Subregular Recourse in Nonlinear Multistage Stochastic Optimization

Darinka Dentcheva
Stevens Institute of Technology

joint work with

Andrzej Ruszczyński
Rutgers University

One World Optimization Seminar June 14, 2021

Outline

Darinka Dentcheva and Andrzej Ruszczyński. "Subregular recourse in nonlinear multistage stochastic optimization." *Mathematical Programming* (2021): 1–22
DOI https://doi.org/10.1007/s10107-020-01612-z

Objective

Given a probability space $(\Omega, \mathcal{F}, P)$, with $p \in [1, \infty)$, we consider optimization problems with decisions in the space $\mathcal{X} = \mathcal{L}_p(\Omega, \mathcal{F}, P; \mathbb{R}^n)$.

$$
\begin{aligned}
&\min \varphi(x) \\
&\text{ s.t. } F(x) \in Y \quad \text{a.s.}, \\
&\qquad x \in X \quad \text{a.s.}, \\
&\qquad x \in \mathcal{N}.
\end{aligned}
$$

The objective function $\varphi : \mathcal{X} \to \mathbb{R}$ is a Lipschitz continuous functional.
The constraints are given by a nonlinear operator $F : \mathcal{X} \to \mathcal{Y}$, where $\mathcal{Y} = \mathcal{L}_p(\Omega, \mathcal{F}, P; \mathbb{R}^m)$, multifunctions $X : \Omega \rightrightarrows \mathbb{R}^n$ and $Y : \Omega \rightrightarrows \mathbb{R}^m$, and a subspace $\mathcal{N} \subset \mathcal{X}$.
In stochastic optimization and control

$$
F(x)(\omega) = f(x(\omega), \omega), \quad \omega \in \Omega,
$$

where $f : \mathbb{R}^n \times \Omega \to \mathbb{R}^m$ describes the dynamics of the system.

Decomposable and Derivable Sets

Recall that the contingent cone to a closed set $A \subset \mathcal{X}$ at $x \in A$ is the set

$$\mathcal{T}_A(x) = \Big\{ v \in \mathcal{X} : \liminf_{\tau \downarrow 0} \frac{1}{\tau} \operatorname{dist}(x + \tau v, A) = 0 \Big\}.$$

Definition

A set $A \subset \mathcal{X}$ is derivable at $x \in A$ if for every $v \in \mathcal{T}_A(x)$

$$\lim_{\tau \downarrow 0} \frac{1}{\tau} \operatorname{dist}_{\mathcal{X}}(x + \tau v, A) = 0.$$

Definition

A set $\mathcal{K} \subset \mathcal{X}$ is decomposable if a measurable multifunction $K : \Omega \rightrightarrows \mathbb{R}^n$ exists, such that $\mathcal{K} = \big\{ x \in \mathcal{X} : x(\omega) \in K(\omega) \text{ a.s.} \big\}$.

Lemma (Aubin & Frankowska 2009)

Suppose $A \subset \mathcal{X}$ is decomposable and $A(\omega)$ are closed and derivable sets for P-almost all $\omega \in \Omega$. Then

$$\mathcal{T}_A(x) = \Big\{ v \in \mathcal{X} : \text{ for } P\text{-almost all } \omega,\ v(\omega) \in \mathcal{T}_{A(\omega)}\big(x(\omega)\big) \Big\}.$$

Recall that for a cone $\mathcal{K} \subset \mathcal{X}$ its polar cone is defined as follows:

$$\mathcal{K}^\circ = \big\{ y \in \mathcal{X}^* : \langle y, x \rangle \leq 0 \text{ for all } x \in \mathcal{K} \big\}.$$

Here $\mathcal{X}^* = \mathcal{L}_q(\Omega, \mathcal{F}, P; \mathbb{R}^n)$, $1/p + 1/q = 1$ and

$$\langle y, x \rangle = \int_\Omega y(\omega)^\top x(\omega)\ P(d\omega), \quad y \in \mathcal{X}^*, \quad x \in \mathcal{X}.$$

Lemma

The polar cone $\mathcal{K}^\circ$ of a convex decomposable cone $\mathcal{K} \subset \mathcal{X}$ is a convex decomposable cone, and $K^\circ(\omega) = \big(K(\omega)\big)^\circ$ a.s.

Subregular Multifunctions

Consider $0 \in \mathfrak{H}(x)$ where $\mathfrak{H} : \mathcal{X} \rightrightarrows \mathcal{Y}$ and $\mathcal{Y}$ is a Banach space.

