Rockafellian Relaxation in Optimization under Uncertainty: Asymptotically Exact Formulations

Johannes O. Royset
Operations Research Department
Naval Postgraduate School, Monterey, California

with Louis Chen and Eric Eckstrand

Supported by AFOSR
Erice, May 20, 2022

Distributional ambiguity

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \ \mathbb{E}_P\left[g(\boldsymbol{\xi}, x)\right]$$

Distributional ambiguity

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \ \mathbb{E}_P\big[g(\boldsymbol{\xi}, x)\big]$$

If P is replaced by approximating P^ν, would the corresponding solutions be close?

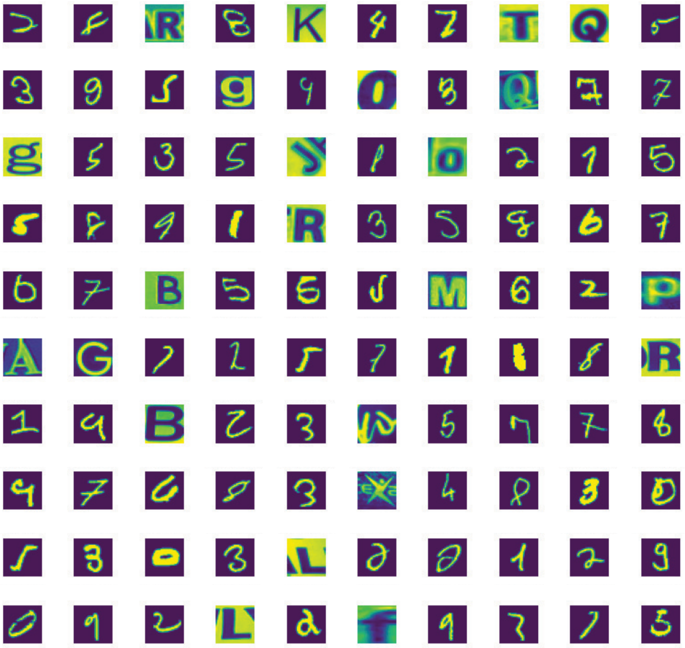
Distributional ambiguity

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \ \mathbb{E}_P\big[g(\boldsymbol{\xi}, x)\big]$$

If P is replaced by approximating P^ν, would the corresponding solutions be close?

- statistical noise; P^ν empirical distribution
- sensitivity analysis
- upweighing, influence functions, adversarial attacks
- corruption, contamination, outliers

MNIST (numbers) corrupted with Chars74k (letters)



Lack of convergence of expectations

Let $g(\xi, x) = \xi x + \frac{1}{2}(1 - x)$

$$\underset{x \in [0,1]}{\text{minimize}} \ \mathbb{E}_P\big[g(\boldsymbol{\xi}, x)\big]$$

P assigns probability 1 to $\xi = 0$

Unique minimizer at 1

Lack of convergence of expectations

Let $g(\xi, x) = \xi x + \frac{1}{2}(1 - x)$

$$\underset{x \in [0,1]}{\text{minimize}} \; \mathbb{E}_P\big[g(\boldsymbol{\xi}, x)\big]$$

P assigns probability 1 to $\xi = 0$

Unique minimizer at 1

$$\underset{x \in [0,1]}{\text{minimize}} \; \mathbb{E}_{P^\nu}\big[g(\boldsymbol{\xi}, x)\big]$$

P^ν assigns prob. $1 - \frac{1}{\nu}$ to $\xi = 0$; assigns prob. $\frac{1}{\nu}$ to $\xi = \nu$

Unique minimizer at 0

Lack of convergence of expectations

Let $g(\xi, x) = \xi x + \frac{1}{2}(1 - x)$

$$\underset{x \in [0,1]}{\text{minimize}}\ \mathbb{E}_P\left[g(\boldsymbol{\xi}, x)\right]$$

P assigns probability 1 to $\xi = 0$

Unique minimizer at 1

$$\underset{x \in [0,1]}{\text{minimize}}\ \mathbb{E}_{P^\nu}\left[g(\boldsymbol{\xi}, x)\right]$$

P^ν assigns prob. $1 - \frac{1}{\nu}$ to $\xi = 0$; assigns prob. $\frac{1}{\nu}$ to $\xi = \nu$

