the decomposition into separate scenario subproblems

- the risk-averse case with CVaR_{α} has **no separability directly**,

but **separability can be achieved,** as will be explained later

Projection Tool for Aggregating Responses

Recalling the structure of the complementary subspaces:

$$\mathcal{N} = \{x(\cdot) \in \mathcal{L} \mid x_k(\xi) \text{ depends only on } \xi_1, \ldots, \xi_{k-1}\}$$

$$\mathcal{M} = \{w(\cdot) \in \mathcal{L} \mid E_{\xi_k,\ldots,\xi_N}[w_k(\xi_1, \ldots, \xi_{k-1}, \xi_k \ldots, \xi_N)] = 0\}$$

Aggregation: let $\mathcal{P}$ = projection onto $\mathcal{N}$

then $\mathcal{I} - \mathcal{P}$ = projection onto $\mathcal{M}$, since $\mathcal{M} = \mathcal{N}^{\perp}$

Execution relative to the information structure:

- Scenarios $\xi = (\xi_1, \ldots, \xi_N)$ and $\xi' = (\xi'_1, \ldots, \xi'_N)$ are at stage k information-equivalent if $(\xi_1, \ldots, \xi_{k-1}) = (\xi'_1, \ldots, \xi'_{k-1})$
- Let $A_k(\xi)$ = kth-stage equivalence class containing ξ
- Then $x(\cdot) = \mathcal{P}(\bar{x}(\cdot))$ has its kth-stage component given by

$$x_k(\xi) = \sum_{\xi' \in A_k(\xi)} p(\xi')\bar{x}_k(\xi') / \sum_{\xi' \in A_k(\xi)} p(\xi')$$

thus $x_k(\xi)$ is the **conditional expectation** of $\bar{x}_k(\xi)$ relative to the kth-stage information-equivalence class containing ξ

Progressive Hedging in Stochastic Programming

Algorithm statement in the risk-neutral case with parameter $r > 0$

Having $x^\nu(\cdot) \in \mathcal{N}$ and $w^\nu(\cdot) \in \mathcal{M}$, get $\hat{x}^\nu(\cdot) \in \mathcal{L}$ by

$$\hat{x}^\nu(\xi) = \operatorname{argmin}_{x \in C(\xi)} \left\{ g(x,\xi) + x \cdot w^\nu(\xi) + \tfrac{r}{2} ||x - x^\nu(\xi)||^2 \right\}$$

Then get $x^{\nu+1}(\cdot) \in \mathcal{N}$ and $w^{\nu+1}(\cdot) \in \mathcal{M}$ by aggregation:

$$x^{\nu+1}(\cdot) = \mathcal{P}(\hat{x}^\nu(\cdot)), \qquad w^{\nu+1}(\cdot) = w^\nu(\cdot) + r[\hat{x}^\nu(\cdot) - x^{\nu+1}(\cdot)]$$

Convergence theorem — when a solution pair $x(\cdot)$, $w(\cdot)$, exists

The sequence $\{(x^\nu(\cdot), w^\nu(\cdot))\}_{\nu=1}^\infty$ generated by the algorithm will always converge to a particular solution pair $(x^*(\cdot), w^*(\cdot))$, with

$$\begin{aligned} &||x^{\nu+1}(\cdot) - x^*(\cdot)||^2 + \tfrac{1}{r^2} ||w^{\nu+1}(\cdot) - w^*(\cdot)||^2 \\ &\qquad \leq \; ||x^\nu(\cdot) - x^*(\cdot)||^2 + \tfrac{1}{r^2} ||w^\nu(\cdot) - w^*(\cdot)||^2 \end{aligned}$$

Adaptation of Progressive Hedging to a Risk-Averse Case

CVaR Minimization formula: Rock.& Uryasev (2000, 2002)

$$\mathrm{CVaR}_\alpha(X) = \min_{z\in R}\left\{ z + \frac{1}{1-\alpha} E[\max\{0, X - z\}] \right\}$$

Consequence: for the random variable $G(x(\cdot)) : \xi \to g(x(\xi), \xi)$,

$$\mathrm{CVaR}_\alpha(G(x(\cdot)) = \min_{z\in R}\left\{ z + \frac{1}{1-\alpha} E_\xi[\max\{0, g(x(\xi), \xi) - z\}] \right\}$$

Risk-Averse stochastic programming problem, reformulated

minimizing $\mathrm{CVaR}_\alpha(G(x(\cdot))$ over $x(\cdot) \in \mathcal{C} \cap \mathcal{N}$ is equivalent to minimizing $E_\xi[\, h(z, x(\xi), \xi)\,]$ over $z \in R$, $x(\cdot) \in \mathcal{C} \cap \mathcal{N}$, where

$$h(z, x(\xi), \xi) = z + \frac{1}{1-\alpha} \max\{0, g(x(\xi), \xi) - z\}$$

Route to computation:

- Incorporate z within $x(\cdot)$ as an extra first-stage variable
- Then just apply the risk-neutral version of the progressive hedging algorithm with h taking the place of g