Definition

The multifunction $\mathfrak{H}$ is subregular at $\hat{x} \in \mathcal{X}$ with $0 \in \mathfrak{H}(\hat{x})$, if $\delta > 0$ and $C > 0$ exist such that for all $x \in \mathcal{X}$ with $\|x - \hat{x}\|_{\mathcal{X}} \le \delta$ a point $\tilde{x}$ exists such that

$$0 \in \mathfrak{H}(\tilde{x}) \quad \text{and} \quad \|\tilde{x} - x\|_{\mathcal{X}} \le C \,\text{dist}(0, \mathfrak{H}(x)).$$

For $F(x) \in Y$ ($\mathfrak{H} = F(x) - Y$) where $\mathcal{Y}$ is an $\mathcal{L}_p$-space, $F : \mathcal{X} \to \mathcal{Y}$ is Lipschitz continuous, and $Y \subset \mathcal{Y}$ is a closed convex set, subregularity means that a constant C exists, such that for all x in a neighborhood of $\hat{x}$,

$$\text{dist}\left(x, F^{-1}(Y)\right) \le C \,\text{dist}\left(F(x), Y\right).$$

It is equivalent to the calmness of the multifunction

$$M(z) = \left\{x : F(x) \in Y - z\right\}$$

at the point $(0, \hat{x})$ with calmness being defined by Robinson in 1979 under the name of the "upper Lipschiz property".

Causal Operators

For a probability space $(\Omega, \mathcal{F}, P)$ with filtration $\{\emptyset, \Omega\} = \mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots \subset \mathcal{F}_T = \mathcal{F}$, we define the spaces $\mathcal{X}_t = \mathcal{L}_p(\Omega, \mathcal{F}_t, P; \mathbb{R}^n)$ and $\mathcal{Y}_t = \mathcal{L}_p(\Omega, \mathcal{F}_t, P; \mathbb{R}^m)$ with $p \in [1, \infty)$.
Let $\mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_T$ and $\mathcal{Y} = \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_T$.
We use $x_{1:t} = (x_1, \ldots, x_t)$, and $\mathcal{X}_{1:t} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_t$.

Definition

An operator $F : \mathcal{X} \to \mathcal{Y}$ is causal, if functions $f_t : \mathbb{R}^{nt} \times \Omega \to \mathbb{R}^m$ exist, such that for all $t = 1 \ldots T$, $f_t(\cdot, \cdot)$ is superpositionally measurable and

$$F_t(x)(\omega) = f_t(x_{1:t}(\omega), \omega), \quad \omega \in \Omega.$$

Assumption A

For all $t = 1, \ldots, T$:

(i) $f_t(\xi, \cdot)$ is an element of $\mathcal{Y}_t$ for all $\xi \in \mathbb{R}^{nt}$;

(ii) For almost all $\omega \in \Omega$, $f_t(\cdot, \omega)$ is continuously differentiable, with the Jacobian $f_t'(\cdot, \omega)$;

(iii) A constant C_f exists, such that $\|f_t'(\cdot, \omega)\| \leq C_f$, almost surely.

Differentiability of Causal Operators

Lemma

A causal operator $F(\cdot)$ satisfying Assumption A is Gâteaux differentiable with the derivative $F'(x)$ defined by

$$[F'(x)\,h](\omega) = f'(x(\omega),\omega)\,h(\omega), \quad \omega \in \Omega.$$

Theorem

Suppose $F(\cdot)$ is a causal operator satisfying Assumption A and $\rho : \mathcal{Y} \to \mathbb{R}$ is a convex subdifferentiable functional. Then the Clarke subdifferential of the composition has the form:

$$\partial(\rho \circ F)(x) = [F'(x)]^* \,\partial\rho(F(x))$$

where $\big[F'(x)\big]^*$ is the adjoint operator to the Gâteaux derivative $F'(x)$.

Corollary Suppose additionally $Y \subset \mathcal{Y}$ is convex and closed, $F(x) \in Y$. Then

$$\partial \operatorname{dist}\big(F(x), Y\big) = \big[F'(x)\big]^* \big(N_Y(F(x)) \cap \mathbb{B}_{\mathcal{Y}^*}\big),$$

where $\mathbb{B}_{\mathcal{Y}^*}$ is the closed unit ball in $\mathcal{Y}^*$.

Example: Nonlinear causal operators are not Fréchet differentiable

Let $\Omega = [0,1]$ and P be the Lebesgue measure on $[0,1]$. We define the spaces $\mathcal{X} = \mathcal{Y} = \mathcal{L}_1(\Omega, \mathcal{F}, P)$, and the operator $F : \mathcal{X} \to \mathcal{Y}$ given by

$$F(x)(\omega) = f(x(\omega), \omega) = \begin{cases} \big(x(\omega)\big)^2 & \text{if } -1 \le x(\omega) \le 1, \\ 2\big|x(\omega)\big| - 1 & \text{otherwise.} \end{cases}$$

Note that $\|F(x)\| \le 2\|x\|$, and thus indeed $F : \mathcal{X} \to \mathcal{Y}$.