Unique minimizer at 0

But, P^ν converges weakly to P

Fairness constraints in learning

Random vector $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$: features, labels, sensitive attributes

$$\begin{aligned} \underset{a,\alpha}{\text{minimize}}\ & \mathbb{E}\Big[\max\big\{0, 1 - \boldsymbol{y}\big(\langle a, \boldsymbol{x}\rangle + \alpha\big)\big\}\Big] + \|a\|_2^2 \\ \text{subject to}\ & \mathbb{E}\Big[(\boldsymbol{z} - \bar{z})\big(\langle a, \boldsymbol{x}\rangle + \alpha\big)\Big] \le \tau \end{aligned}$$

Fairness constraints in learning

Random vector $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$: features, labels, sensitive attributes

$$\begin{aligned} \underset{a,\alpha}{\text{minimize}}\ & \mathbb{E}\Big[\max\big\{0, 1 - \boldsymbol{y}\big(\langle a, \boldsymbol{x}\rangle + \alpha\big)\big\}\Big] + \|a\|_2^2 \\ \text{subject to}\ & \mathbb{E}\Big[(\boldsymbol{z} - \bar{z})\big(\langle a, \boldsymbol{x}\rangle + \alpha\big)\Big] \le \tau \end{aligned}$$

Distribution of $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$:

$$\begin{cases} (-1, -1, 0) & \text{with probability } 1/2 \\ (1, 1, 1) & \text{with probability } 1/2 \end{cases}$$

Produces set of minimizers: $\{1/2\} \times [-1/2, 1/2]$

Fairness constraints in learning

Random vector $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$: features, labels, sensitive attributes

$$\begin{aligned} \underset{a,\alpha}{\text{minimize}} \; & \mathbb{E}\Big[\max\big\{0, 1 - \boldsymbol{y}\big(\langle a, \boldsymbol{x}\rangle + \alpha\big)\big\}\Big] + \|a\|_2^2 \\ \text{subject to} \; & \mathbb{E}\Big[(\boldsymbol{z} - \bar{z})\big(\langle a, \boldsymbol{x}\rangle + \alpha\big)\Big] \le \tau \end{aligned}$$

Distribution of $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$:

$$\begin{cases} (-1, -1, 0) & \text{with probability } 1/2 \\ (1, 1, 1) & \text{with probability } 1/2 \end{cases}$$

Produces set of minimizers: $\{1/2\} \times [-1/2, 1/2]$

Approximating distribution of $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$:

$$\begin{cases} (-1, -1, 0) & \text{with probability } 1/2 \\ (1, 1, 1) & \text{with probability } 1/2 - 1/\nu \\ (\nu, 1, 1) & \text{with probability } 1/\nu \end{cases}$$

Produces minimizer: $(1/4, -3/4)$

More trouble...

$$\underset{x\in\mathbb{R}^n}{\text{minimize}}\ f_0(x) + \sum_{i=1}^{s} p_i f_i(x)$$

Two-stage stochastic problems without complete recourse

If $f_i(x) = \infty$: changing $p_i = 0$ has big effect

More trouble...

$$\underset{x \in \mathbb{R}^n}{\text{minimize}}\ f_0(x) + \sum_{i=1}^{s} p_i f_i(x)$$

Two-stage stochastic problems without complete recourse

If $f_i(x) = \infty$: changing $p_i = 0$ has big effect

Classification problems with a cross-entropy loss

$g(x; \xi, \eta) =$ prob. that model x assigns to label η given feature ξ

Then, $f_i(x) = -\log g(x; \xi_i, \eta_i)$

More trouble...

$$\underset{x\in\mathbb{R}^n}{\text{minimize}}\ f_0(x) + \sum\nolimits_{i=1}^{s} p_i f_i(x)$$

Two-stage stochastic problems without complete recourse

If $f_i(x) = \infty$: changing $p_i = 0$ has big effect

Classification problems with a cross-entropy loss

$g(x;\xi,\eta) =$ prob. that model x assigns to label η given feature ξ

Then, $f_i(x) = -\log g(x;\xi_i,\eta_i)$

Counting losses, AUC, reliability analysis produce

$$f_i(x) = \begin{cases} 1 & \text{if state}(\xi_i, x) > 0 \\ 0 & \text{otherwise} \end{cases}$$