Example: The One-Stage Case

Simplified pattern of decisions and observations:

$x \in \boldsymbol{R}^n$ followed by $\xi \in \Xi$ yielding cost $g(x, \xi)$

Response functions: $x(\cdot) \in \mathcal{C} \cap \mathcal{N}$,

$x(\xi) \in C(\xi)$ but also $x(\xi) \equiv$ constant

Risk-averse optimization problem:

minimize $\mathrm{CVaR}_\alpha(g(x(),\cdot))$ over $x(\cdot) \in \mathcal{C} \cap \mathcal{N}$

Progressive hedging in this setting

In iteration ν with x^ν and z^ν along with $w^\nu(\cdot)$, $u^\nu(\cdot)$, having $E_\xi[w^\nu(\xi)] = 0$, $E_\xi[u^\nu(\xi)] = 0$, get $(\hat{x}^\nu(\xi), \hat{z}^\nu(\xi))$ for each ξ from

$$\min_{x(\xi), z(\xi)} \Big\{ z(\xi) + \tfrac{1}{1-\alpha} \max\{0, g(x(\xi), \xi) - z(\xi)\} - w^\nu(\xi) \cdot x(\xi) \\ - u^\nu(\xi) z(\xi) + \tfrac{r}{2} ||x(\xi) - x^\nu||^2 + \tfrac{r}{2} |z(\xi) - z^\nu|^2 \Big\}$$

Then update by taking

$$x^{\nu+1} = E_\xi[\hat{x}^\nu(\xi)], \quad w^{\nu+1}(\xi) = w^\nu(\xi) + r[\hat{x}^\nu(\xi) - x^{\nu+1}],$$
$$z^{\nu+1} = E_\xi[\hat{z}^\nu(\xi)], \quad u^{\nu+1}(\xi) = u^\nu(\xi) + r[\hat{z}^\nu(\xi) - z^{\nu+1}].$$

Extension to Other Measures of Risk Than CVaR

Risk measures of expectation type: on r.v.'s $X : \Xi \to R$

$$\mathcal{R}(X) = \min_{z \in R} \{ z + E_\xi[\, v(\, X(\xi) - z\,)\,] \}$$ for a "regret" function v

CVaR_α case: $v(t) = \frac{1}{1-\alpha} \max\{0, t\}$

$\longrightarrow$ adaptation proceeds just with this different v!

Mixtures of such measures: weights $\lambda_i > 0$, $\sum_{i=1}^m \lambda_i = 1$

$\mathcal{R} = \sum_{i=1}^m \lambda_i \mathcal{R}_i$, where $\mathcal{R}_i(X) = \min_{z_i \in R} \{ z_i + E_\xi[\, v_i(\, X(\xi) - z_i\,)\,] \}$

$$\Longrightarrow \quad \mathcal{R}(X) = \min_{z_1,\ldots,z_m} E_\xi\Big[\sum_{i=1}^m \lambda_i [z_i + v_i(\, X(\xi) - z_i)\,] \Big]$$

$\longrightarrow$ adaptation proceeds similarly with m auxiliary variables!

Extension to Nested Risk in Ruszczyński's Sense

two-stage structure, for simplicity: pattern x_1, ξ_1, x_2, ξ_2

First-stage risk: a risk measure $\mathcal{R}_1$ for r.v.'s in ξ_1

Second-stage risk: risk measures $\mathcal{R}_{2,\xi_1}$ for r.v.'s in ξ_2

Formulation of objective

For nonanticipative $(x_1, x_2(\cdot))$ consider

- cost r.v.'s $\xi_2 \to g_2(x_1, x_2(\xi_1), \xi_1, \xi_2)$ and get an r.v. in ξ_1 by applying $\mathcal{R}_{2,\xi_1}$ to them.
- add that r.v. to the cost r.v. $\xi_1 \to g_1(x_1, \xi)$ and then apply $\mathcal{R}_1$ to get a numerical value for it.
- that is the value to be minimized with respect to $(x_1, x_2(\cdot))$.

Corresponding adaptation to progressive hedging:

first-stage auxiliary parameter introduced for first-stage

second-stage auxiliary parameters introduced for second-stage

Some References

[1] R.T. Rockafellar (2017) "Solving stochastic optimization problems with risk measures by the progressive hedging algorithm," Mathematical Programming B (submitted).

[2] R.T. Rockafellar and R.J-B Wets (1991) "Scenarios and policy aggregation in optimization under uncertainty," *Mathematics of Operations Research* 16, 119–147.

[3] R.T. Rockafellar and S. Uryasev (2002) "Conditional value-at-risk for general loss distributions," *Journal of Banking and Finance* 26, 1443–1471.

[4] R.T. Rockafellar and S. Uryasev (2013) "The fundamental risk quadrangle in risk management, optimization and statistical estimation," *Surveys in Operations Research and Management science* 18, 33–53.

website: www.math.washington.edu/∼rtr/mypage.html