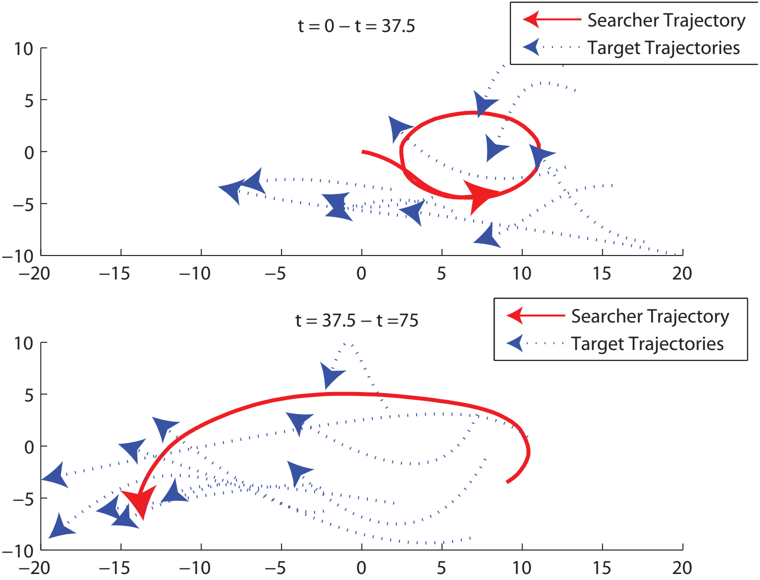
Consistent Approximations in Optimization

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

Supported in part by AFOSR, ONR, and DARPA

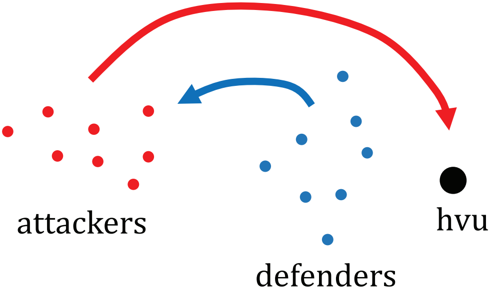
Linz, Austria, November 2019



Phelps, Royset & Gong, "Optimal Control of Uncertain Systems using Sample Average Approximations," SIAM J. Control and Optimization, 2016

Stone, Royset & Washburn, Optimal Search for Moving Targets, Springer, 2016

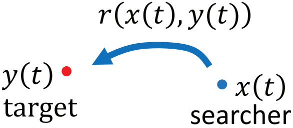
Maximize probability of HVU survival



Walton, Lambrianides, Kaminer, Royset & Gong, "Optimal Motion Planning in Rapid-Fire Combat Situations with Attacker Uncertainty," Naval Research Logistics, 2018

Seven defenders vs 100 attackers

Modeling probability of detection



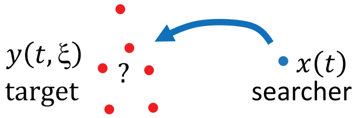
$r(x(t), y(t))\Delta t$: probability of detection during $[t, t + \Delta t)$

$q(t)$: probability of no detection during $[0, t]$

$$q(t + \Delta t) = q(t)(1 - r(x(t), y(t))\Delta t)$$

$$\dot{q}(t) = -q(t) r(x(t), y(t)), \qquad q(0) = 1$$

Target uncertainty



$\{y(t,\boldsymbol{\xi}), t \in [0,1]\}$ uncertain track of target; $\boldsymbol{\xi}$ random vector

$q(t,\xi)$: prob. of no detection during $[0,t]$ given ξ

$\dot{q}(t,\xi) = -q(t,\xi) r(x(t), y(t,\xi), \xi),\ q(0,\xi) = 1$

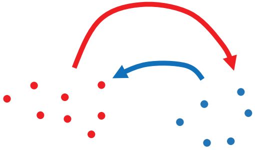
$\mathbb{E}\big[q(1,\boldsymbol{\xi})\big]$ probability of no detection during $[0,1]$

Combine $q(t,\xi)$ with searcher state $x(t)$ to get state $x(t,\xi)$

$$\underset{u \in U}{\text{minimize}}\, \mathbb{E}\big[\varphi(x^u(1,\boldsymbol{\xi}),\boldsymbol{\xi})\big]$$

with $x^u(\cdot,\xi)$ solving $\dot{x}(t,\xi) = f\big(x(t,\xi), u(t), \xi\big);\ x(0,\xi) = x_0(\xi)$ a.s.