Summary

Replacing

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \ \mathbb{E}_P\big[g(\boldsymbol{\xi}, x)\big]$$

by

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \ \mathbb{E}_{P^\nu}\big[g(\boldsymbol{\xi}, x)\big]$$

may have outsized effect

Consider alternatives based on Rockafellian relaxation

Perspectives

Conservative approaches under distributional ambiguity:

Robust optimization, adversarial training, diametrical risk

Motivation: downward bias in sampling; attacks

Perspectives

Conservative approaches under distributional ambiguity:

Robust optimization, adversarial training, diametrical risk

Motivation: downward bias in sampling; attacks

Our approach is **optimistic** based on **relaxation**

Recall: for nonnegative X^ν converging in distribution to X:

$$\liminf \mathbb{E}[X^\nu] \geq \mathbb{E}[X] \qquad \text{(possibly strict)}$$

Perspectives

Conservative approaches under distributional ambiguity:

Robust optimization, adversarial training, diametrical risk

Motivation: downward bias in sampling; attacks

Our approach is **optimistic** based on **relaxation**

Recall: for nonnegative X^ν converging in distribution to X:

$$\liminf \mathbb{E}[X^\nu] \geq \mathbb{E}[X] \qquad \text{(possibly strict)}$$

Approximation is too high: need relaxation not restriction

Rockafellians

For $\varphi : \mathbb{R}^n \to \overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$,

$$\operatorname*{minimize}_{x \in \mathbb{R}^n} \varphi(x)$$

Rockafellians

For $\varphi : \mathbb{R}^n \to \overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$,

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \; \varphi(x)$$

A (bivariate) function $f : \mathbb{R}^m \times \mathbb{R}^n \to \overline{\mathbb{R}}$ is a *Rockafellian* for the problem, anchored at $\bar{u}$, when

$$f(\bar{u}, x) = \varphi(x) \quad \forall x \in \mathbb{R}^n$$

Example: stochastic optimization

Concerns about changes to $p = (p_1, \ldots, p_s)$ in

$$\underset{x \in \mathbb{R}^n}{\text{minimize}}\ \varphi(x) = f_0(x) + \sum_{i=1}^{s} p_i f_i(x)$$

Example: stochastic optimization

Concerns about changes to $p = (p_1, \ldots, p_s)$ in

$$\operatorname*{minimize}_{x \in \mathbb{R}^n} \varphi(x) = f_0(x) + \sum\nolimits_{i=1}^{s} p_i f_i(x)$$

May lead to study of the Rockafellian

$$f(u, x) = f_0(x) + \sum\nolimits_{i=1}^{s} (p_i + u_i) f_i(x)$$

with anchor at $\bar{u} = 0$

Example: stochastic optimization

Concerns about changes to $p = (p_1, \ldots, p_s)$ in

$$\underset{x \in \mathbb{R}^n}{\text{minimize}}\ \varphi(x) = f_0(x) + \sum\nolimits_{i=1}^{s} p_i f_i(x)$$

May lead to study of the Rockafellian

$$f(u, x) = f_0(x) + \sum\nolimits_{i=1}^{s} (p_i + u_i) f_i(x)$$

with anchor at $\bar{u} = 0$

How do solutions of $\text{minimize}_x\, f(u, x)$ change for u near 0?

General framework

Actual problem: $\underset{x\in\mathbb{R}^n}{\text{minimize}}\ \varphi(x)$

General framework

$$\textbf{Actual problem}: \quad \underset{x \in \mathbb{R}^n}{\text{minimize}}\ \varphi(x)$$

Rockafellian $f : \mathbb{R}^m \times \mathbb{R}^n \to \overline{\mathbb{R}}$ with anchor at $\bar{u}$; vector $\bar{y}$

$$\textbf{Rockafellian relaxation}: \quad \underset{u \in \mathbb{R}^m, x \in \mathbb{R}^n}{\text{minimize}}\ f(u, x) - \langle \bar{y}, u - \bar{u} \rangle$$

General framework

Actual problem: $\underset{x \in \mathbb{R}^n}{\text{minimize}}\ \varphi(x)$

Rockafellian $f : \mathbb{R}^m \times \mathbb{R}^n \to \overline{\mathbb{R}}$ with anchor at $\bar{u}$; vector $\bar{y}$

Rockafellian relaxation: $\underset{u \in \mathbb{R}^m, x \in \mathbb{R}^n}{\text{minimize}}\ f(u, x) - \langle \bar{y}, u - \bar{u} \rangle$

