THE PROGRESSIVE HEDGING ALGORITHM IN NONCONVEX STOCHASTIC OPTIMIZATION

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

Risk Management Approaches in Engineering Applications
Workshop, University of Florida, Gainesville
1–2 October 2018

A Basic Model in Stochastic Optimization

Information pattern: here single-stage

decision $x \in \boldsymbol{R}^n$ followed by observing $\xi \in \Xi$ (prob. space)

multistage extension: repeated interplay —- bypassed here

Problem

minimize $E_\xi\left[f_0(x,\xi)\right]$ subject to $F(x,\xi) \in K \subset \boldsymbol{R}^m$, where

K = closed convex cone, $F(x,\xi) = (f_1(x,\xi), \ldots, f_m(x,\xi))$

Alternative objectives: to just minimizing an "expected cost"

- minimizing some measure of risk, or
- minimizing bPOE at some threshold

these extensions can be subsumed into the expectation model!

Simplifying Assumption

there are finitely many scenarios $\xi \in \Xi$, probabilities $p(\xi) > 0$

Problem restatement: in a simplified form

$$\text{minimize} \quad \varphi(x) = \sum_\xi p(\xi) f(x,\xi) \text{ over } x \in \boldsymbol{R}^n,$$

$$\text{where} \qquad f(x,\xi) = \begin{cases} f_0(x,\xi) & \text{if } F(x,\xi) \in K, \\ \infty & \text{if } \notin K \end{cases}$$

Convex case: occurs if each $f(x,\xi)$ is convex in x, as when

- $f_0(x,\xi)$ is convex with respect to x
- the set $C(x,\xi) = \{x \mid F(x,\xi) \in K\}$ is convex

Optimality condition: on $\bar{x}$ from subgradient rule $0 \in \partial\varphi(\bar{x})$

$$\exists\, \bar{w}(\xi) \in \partial f(\bar{x},\xi) \text{ such that } \sum_\xi p(\xi)\bar{w}(\xi) = 0$$

Status: sufficient for global optimality in the convex case, and necessary for local optimality in general under a constraint qual.

Computational focus: find $\bar{x}$ and $\bar{w}(\xi)$ satisfying this condition

Progressive Hedging Background

Aim: reduce computations to iteratively solving subproblems which depend only on the individual scenarios $\xi \in \Xi$

Original algorithm (convex case) — with proximal parameter $r > 0$

In iteration k, having x^k and $w^k(\xi)$ with $\sum_\xi p(\xi) w^k(\xi) = 0$, get

$$\widehat{x}^k(\xi) = \operatorname*{argmin}_{x \in \boldsymbol{R}^n} \Big\{ f(x,\xi) - w^k(\xi) \cdot x + \tfrac{r}{2} ||x - x^k||^2 \Big\}$$
$$= \operatorname*{argmin}_{F(x,\xi) \in K} \Big\{ f_0(x,\xi) - w^k(\xi) \cdot x + \tfrac{r}{2} ||x - x^k||^2 \Big\}$$

(taking advantage of strong convexity in x), and then update by

$$x^{k+1} = \sum_\xi p(\xi)\, \widehat{x}^k(\xi), \qquad w^{k+1}(\xi) = w^k(\xi) - r\big[\widehat{x}^k(\xi) - x^{k+1}\big]$$

Convergence: in convex case, global from any initial x^0, $w^0(\xi)$

Challenge: how to adapt this now to a **nonconvex** setting?

$f_0(x,\xi)$ not convex? $C(x,\xi) = \big\{ x \,\big|\, F(x,\xi) \in K \big\}$ not convex?