$F(\cdot)$ is not Fréchet differentiable at 0.

The Gâteaux derivative of F at 0 is 0, and, thus, the Fréchet derivative, if it existed, would be $F'(0) = 0$. Consider the sequence of functions

$$x_n(\omega) = \begin{cases} 1 & \text{if } 0 \le \omega \le \frac{1}{n}, \\ 0 & \text{otherwise,} \end{cases} \qquad n = 1, 2, \ldots$$

We have $\big\|x_n\big\| = \frac{1}{n}$ and thus $x_n \to 0$. By construction, $F(x_n) = x_n$, $F(0) = 0$, and then, by the definition of the Fréchet derivative, we would have

$$0 = \lim_{n\to\infty} \frac{F(x_n) - F(0) - F'(0)x_n}{\|x_n\|} = \lim_{n\to\infty} \frac{x_n}{\|x_n\|}.$$

This is a contradiction.

Multistage Stochastic Optimization with Built-In Nonanticipativity

$$\min \varphi(x_{1:T}) \tag{1}$$

$$\text{s.t. } F_t(x_{1:t}) \in Y_t \quad \text{a.s.}, \quad t = 1, \ldots, T, \tag{2}$$

$$x_t \in X_t \quad \text{a.s.}, \quad t = 1, \ldots, T. \tag{3}$$

Due to the causality of $F(\cdot)$ and the decomposability of Y, (2) can be written as

$$f_t(x_{1:t}(\omega), \omega) \in Y_t(\omega), \quad t = 1, \ldots, T, \quad \omega \in \Omega.$$

$X_t : \Omega \rightrightarrows \mathbb{R}^n$ are $\mathcal{F}_t$-measurable mulitifunctions with closed convex images.

Sub-regular recourse definition

$$f_t(\zeta_{1:t-1}, \xi, \omega) \in Y_t(\omega), \tag{4}$$

$$\xi \in X_t(\omega). \tag{5}$$

admits a complete subregular recourse, if a constant C exist, such that for almost all $\omega \in \Omega$, every $\zeta_{1:t-1} \in X_{1:t-1}(\omega)$ and every $\eta \in \mathbb{R}^n$, a solution ξ exists, satisfying

$$\|\xi - \eta\| \leq C\big(\mathrm{d}(f_t(\zeta_{1:t-1}, \eta, \omega), Y_t(\omega)) + \mathrm{d}(\eta, X_t(\omega))\big)$$

$\zeta_{1:t-1}$ represents the history of decisions at the particular elementary event.

Optimality conditions

Under Assumption A, we denote:
$F_t'(\hat{x}_{1:t}) = A_t = \big(A_{t,1}, \dots, A_{t,t}\big), \quad t = 1, \dots, T$, with partial Jacobians
$A_{t,\ell} : \mathcal{X}_\ell \to \mathcal{Y}_t$, $A_{t,\ell} = \frac{\partial F_t(\hat{x}_{1:t})}{\partial x_\ell}, \quad \ell = 1, \dots, t, \quad t = 1, \dots, T.$
These linear operators are defined pointwise:

$$A_{t,\ell}(\omega) = \frac{\partial f_t(\hat{x}_{1:t}(\omega), \omega)}{\partial x_\ell(\omega)}, \quad \ell = 1, \dots, t, \quad t = 1, \dots, T, \quad \omega \in \Omega.$$

Theorem

If the system (4)–(5) admits complete subregular recourse, then the system (2)–(3) is subregular at any feasible point $\hat{x} = (\hat{x}_1, \dots, \hat{x}_T)$. If the policy $\hat{x}$ is a local minimum of problem (1)–(3), then a subgradient $\hat{g} \in \partial\varphi(\hat{x})$, multipliers $\hat{\psi}_t \in N_{Y_t}(F_t(\hat{x}_{1:t}))$, and normal elements $\hat{n}_t \in N_{X_t}(\hat{x}_t)$, exist, such that for P-almost all $\omega \in \Omega$ we have:

$$\hat{g}_t + A_{t,t}^\top \hat{\psi}_t + \mathbb{E}_t\Bigg[\sum_{\ell=t+1}^{T} A_{\ell,t}^\top \hat{\psi}_\ell\Bigg] + \hat{n}_t = 0, \quad t = 1, \dots, T. \tag{6}$$