Approximating $f^\nu : \mathbb{R}^m \times \mathbb{R}^n \to \overline{\mathbb{R}}$; approximating vector y^ν

Approximating problem: $\underset{u \in \mathbb{R}^m, x \in \mathbb{R}^n}{\text{minimize}}\ f^\nu(u, x) - \langle y^\nu, u - \bar{u} \rangle$

Exact Rockafellian

The Rockafellian f is *exact* for $\bar{y} \in \mathbb{R}^m$ when

$$x^\star \in \operatorname{argmin}_x \varphi(x) \implies (\bar{u}, x^\star) \in \operatorname{argmin}_{u,x} f(u, x) - \langle \bar{y}, u - \bar{u} \rangle$$

Exact Rockafellian

The Rockafellian f is *exact* for $\bar{y} \in \mathbb{R}^m$ when

$$x^\star \in \operatorname{argmin}_x \varphi(x) \implies (\bar{u}, x^\star) \in \operatorname{argmin}_{u,x} f(u,x) - \langle \bar{y}, u - \bar{u} \rangle$$

The Rockafellian is *strictly exact* for $\bar{y}$ when, in addition, it satisfies

$$(u^\star, x^\star) \in \operatorname{argmin}_{u,x} f(u,x) - \langle \bar{y}, u - \bar{u} \rangle \implies u^\star = \bar{u},\ x^\star \in \operatorname{argmin}_x \varphi(x)$$

Asymptotically exact Rockafellians

The approximations $\{f^\nu, \nu \in \mathbb{N}\}$ are *asymptotically exact Rockafellians* if they **epi-converge** to an exact Rockafellian f for the actual problem

Asymptotically exact Rockafellians

The approximations $\{f^\nu, \nu \in \mathbb{N}\}$ are *asymptotically exact Rockafellians* if they **epi-converge** to an exact Rockafellian f for the actual problem

The approximations $\{f^\nu, \nu \in \mathbb{N}\}$ are *asymptotically strictly exact Rockafellians* if f is strictly exact

Convergence under asymptotic exactness

Suppose that $\operatorname{dom}\varphi \neq \emptyset$ and that $\{f^\nu, \nu \in \mathbb{N}\}$ are asymptotically strictly exact Rockafellians for $\bar{y} \in \mathbb{R}^m$. Let $y^\nu \to \bar{y}$, $\varepsilon^\nu \searrow 0$, and

$$(u^\nu, x^\nu) \in \varepsilon^\nu\text{-}\operatorname{argmin}_{u,x} \left\{ f^\nu(u,x) - \langle y^\nu, u \rangle \right\}.$$

Then, every cluster point $(\hat{u}, \hat{x})$ of $\{(u^\nu, x^\nu), \nu \in \mathbb{N}\}$ satisfies

$$\hat{x} \in \operatorname{argmin}_x \varphi(x)$$

Convergence under asymptotic exactness

Suppose that $\operatorname{dom}\varphi \neq \emptyset$ and that $\{f^\nu, \nu \in \mathbb{N}\}$ are asymptotically strictly exact Rockafellians for $\bar{y} \in \mathbb{R}^m$. Let $y^\nu \to \bar{y}$, $\varepsilon^\nu \searrow 0$, and

$$(u^\nu, x^\nu) \in \varepsilon^\nu\text{-}\operatorname{argmin}_{u,x} \left\{ f^\nu(u,x) - \langle y^\nu, u \rangle \right\}.$$

Then, every cluster point $(\hat{u}, \hat{x})$ of $\{(u^\nu, x^\nu), \nu \in \mathbb{N}\}$ satisfies

$$\hat{x} \in \operatorname{argmin}_x \varphi(x)$$

If asymptotic strict exactness is replaced by asymptotic exactness, then $(\hat{u}, \hat{x})$ satisfies a necessary optimality condition for the actual problem

Stochastic optimization with distributional ambiguity

$$\underset{x \in \mathbb{R}^n}{\text{minimize}}\ \varphi(x) = f_0(x) + \sum_{i=1}^{s} p_i f_i(x) \qquad \text{replace } p \text{ by } p^\nu?$$

Stochastic optimization with distributional ambiguity

$$\underset{x \in \mathbb{R}^n}{\text{minimize}}\ \varphi(x) = f_0(x) + \sum_{i=1}^{s} p_i f_i(x) \qquad \text{replace } p \text{ by } p^\nu?$$