Reformulation Toward Accommodating Nonconvexity

Linkage problem format: Rock. 2018

minimize a function φ over some "linkage" subspace $\mathcal{S}$

$\longrightarrow$ "progressive decoupling algorithm" that can "elicit" convexity

New context: the space $\mathcal{L}$ = all $(x(\cdot), u(\cdot)) = (x(\xi), u(\xi))_{\xi \in \Xi}$

Extended problem statement

$$\min \ \varphi(x(\cdot), u(\cdot)) = \sum_\xi p(\xi) \Big[f_0(x(\xi), \xi) + \delta_K\big(F(x(\xi), \xi) + u(\xi)\big) \Big]$$

over the subspace $\mathcal{S}$ of the space $\mathcal{L}$ defined by

for all $\xi \in \Xi$, $x(\xi)$ = the same $x \in \boldsymbol{R}^n$, while $u(\xi) = 0$

Complementary subspace: orthogonal to $\mathcal{S}$ in $\mathcal{L}$

$$\mathcal{S}^\perp = \Big\{ (w(\cdot), y(\cdot)) = (w(\xi), y(\xi))_{\xi \in \Xi} \,\Big|\, \sum_\xi p(\xi) w(\xi) = 0 \Big\}$$

$$\Big\langle (x(\cdot), u(\cdot)), (w(\cdot), y(\cdot)) \Big\rangle = \sum_\xi p(\xi) \, (x(\xi), u(\xi)) \cdot (w(\xi), y(\xi))$$

Progressive Decoupling in this Stochastic Setting

specializing a new, very general procedure of Rock. 2018

Algorithm in "raw" form — with parameters $r > e \geq 0$

Having $(x^k(\xi), u^k(\xi))_{\xi\in\Xi} \in \mathcal{S}$ and $(w^k(\xi), y^k(\xi))_{\xi\in\Xi} \in \mathcal{S}^\perp$ find

$$(\widehat{x}^k(\xi), \widehat{u}^k(\xi)) \in \operatorname*{argmin}_{x,u} \varphi^k(x, u, \xi) \text{ for each } \xi \in \Xi$$

where $\varphi^k(x, u, \xi) = f_0(x, \xi) + \delta_K(F(x, \xi) + u)$
$$-w^k(\xi)\cdot x - y^k(\xi)\cdot u + \tfrac{r}{2}||x - x^k(\xi)||^2 + \tfrac{r}{2}||u - u^k(\xi)||^2$$

and then update by

$$(x^{k+1}(\xi), u^{k+1}(\xi))_{\xi\in\Xi} = \text{ projection of } \left(\widehat{x}^k(\xi), \widehat{u}^k(\xi)\right)_{\xi\in\Xi} \text{ on } \mathcal{S},$$
$$(w^{k+1}(\xi), y^{k+1}(\xi)) = (w^k(\xi), y^k(\xi)) - (r - e)\left[(\widehat{x}^k(\xi), \widehat{u}^k(\xi)) - (x^{k+1}(\xi), u^{k+1}(\xi))\right]$$

$e =$ <u>elicitation parameter</u> which needs to be "<u>high enough</u>"

Consolidation With the Specifics of $\mathcal{S}$ and $\mathcal{S}^\perp$

here $x^k(\xi) =$ same $x^k \in \boldsymbol{R}^n$ for all ξ, while $u^k(\xi) = 0$ for all ξ

Having x^k, $y^k(\xi)$, and $w^k(\xi)$ with $\sum_\xi w^k(\xi) = 0$, calculate

$$(\widehat{x}^k(\xi), \widehat{u}^k(\xi)) \in \operatorname*{argmin}_{x,u} \varphi^k(x, u, \xi) \quad \text{for each } \xi \in \Xi$$

where
$$\varphi^k(x, u, \xi) = f_0(x, \xi) + \delta_K(F(x, \xi) + u) - w^k(\xi)\cdot x - y^k(\xi)\cdot u + \tfrac{r}{2}||x - x^k||^2 + \tfrac{r}{2}||u||^2$$

and then update by

$$x^{k+1} = \textstyle\sum_\xi p(\xi)\widehat{x}^k(\xi), \qquad y^{k+1}(\xi) = y^k(\xi) - (r - e)\widehat{u}^k(\xi)$$

$$w^{k+1}(\xi) = w^k(\xi) - (r - e)\left[\widehat{x}^k(\xi) - x^{k+1}\right]$$

Further consolidation: carry out the min in u in "closed form"