Attacker-Defender



$$
\begin{aligned}
\dot{p}_0(t,\xi) &= -r(x(t),y(t,\xi),\xi)p_0(t,\xi)Q(t) \\
\dot{p}_1(t,\xi) &= -r(x(t),y(t,\xi),\xi)\big(p_1(t,\xi)-p_0(t,\xi)\big)Q(t) \\
&\vdots \\
\dot{p}_{N-1}(t,\xi) &= -r(x(t),y(t,\xi),\xi)\big(p_{N-1}(t,\xi)-p_{N-2}(t,\xi)\big)Q(t) \\
\dot{q}_0(t,\xi) &= -s(x(t),y(t,\xi),\xi)q_0(t,\xi)P(t) \\
\dot{q}_1(t,\xi) &= -s(x(t),y(t,\xi),\xi)\big(q_1(t,\xi)-q_0(t,\xi)\big)P(t) \\
&\vdots \\
\dot{q}_{N-1}(t,\xi) &= -s(x(t),y(t,\xi),\xi)\big(q_{N-1}(t,\xi)-q_{N-2}(t,\xi)\big)P(t)
\end{aligned}
$$

$$
P(t) = \sum_{n=0}^{N-1} p_n(t) \qquad Q(t) = \sum_{n=0}^{N-1} q_n(t)
$$

Setting for presentation

(X, d) metric space
$f^\nu, f : X \to [-\infty, \infty]$, usually lower semicontinuous (lsc)

$$\begin{aligned} \text{Actual problem:} \quad & \operatorname*{minimize}_{x \in X} f(x) \\ \text{Approximating problem:} \quad & \operatorname*{minimize}_{x \in X} f^\nu(x) \end{aligned}$$

Constraints often handled abstractly:
Setting objective function to ∞ if x infeasible (wlog)

Setting for presentation

(X, d) metric space
$f^\nu, f : X \to [-\infty, \infty]$, usually lower semicontinuous (lsc)

$$\text{Actual problem: } \underset{x \in X}{\text{minimize}}\, f(x)$$
$$\text{Approximating problem: } \underset{x \in X}{\text{minimize}}\, f^\nu(x)$$

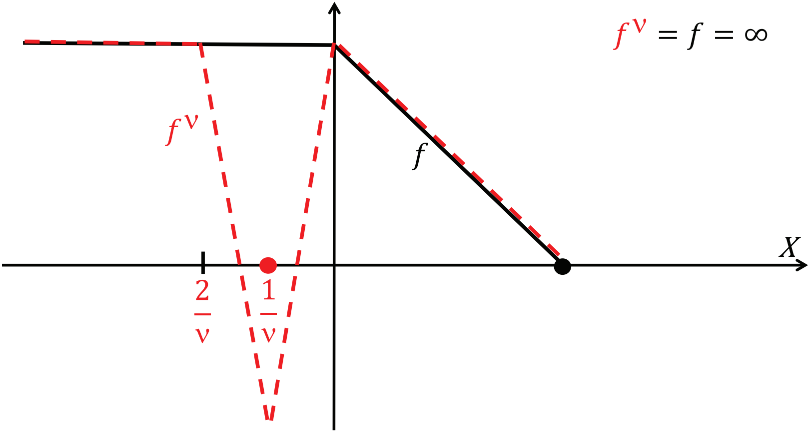
Constraints often handled abstractly:
Setting objective function to ∞ if x infeasible (wlog)

What constitutes a consistent approximation?