Adopt the Rockafellian

$$f(u, x) = f_0(x) + \sum_{i=1}^{s} (p_i + u_i) f_i(x) + \iota_{\{0\}^s}(u)$$

with anchor at $\bar{u} = 0$ $\quad$ ($\iota_C(u) = 0$ if $u \in C$; infinity otherwise)

Stochastic optimization with distributional ambiguity

$$\underset{x \in \mathbb{R}^n}{\text{minimize}}\ \varphi(x) = f_0(x) + \sum_{i=1}^{s} p_i f_i(x) \qquad \text{replace } p \text{ by } p^\nu?$$

Adopt the Rockafellian

$$f(u, x) = f_0(x) + \sum_{i=1}^{s} (p_i + u_i) f_i(x) + \iota_{\{0\}^s}(u)$$

with anchor at $\bar{u} = 0$ ($\iota_C(u) = 0$ if $u \in C$; infinity otherwise)

Adopt approximations

$$f^\nu(u, x) = f_0(x) + \sum_{i=1}^{s} (p_i^\nu + u_i) f_i(x) + \tfrac{1}{2}\theta^\nu \|u\|_2^2 + \iota_\Delta(p^\nu + u)$$

where $\Delta = \{q \in \mathbb{R}^s \mid \sum_{i=1}^{s} q_i = 1,\ q_i \geq 0\}$

Stochastic optimization with distributional ambiguity

$$\operatorname*{minimize}_{x\in\mathbb{R}^n} \varphi(x) = f_0(x) + \sum_{i=1}^{s} p_i f_i(x) \qquad \text{replace } p \text{ by } p^\nu?$$

Adopt the Rockafellian

$$f(u,x) = f_0(x) + \sum_{i=1}^{s}(p_i + u_i)f_i(x) + \iota_{\{0\}^s}(u)$$

with anchor at $\bar{u} = 0$ $\quad(\iota_C(u) = 0$ if $u \in C$; infinity otherwise)

Adopt approximations

$$f^\nu(u,x) = f_0(x) + \sum_{i=1}^{s}(p_i^\nu + u_i)f_i(x) + \tfrac{1}{2}\theta^\nu\|u\|_2^2 + \iota_\Delta(p^\nu + u)$$

where $\Delta = \{q \in \mathbb{R}^s \mid \sum_{i=1}^{s} q_i = 1,\ q_i \geq 0\}$

Solve the approximating problem

$\operatorname*{minimize}_{u\in\mathbb{R}^s, x\in\mathbb{R}^n} f^\nu(u,x) - \langle y^\nu, u\rangle$ **instead of** $\operatorname*{minimize}_{x\in\mathbb{R}^n} f_0(x) + \sum_{i=1}^{s} p_i^\nu f_i(x)$

Exactness and asymptotic exactness

If $f_i : \mathbb{R}^n \to \overline{\mathbb{R}}$, $i = 0, 1, \ldots, s$ proper lsc, then

- f is strictly exact for any $\bar{y}$
- $\{f^\nu, \nu \in \mathbb{N}\}$ are asymptotically exact provided that $\theta^\nu \to \infty$ and $\theta^\nu \|p^\nu - p\|_2^2 \to 0$

Rate of convergence

If (u^ν, x^ν) minimizes approximating problem,

then there is a positive constant σ such that

$$\mathrm{dist}\left(x^\nu,\ 2\eta^\nu\text{-}\operatorname{argmin}_x f_0(x) + \sum_{i=1}^{s} p_i f_i(x)\right) \le \eta^\nu = \sigma \| p^\nu - p \|_2^{2/3}$$

for all sufficiently large ν

Interpretation

Approximating problem

$$\underset{u\in\mathbb{R}^s, x\in\mathbb{R}^n}{\text{minimize}}\ f^\nu(u,x) - \langle y^\nu, u\rangle$$

reduces to

$$\underset{x\in\mathbb{R}^n}{\text{minimize}}\ f_0(x) + \sum\nolimits_{i=1}^{s} p_i^\nu f_i(x) - r^\nu(x)$$

