PROGRESSIVE HEDGING IN NONCONVEX STOCHASTIC OPTIMIZATION

Terry Rockafellar
University of Washington, Seattle

Smart Energy and Stochastic Optimization
CMO-BIRS — Oaxaca, Mexico
25 September 2019

A Basic Model in Stochastic Optimization

Information pattern: *here single-stage at first*

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

multistage extension: repeated interplay — coming later

Problem (in simplified initial formulation)

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

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

functions $f_i(x,\xi)$ continuous with respect to x

Alternative objectives: (to just minimizing an "expected cost")

- minimizing a CVaR-type measure of risk, or
- minimizing buffered probability of failure at some threshold

these extensions can be subsumed into the expectation model!

Scenario Framework

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

Problem restatement: in reduced form with ∞ penalization

$$\text{minimize } \Phi(x) = E_\xi[\varphi(x,\xi)] = \sum_\xi p(\xi)\varphi(x,\xi) \text{ over } x \in \boldsymbol{R}^n$$

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

The convex case: Φ is lsc convex function on $\boldsymbol{R}^n$ when, for all ξ,

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

but here the nonconvex case will be targeted as well

Relaxation in Terms of Subgradients

Fermat's rule: for minimizing Φ the condition $0 \in \partial\Phi(\bar{x})$ is

- necessary for local optimality at $\bar{x}$ in general,
- sufficient for global optimality at $\bar{x}$ in the convex case

Subgradient calculus: under a minor constraint qualification,

$$\Phi(x) = \sum_\xi p(\xi)\varphi(x,\xi) \quad \Longrightarrow \quad \partial\Phi(x) = \sum_\xi p(\xi)\partial\varphi(x,\xi)$$

Associated first-order optimality condition

$$\forall\xi,\ \exists\, \bar{w}(\xi) \in \partial\varphi(\bar{x},\xi) \text{ such that } 0 = \sum_\xi p(\xi)\bar{w}(\xi) =: E_\xi[\bar{w}(\xi)]$$

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

Computational focus in progressive hedging

find vectors $\bar{x} \in \boldsymbol{R}^n$ and $\bar{w}(\xi) \in \boldsymbol{R}^n$ 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 $E_\xi[w^k(\xi)] = 0$, get

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

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

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

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

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

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

A Special Motivation for Admitting Nonconvexity

Decision-influenced probabilities: $p(\xi) \longrightarrow p(x,\xi)$

$\min_x \sum_\xi p(\xi)\varphi(x,\xi)$ replaced by $\min_x \sum_\xi p(x,\xi)\varphi(x,\xi)$

Example: promotion can affect the demand for a product

Transformation back to the influence-free format:

- let $\widetilde{p}(\xi) = \frac{1}{S}$, where S = the total number of scenarios $\xi \in \Xi$
- introduce $\widetilde{\varphi}(x,\xi) = Sp(x,\xi)\varphi(x,\xi)$, so that

$$\widetilde{p}(\xi)\widetilde{\varphi}(x,\xi) = p(x,\xi)\varphi(x,\xi)$$

- the given problem becomes $\min_x \sum_\xi \widetilde{p}(\xi)\widetilde{\varphi}(x,\xi)$

but this transformation won't preserve convexity!

Conclusion: the capability of solving nonconvex stochastic programming problems will open up treatment of this case

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 — with perturbation vectors

minimize $\Psi(x(\cdot), u(\cdot)) = E_\xi\Big[f_0(x(\xi),\xi) + \delta_K\Big(F(x(\xi),\xi) + u(\xi)\Big)\Big]$

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

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

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

$$\mathcal{S}^\perp = \Big\{(w(\cdot), y(\cdot)) = (w(\xi), y(\xi))_{\xi\in\Xi} \;\Big|\; E_\xi[w(\xi)] = 0\Big\}$$

expectational inner product:

$$\Big\langle (x(\cdot), u(\cdot)), (w(\cdot), y(\cdot)) \Big\rangle = E_\xi\Big[(x(\xi), u(\xi))\cdot(w(\xi), y(\xi))\Big]$$

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

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

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

and then update by

$(x^{k+1}(\xi), u^{k+1}(\xi))_{\xi\in\Xi} =$ projection of $\big(\widehat{x}^k(\xi), \widehat{u}^k(\xi)\big)_{\xi\in\Xi}$ onto $\mathcal{S}$,

$$(w^{k+1}(\xi), y^{k+1}(\xi)) = (w^k(\xi), y^k(\xi)) - (r-e)\big[(\widehat{x}^k(\xi), \widehat{u}^k(\xi)) - (x^{k+1}(\xi), u^{k+1}(\xi))\big]$$

$e =$ elicitation parameter which needs to be "high enough"

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 $E_\xi[w^k(\xi)] = 0$, calculate

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

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

and then update by

$$x^{k+1} = E_\xi[\widehat{x}^k(\xi)], \qquad y^{k+1}(\xi) = y^k(\xi) - (r-e)\widehat{u}^k(\xi)$$