Level 0: convergence of minimizers, minima

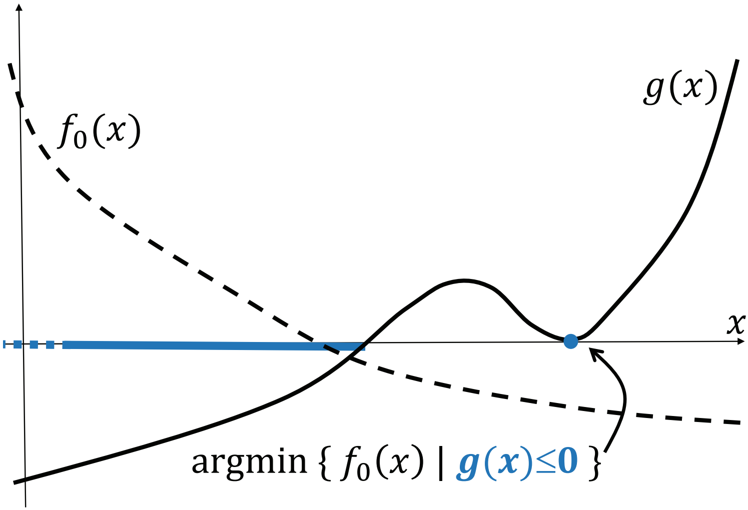
Level 1: convergence of first-order stationary points

Would pointwise convergence suffice?

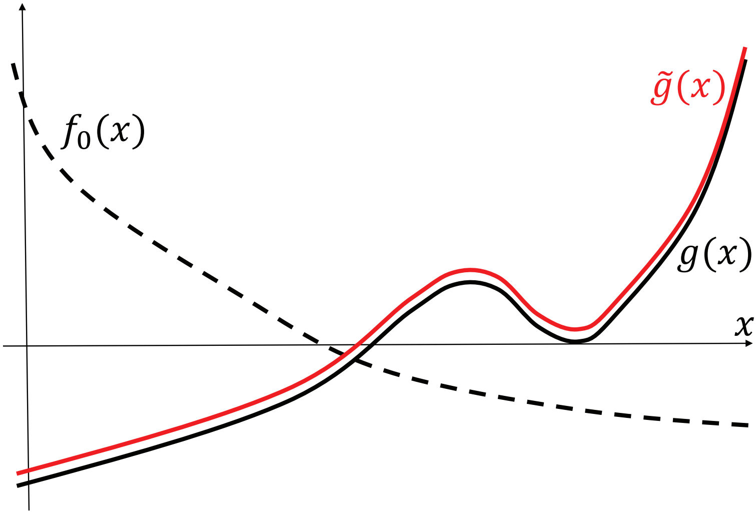


Pointwise convergence not sufficient for convergence of minimizers

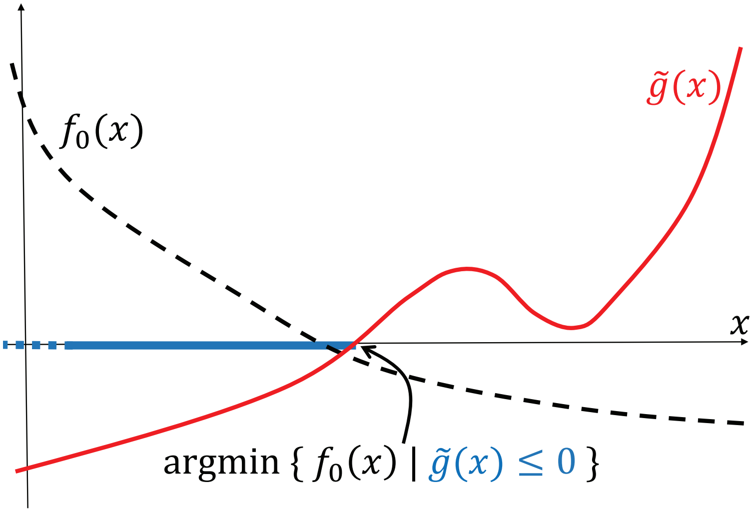
What about uniform convergence?



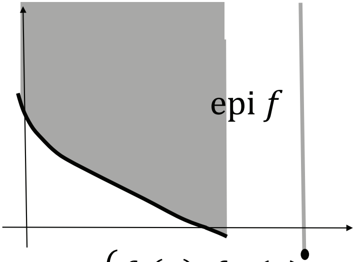
What about uniform convergence?