Interpretation

Approximating problem

$$\operatorname*{minimize}_{u\in\mathbb{R}^s, x\in\mathbb{R}^n} f^\nu(u,x) - \langle y^\nu, u\rangle$$

reduces to

$$\operatorname*{minimize}_{x\in\mathbb{R}^n} f_0(x) + \sum_{i=1}^{s} p_i^\nu f_i(x) - r^\nu(x)$$

with "regularizer"

$$r^\nu(x) = \min_{w\in\mathbb{R}^s} \left\{ \max_{i=1,\dots,s} w_i - \langle p^\nu, w\rangle + \frac{1}{2\theta^\nu}\left\|y^\nu - F(x) - w\right\|_2^2 \right\}$$

where $F(x) = \big(f_1(x), \dots, f_s(x)\big)$

Interpretation

Approximating problem

$$\operatorname*{minimize}_{u\in\mathbb{R}^s, x\in\mathbb{R}^n} f^\nu(u,x) - \langle y^\nu, u\rangle$$

reduces to

$$\operatorname*{minimize}_{x\in\mathbb{R}^n} f_0(x) + \sum_{i=1}^{s} p_i^\nu f_i(x) - r^\nu(x)$$

with "regularizer"

$$r^\nu(x) = \min_{w\in\mathbb{R}^s}\left\{ \max_{i=1,\dots,s} w_i - \langle p^\nu, w\rangle + \frac{1}{2\theta^\nu}\left\|y^\nu - F(x) - w\right\|_2^2\right\}$$

where $F(x) = \big(f_1(x), \dots, f_s(x)\big)$

If F smooth, then r^ν is continuously differentiable

Refinement 1: Φ-divergence

New approximation

$$f^{\nu}(u,x) = f_0(x) + \sum_{i=1}^{s} (p_i^{\nu} + u_i) f_i(x) + \theta^{\nu} d_{\Phi}(p^{\nu} + u | p^{\nu}) + \iota_{\Delta}(p^{\nu} + u)$$

instead of

$$f^{\nu}(u,x) = f_0(x) + \sum_{i=1}^{s} (p_i^{\nu} + u_i) f_i(x) + \tfrac{1}{2}\theta^{\nu} \|u\|_2^2 + \iota_{\Delta}(p^{\nu} + u)$$

Refinement 1: Φ-divergence

New approximation

$$f^\nu(u,x) = f_0(x) + \sum_{i=1}^{s} (p_i^\nu + u_i) f_i(x) + \theta^\nu d_\Phi(p^\nu + u | p^\nu) + \iota_\Delta(p^\nu + u)$$

instead of

$$f^\nu(u,x) = f_0(x) + \sum_{i=1}^{s} (p_i^\nu + u_i) f_i(x) + \tfrac{1}{2}\theta^\nu \|u\|_2^2 + \iota_\Delta(p^\nu + u)$$

Asymptotic strict exactness when

$\theta^\nu \to \infty$

$\theta^\nu \|p^\nu - p\|_2 \to 0$

Refinement 2: support ambiguity

Actual problem

$$\operatorname*{minimize}_{x\in\mathbb{R}^n} f_0(x) + \sum\nolimits_{i=1}^{s} p_i g(\xi_i, x)$$

What if p replaced by p^ν and ξ_i by ξ_i^ν?

Refinement 2: support ambiguity

Actual problem

$$\underset{x\in\mathbb{R}^n}{\text{minimize}}\, f_0(x) + \sum\nolimits_{i=1}^{s} p_i g(\xi_i, x)$$

What if p replaced by p^ν and ξ_i by ξ_i^ν?

Rockafellian

$$f\big((u,v),x\big) = f_0(x) + \sum\nolimits_{i=1}^{s} (p_i + u_i) g(\xi_i + v_i, x) + \iota_{\{0\}^s}(u) + \iota_{\{0\}^{sm}}(v)$$

and approximation

$$f^\nu\big((u,v),x\big) = f_0(x) + \sum\nolimits_{i=1}^{s} (p_i^\nu + u_i) g(\xi_i^\nu + v_i, x) + \tfrac{1}{2}\theta^\nu \|u\|_2^2 + \tfrac{1}{2}\lambda^\nu \|v\|_2^2 + \iota_\Delta(p^\nu + u)$$

strictly exact and asymptotically strictly exact as before

Refinement 3: rate-independent Rockafellian

Return to

$$\underset{x\in\mathbb{R}^n}{\text{minimize}}\, f_0(x) + \sum\nolimits_{i=1}^{s} p_i f_i(x)$$

but with Rockafellian

$$f(u,x) = f_0(x) + \sum\nolimits_{i=1}^{s} (p_i + u_i) f_i(x) + \theta\|u\|_1 + \iota_\Delta(p+u)$$