Refinement Utilizing Augmented Lagrangians

Lagrangian: in minimizing $f_0(x,\xi)$ subject to $F(x,\xi) \in K$

$$L(x,y,\xi) = f_0(x,\xi) + y{\cdot}F(x,\xi) - \delta_Y(y) \quad \text{for } Y = \textbf{polar cone } K^*$$
$$= \min_u \left\{ f_0(x,\xi) + \delta_K(F(x,\xi)+u) - y{\cdot}u \right\}$$

Augmented Lagrangian: for $r > 0$ and $d_Y(y) = \mathrm{dist}(y, Y)$

$$L_r(x,y,\xi) = L(x,y,\xi) + \tfrac{r}{2}||F(x,\xi)||^2 - \tfrac{1}{2r} d_Y^2(y^k(\xi) + rF(x,\xi))$$
$$= \min_u \left\{ f_0(x,\xi) + \delta_K(F(x,\xi)+u) - y{\cdot}u + \tfrac{r}{2}||u||^2 \right\}$$

where moreover $\nabla_y L_r(x,y,\xi) =$ the unique u giving this min

Subminimization in the subproblems:

since $\varphi^k(x,u,\xi) = f_0(x,\xi) + \delta_K(F(x,\xi)+u) - y^k(\xi){\cdot}u + \frac{r}{2}||u||^2$
$$-w^k(\xi){\cdot}x + \tfrac{r}{2}||x - x^k||^2,$$

$$\min_u \varphi^k(x,u,\xi) \;=\; L_r(x, y^k(\xi), \xi) - w^k(\xi){\cdot}x + \tfrac{r}{2}||x - x^k||^2$$

Residual computation:

- minimize the latter expression in x to get $\widehat{x}^k(\xi)$
- then get $\widehat{u}^k(\xi)$ as the gradient $\nabla_y L_r(\widehat{x}^k(\xi), \widehat{y}^k(\xi), \xi)$

Resulting Procedure and its Characteristics

Augmented progressive hedging — parameters $r > e \geq 0$

Having x^k, $y^k(\xi)$, and $w^k(\xi)$ with $\sum_\xi w^k(\xi) = 0$, calculate

$$\widehat{x}^k(\xi) \in \operatorname{argmin}_x \left\{ L_r(x, y^k(\xi), \xi) - w^k(\xi) \cdot x + \tfrac{r}{2} ||x - x^k||^2 \right\},$$
$$\widehat{u}^k(\xi) = \nabla_y L_r(\widehat{x}^k(\xi), \widehat{y}^k(\xi), \xi)$$

and then update by

$$x^{k+1} = \textstyle\sum_\xi p(\xi) \widehat{x}^k(\xi), \qquad y^{k+1}(\xi) = y^k(\xi) - (r - e) \widehat{u}^k(\xi)$$
$$w^{k+1}(\xi) = w^k(\xi) - (r - e) \left[\widehat{x}^k(\xi) - x^{k+1} \right]$$

Key observation: around solution elements $\bar{x}$, $\bar{y}(\xi)$, $\bar{w}(\xi)$ second-order optimality conditions guarantee $\exists\, e$ such that, when $r > e$, the augmented Lagrangian $L_r(x, y, \xi)$ will be convex-concave on a neighborhood of $(\bar{x}, \bar{y}(\xi))$

then the algorithm will converge <u>locally</u> as if in the convex case

References

[1] R.T. Rockafellar (2018) "Progressive decoupling of linkages in optimization and variational inequalities with elicitable convexity or monotonicity," *Set-valued and Variational Analysis*, accepted.

[2] R.T. Rockafellar and S. Uryasev (2018) "Minimizing buffered probability of exceedance by progressive hedging," *Mathematical Programming B*, submitted.

[3] R.T. Rockafellar (2017) "Solving stochastic programming problems with risk measures by progressive hedging," *Set-valued and Variational Analysis* (already published online).

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