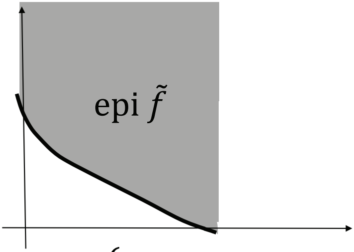
Uniform “approximation,” but large error in argmin



Passing to epigraphs of the effective functions

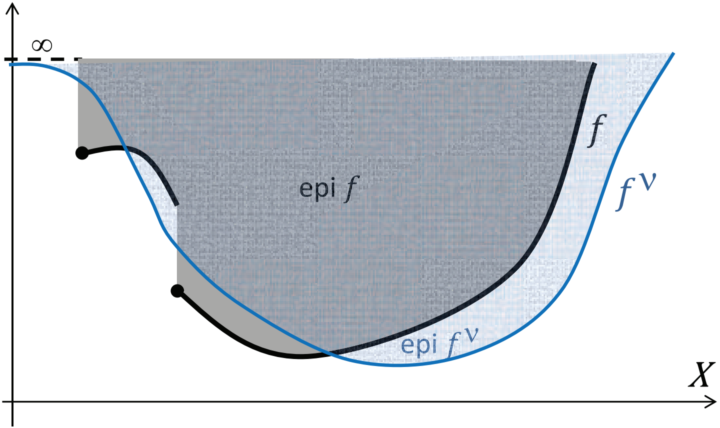


$$f(x) = \begin{cases} f_0(x) \text{ if } g(x) \leq 0 \\ \infty \quad \text{otherwise} \end{cases}$$



$$\tilde{f}(x) = \begin{cases} f_0(x) \text{ if } \tilde{g}(x) \leq 0 \\ \infty \quad \text{otherwise} \end{cases}$$

Epi-convergence

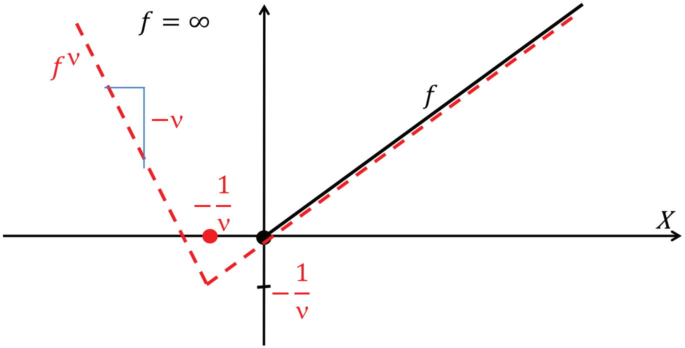


f^ν epi-converges to f $\Longleftrightarrow$ epi f^ν set-converges to epi f

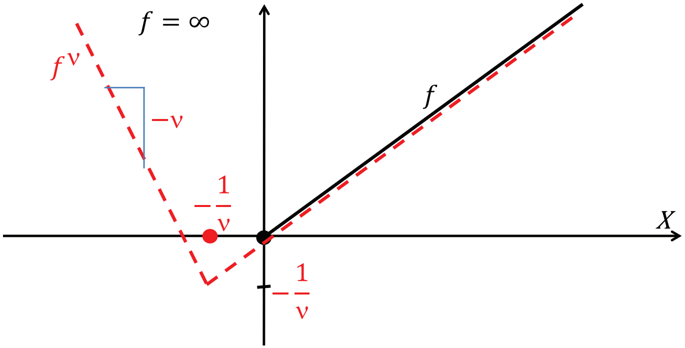
Main consequence:

f^ν epi-converges to f and $x^\nu \in \text{argmin}\, f^\nu \to \bar{x} \Longrightarrow \bar{x} \in \text{argmin}\, f$