Refinement 3: rate-independent Rockafellian

Return to

$$\underset{x \in \mathbb{R}^n}{\text{minimize}}\, f_0(x) + \sum_{i=1}^{s} p_i f_i(x)$$

but with Rockafellian

$$f(u, x) = f_0(x) + \sum_{i=1}^{s} (p_i + u_i) f_i(x) + \theta \|u\|_1 + \iota_\Delta(p + u)$$

and approximation

$$f^\nu(u, x) = f_0(x) + \sum_{i=1}^{s} (p_i^\nu + u_i) f_i(x) + \theta \|u\|_1 + \iota_\Delta(p^\nu + u)$$

Refinement 3: rate-independent Rockafellian

Return to

$$\underset{x\in\mathbb{R}^n}{\text{minimize}}\, f_0(x) + \sum\nolimits_{i=1}^{s} p_i f_i(x)$$

but with Rockafellian

$$f(u,x) = f_0(x) + \sum\nolimits_{i=1}^{s}(p_i + u_i)f_i(x) + \theta\|u\|_1 + \iota_\Delta(p+u)$$

and approximation

$$f^\nu(u,x) = f_0(x) + \sum\nolimits_{i=1}^{s}(p_i^\nu + u_i)f_i(x) + \theta\|u\|_1 + \iota_\Delta(p^\nu+u)$$

Strictly exact and asymptotically strictly exact if

$f_i : \mathbb{R}^n \to \overline{\mathbb{R}}$, $i = 0, 1, \ldots, s$ proper lsc; θ sufficiently large

$\|p^\nu - p\|_2 \to 0$

$\exists\bar{x}$ such that $f_i(\bar{x}) < \infty$ and $\inf_x f_i(x)$ finite for $i = 0, 1, \ldots, s$

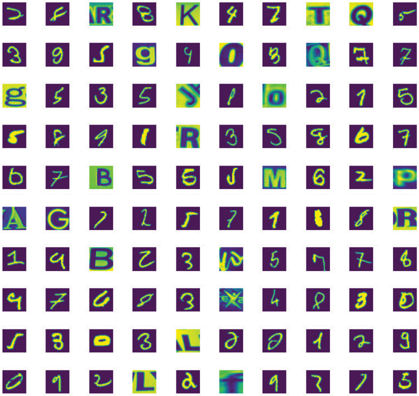
Outliers

60000 MNIST images; 600 Chars74k images (random labels)

"Incorrect" probabilities: $p_i^\nu = 1/60600 = 1.65 \cdot 10^{-5}$

"Correct" prob.: $p_i = 1/60000$ for MNIST; $p_i = 0$ for Chars74k

Solve approximating problem instead of $\underset{x \in \mathbb{R}^n}{\text{minimize}} \sum_{i=1}^{s} p_i^\nu f_i(x)$



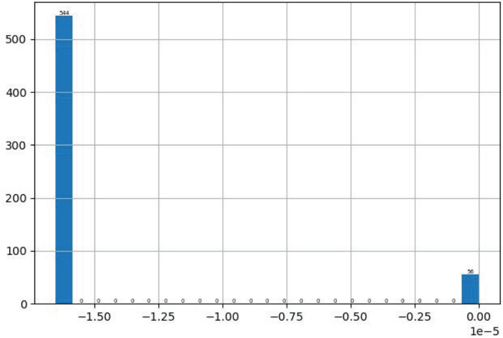
Outlier identification

Solve approximating problem using alternating heuristic

$$\underset{u\in\mathbb{R}^m, x\in\mathbb{R}^n}{\text{minimize}} \sum_{i=1}^{s} (p_i^\nu + u_i) f_i(x) + \theta \|u\|_1 + \iota_\Delta(p^\nu + u)$$

u-minimization: linear programming

x-minimization: stochastic gradient descent



544 of 600 Chars74k images "knocked out" with $u_i = -1.65 \cdot 10^{-5}$

Main references

Chen & Royset, “Rockafellian Relaxation in Optimization under Uncertainty: Asymptotically Exact Formulations,” arXiv:2204.04762

Royset, 2021, “Good and Bad Optimization: Insight from Rockfellians,” in INFORMS Tutorials; J. Carlsson (Ed.), INFORMS

Royset & Wets, 2021, An Optimization Primer, Springer