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

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

this will bring augmented Lagrangians into the picture

Toward Refinement Using Augmented Lagrangians

Consider pure scenario problems as auxiliaries:

$$\min\ f_0(x,\xi) \text{ subject to } (f_1(x,\xi),\ldots,f_m(x,\xi)) = F(x,\xi) \in K$$

let Y = polar cone K^* and let $d_Y(y) = \mathrm{dist}(y, Y)$

Associated Lagrangian:

$$L(x,y,\xi) = f_0(x,\xi) + y{\cdot}F(x,\xi) - \delta_Y(y)$$
$$= \min_u\left\{f_0(x,\xi) + \delta_K\big(F(x,\xi)+u\big) - y{\cdot}u\right\}$$

Augmented Lagrangian: with parameter $r > 0$

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

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

often there's a direct formula for this gradient

Example: the case of $K = R^m_-$ and its polar $Y = R^m_+$ has

$$-u_i = \frac{\partial L_r}{\partial y_i}(x,y,\xi) = \begin{cases} f_i(x,\xi) & \text{if } y_i + rf_i(x,\xi) \le 0 \\ -r^{-1}y_i & \text{if } y_i + rf_i(x,\xi) \ge 0 \end{cases}$$

Application to the Algorithm's Subproblems

Augmented Lagrangian formula to utilize:

$$L_r(x,y,\xi) = \min_u \{ f_0(x,\xi) + \delta_K(F(x,\xi)+u) - y{\cdot}u + \frac{r}{2}||u||^2 \}$$

Subminimization in the subproblems: with respect to u

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

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

Residual computation: in executing the (x,u) minimization

- minimize this Lagrangian 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 — with parameters $r > e \geq 0$

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

$$\widehat{x}^k(\xi) \in \mathrm{argmin}_x \Big\{ L_r(x, y^k(\xi), \xi) - w^k(\xi){\cdot}x + \tfrac{r}{2}||x - x^k||^2 \Big\},$$

$$\widehat{u}^k(\xi) = -\nabla_y L_r(\widehat{x}^k(\xi), \widehat{y}^k(\xi), \xi)$$

and then update by

$$x^{k+1} = E_\xi[\widehat{x}^k(\xi)], \qquad y^{k+1}(\xi) = y^k(\xi) - (r - e)\widehat{u}^k(\xi)$$

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

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 locally as if in the convex case

Extension to a Multistage Model

"Decisions" and "observations" in stages $s = 1, \ldots, N$:

$$x_1,\ \xi_1,\ x_2,\ \xi_2, \ldots, x_N,\ \xi_N \quad \text{with} \quad x_s \in \boldsymbol{R}^{n_s},\ \xi_s \in \Xi_s$$

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

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

Nonanticipativity of decisions

x_s can respond to $\xi_1, \ldots, \xi_{s-1}$ but not to $\xi_s, \ldots, \xi_N$:

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

Embedding: $\mathcal{L} =$ **all** functions $x(\cdot)$ from $\xi \in \Xi$ to $x(\xi) \in \boldsymbol{R}^n$

Nonanticipativity subspace: and its complement in $\mathcal{L}$

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

$$\mathcal{N}^\perp = \left\{ w(\cdot) \in \mathcal{L} \,\middle|\, E_{\xi_s, \ldots, \xi_N}[\, w_s(\xi_1, \ldots, \xi_{s-1}, \xi_s \ldots, \xi_N) \,] = 0 \right\}$$

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

Multistage Objective Structure

Relaxation elements: serving as "perturbations"

$$u(\xi) = (u_1(\xi), \ldots, u_N(\xi)) \in \mathbb{R}^{m_1} \times \cdots \times \mathbb{R}^{m_N}$$

Constraint cones: $K_s \subset \mathbb{R}^{m_s}$ in stage s, closed and convex

Objective function: $\Psi(x(\cdot), u(\cdot)) = E_\xi\Big[\sum_{s=1}^N \psi_s(x(\xi), u(\xi), \xi)\Big]$

where $\psi_s(x(\xi), u(\xi), \xi) = f_{s0}(x_1(\xi), \ldots, x_s(\xi), \xi) + \delta_{K_s}\Big(F_s(x_1(\xi), \ldots, x_s(\xi), \xi) + u_s(\xi)\Big)$

Problem

minimize $\Psi(x(\cdot), u(\cdot))$ subject to $x(\cdot) \in \mathcal{N}$, $u(\cdot) = 0$

Treatment: everything in the single-stage case of progressive hedging can be extended to this multistage pattern, including execution with stage-dependent augmented Lagrangians

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*.

[2] R.T. Rockafellar and Jie Sun (2018) "Solving Lagrangian variational inequalities with applications to stochastic programming," *Mathematical Programming B*.

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

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

website: sites.washington.edu/∼rtr/mypage.html