Approximation of constraints



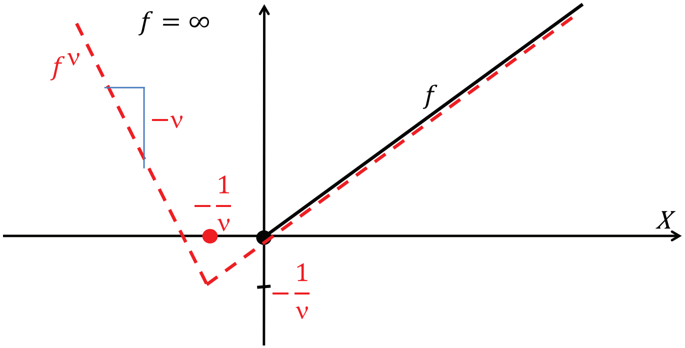
Approximation of constraints



If C^ν set-converges to C and f_0 continuous, then

$$f^\nu(x) = \begin{cases} f_0(x) & \text{if } x \in C^\nu \\ \infty & \text{otherwise} \end{cases} \quad \text{epi-conv to } f(x) = \begin{cases} f_0(x) & \text{if } x \in C \\ \infty & \text{otherwise} \end{cases}$$

Approximation of constraints



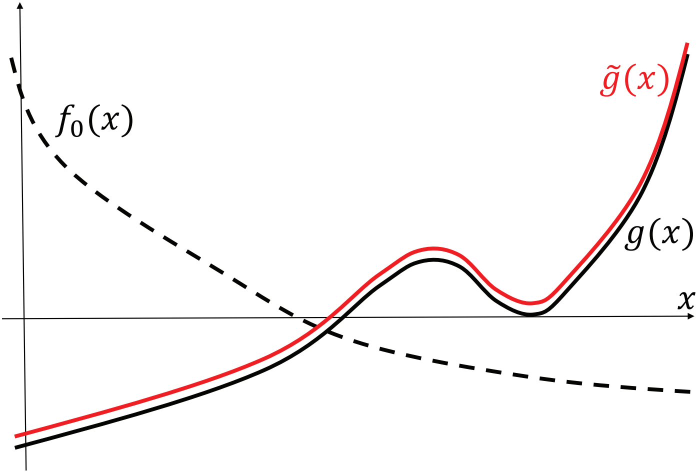
If C^ν set-converges to C and f_0 continuous, then

$$f^\nu(x) = \begin{cases} f_0(x) & \text{if } x \in C^\nu \\ \infty & \text{otherwise} \end{cases} \quad \text{epi-conv to } f(x) = \begin{cases} f_0(x) & \text{if } x \in C \\ \infty & \text{otherwise} \end{cases}$$

Example: $C^1, C^2, \ldots$ dense in $C = X \implies C^\nu$ set-converges to C

Recall failure under uniform convergence

What can be done in this case?



Constraint softening

$$\underset{x \in X}{\text{minimize}}\, f_0(x) \text{ subject to } g_i(x) \leq 0,\ \ i = 1, \ldots, q$$

$$\sup_{x \in X} |f_0^\nu(x) - f_0(x)| \leq \alpha^\nu \quad \text{and} \quad \sup_{x \in X} \max_{i=1,\ldots,q} |g_i^\nu(x) - g_i(x)| \leq \alpha^\nu$$

Constraint softening

$$\underset{x \in X}{\text{minimize}}\, f_0(x) \text{ subject to } g_i(x) \leq 0,\ \ i = 1, \ldots, q$$

$$\sup_{x \in X} |f_0^\nu(x) - f_0(x)| \leq \alpha^\nu \quad \text{and} \quad \sup_{x \in X} \max_{i=1,\ldots,q} |g_i^\nu(x) - g_i(x)| \leq \alpha^\nu$$

$$\underset{x \in X, y \in \mathbb{R}^q}{\text{minimize}}\ f_0^\nu(x) + \theta^\nu \sum_{i=1}^{q} y_i \ \text{ subject to } \ g_i^\nu(x) \leq y_i,\ \ 0 \leq y_i,\ \ i = 1, \ldots, q$$

Constraint softening

$$\underset{x \in X}{\text{minimize}}\, f_0(x) \text{ subject to } g_i(x) \le 0,\;\; i = 1, \ldots, q$$