Multistage Optimization with Nonanticipativity Constraints

We consider extended spaces $\widetilde{\mathcal{X}}_t = \mathcal{L}_p(\Omega, \mathcal{F}, P; \mathbb{R}^n)$, $t = 1, \dots, T$ and a relaxed policy $x = (x_1, \dots, x_T) \in \widetilde{\mathcal{X}}_1 \times \cdots \times \widetilde{\mathcal{X}}_T = \widetilde{\mathcal{X}}$.
The nonaticipativity constraint defines a closed subspace $\mathcal{N}$ in $\widetilde{\mathcal{X}}$:

$$x_t = \mathbb{E}[x_t | \mathcal{F}_t], \quad t = 1, \dots, T. \tag{7}$$

We denote by $\widetilde{\varphi} : \widetilde{\mathcal{X}} \to \mathbb{R}$ a Lipschitz continuous extension of φ, that is, $\widetilde{\varphi}(x) = \varphi(x)$ for all $x \in \mathcal{N}$.
A causal operator $F(\cdot)$ has value space $\widetilde{\mathcal{Y}} = \widetilde{\mathcal{Y}}_1 \times \cdots \times \widetilde{\mathcal{Y}}_T$ with $\widetilde{\mathcal{Y}}_t = \mathcal{L}_p(\Omega, \mathcal{F}, P; \mathbb{R}^m)$. The decomposable sets X_t and Y_t are viewed as subsets $\widetilde{X}_t$ of $\widetilde{\mathcal{X}}_t$ and $\widetilde{Y}_t$ of $\widetilde{\mathcal{Y}}_t$:

$$\widetilde{X}_t = \{x_t \in \widetilde{\mathcal{X}}_t : x_t(\omega) \in X_t(\omega) \text{ a.s.}\},$$
$$\widetilde{Y}_t = \{y_t \in \widetilde{\mathcal{Y}}_t : y_t(\omega) \in Y_t(\omega) \text{ a.s.}\}, \quad t = 1, \dots, T.$$

$$\min \widetilde{\varphi}(x_1, \dots, x_T) \tag{8}$$
$$\text{s.t. } x_t - \mathbb{E}_t x_t = 0 \quad \text{a.s.}, \quad t = 1, \dots, T, \tag{9}$$
$$F_t(x_{1:t}) \in \widetilde{Y}_t \quad \text{a.s.}, \quad t = 1, \dots, T, \tag{10}$$
$$x_t \in \widetilde{X}_t \quad \text{a.s.}, \quad t = 1, \dots, T. \tag{11}$$

Optimality conditions

Theorem

If the system (4)–(5) admits complete subregular recourse, then the system (9)–(11) is subregular at any feasible point $\hat{x} = (\hat{x}_1, \dots, \hat{x}_T)$. Additionally, if a policy $\hat{x}$ is a local minimum of problem (8)–(11), then a subgradient $\tilde{g} \in \partial\widetilde{\varphi}(\hat{x})$, multipliers $\lambda_t \in \widetilde{\mathcal{X}}_t^*$, $\widetilde{\psi}_t \in N_{\widetilde{Y}_t}(F_t(\hat{x}_{1:t}))$, $t = 1, \dots, T$, and normal elements $\tilde{n}_t \in N_{\widetilde{X}_t}(\hat{x}_t)$, $t = 1, \dots, T$, exist, such that for P-almost all $\omega \in \Omega$ we have:

$$\tilde{g}_t + \lambda_t + \sum_{\ell=t}^{T} A_{\ell,t}^\top \widetilde{\psi}_\ell + \tilde{n}_t = 0, \quad t = 1, \dots, T,$$

$$\mathbb{E}_t[\lambda_t] = 0, \quad t = 1, \dots, T.$$

Corollary

The subgradient $\hat{g} \in \partial\varphi(\hat{x})$ given by $\hat{g}_t = \mathbb{E}_t[\tilde{g}_t]$, $t = 1, \dots, T$, together with the multipliers $\hat{\psi}_t = \mathbb{E}_t[\widetilde{\psi}_t]$, $t = 2, \dots, T$, and normal vectors $\hat{n}_t = \mathbb{E}_t[\tilde{n}_t]$ satisfy the optimality conditions (6).

Conclusion and Future Work

- The concept of subregular recourse allows for the verification of the subregularity in abstract spaces by establishing subsegularity of finite-dimensional systems associated with each stage and each elementary event.
- The Clarke subdifferential of a composition of the distance function and a causal operator allows for exact subdifferentiation of the penalty function associated with system's dynamics.
- This approach has potential in addressing nonlinear stochastic dynamic optimization problems.
- Focus on specific dynamic risk measures and exploit their specific structure to obtain specific optimality conditions.
- Adopt Kruger-Mordukhovich calculus by restricting the spaces $\mathcal{X}$ and $\mathcal{Y}$ to those with $p \in (1, \infty)$. This would allow for the treatment of non-convex sets X_t and Y_t, and more accurate subdifferentials of non-convex objective functionals. The main challenge is to derive the explicit form of the coderivative of the operator $F(\cdot)$ describing the dynamics of the system.
- The necessary conditions of optimality are a prerequisite for the development of numerical methods.