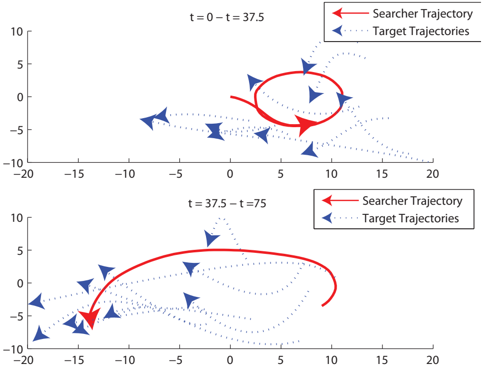
$$\sup_{x \in X} |f_0^\nu(x) - f_0(x)| \le \alpha^\nu \;\;\text{ and }\;\; \sup_{x \in X} \max_{i=1,\ldots,q} |g_i^\nu(x) - g_i(x)| \le \alpha^\nu$$

$$\underset{x \in X, y \in \mathbb{R}^q}{\text{minimize}}\;\; f_0^\nu(x) + \theta^\nu \sum_{i=1}^{q} y_i \;\text{ subject to }\; g_i^\nu(x) \le y_i,\;\; 0 \le y_i,\;\; i = 1, \ldots, q$$

f_0 continuous
g_i lsc, $i = 1, \ldots, q$
$\theta^\nu \to \infty$, $\alpha^\nu \to 0$, $\theta^\nu \alpha^\nu \to 0$

Then, approximation epi-converges to actual

Epi-convergence under sampling and forward Euler



$$\underset{u \in U}{\text{minimize}}\, \mathbb{E}\left[\varphi(x^u(1,\boldsymbol{\xi}),\boldsymbol{\xi})\right]$$

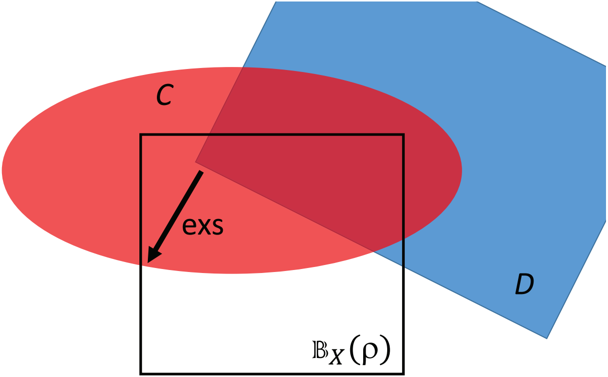
with $x^u(\cdot,\xi)$ solving $\dot{x}(t,\xi) = f\left(x(t,\xi), u(t), \xi\right);\ x(0,\xi) = x_0(\xi)$ a.s.

Sampling and Forward Euler result in epi-convergence

Phelps, Royset & Gong, "Optimal Control of Uncertain Systems using Sample Average Approximations," SIAM J. Control and Optimization, 2016

Truncated Hausdorff distance between sets

For $C, D \subset X$ (metric space)



$$d\hat{l}_{\rho}(C, D) = \max\left\{ \operatorname{exs}\left(C \cap \mathbb{B}_X(\rho); D\right), \ \operatorname{exs}\left(D \cap \mathbb{B}_X(\rho); C\right) \right\}$$

Consequence for minima and near-minimizers

For $f, g : X \to [-\infty, \infty]$,

$$|\inf f - \inf g| \leq \hat{dl}_\rho(\text{epi} f, \text{epi} g)$$

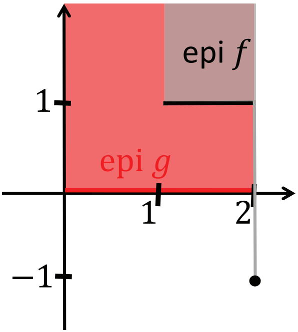
$$\text{exs}\left(\varepsilon\text{-}\operatorname{argmin} g \cap \mathbb{B}_X(\rho);\ \delta\text{-}\operatorname{argmin} f\right) \leq \hat{dl}_\rho(\text{epi} f, \text{epi} g)$$
$$\text{if } \delta > \varepsilon + 2\hat{dl}_\rho(\text{epi} f, \text{epi} g)$$

(product metric is used on $X \times \mathbb{R}$ and ρ large enough)

Replace $>$ by $\geq$ when f and g lsc and X has compact balls

Bounds are sharp

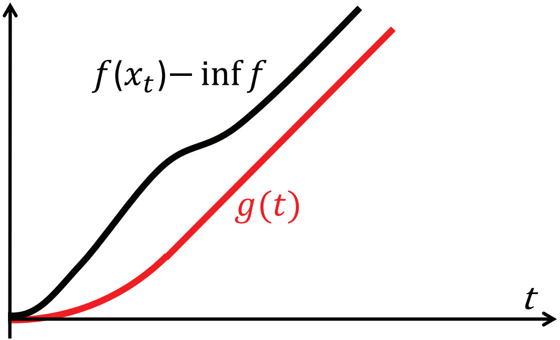
$$\text{exs}\left(\varepsilon\text{-}\operatorname{argmin} g \cap \mathbb{B}_X(\rho);\ \delta\text{-}\operatorname{argmin} f\right) \le \hat{dl}_\rho(\text{epi} f, \text{epi} g)$$
$$\text{if } \delta \ge \varepsilon + 2\hat{dl}_\rho(\text{epi} f, \text{epi} g)$$



What about minimizers?

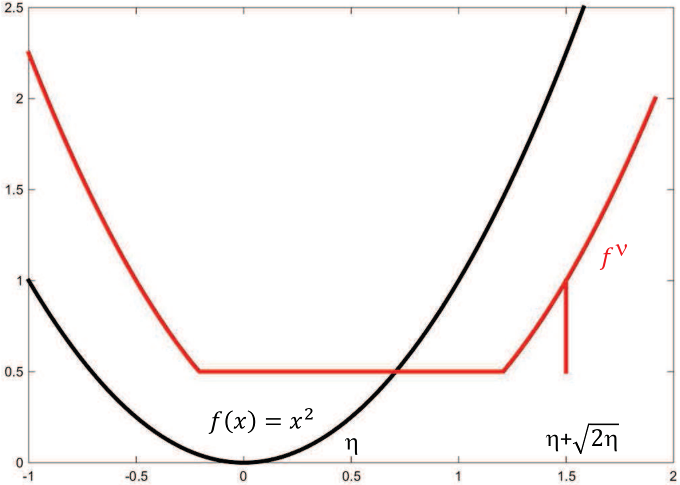
When $f(x) - \inf f \geq g(\mathrm{dist}(x, \mathrm{argmin}\, f))\ \forall x \in X$ for incr g

$$\mathrm{exs}\left(\mathrm{argmin}\, f^\nu \cap \mathbb{B}_X(\rho), \mathrm{argmin}\, f\right) \leq d\hat{l}_\rho(\mathrm{epi}\, f, \mathrm{epi}\, f^\nu) + g^{-1}\left(2d\hat{l}_\rho(\mathrm{epi}\, f, \mathrm{epi}\, f^\nu)\right)$$



Sharpness of bound on minimizers

$d\hat{l}_\rho(\text{epi}\, f, \text{epi}\, f^\nu) = \eta = 1/2$; f has growth $g(t) = t^2$



$$\text{exs}\left(\text{argmin}\, f^\nu \cap \mathbb{B}_X(\rho), \text{argmin}\, f\right) \leq \eta + g^{-1}(2\eta)$$

Computing distances for compositions

For κ-Lipschitz $f : Y \to \mathbb{R}$ and $F, G : X \to Y$,

$$d\hat{l}_\rho\big(\operatorname{epi}(f \circ F),\ \operatorname{epi}(f \circ G)\big) \le \max\{1, \kappa\} d\hat{l}_{\bar\rho}(\operatorname{gph} F, \operatorname{gph} G)$$

provided that $\bar\rho$ large enough

Distances for sums

$f_i, g_i : X \to [-\infty, \infty]$, $i = 1, 2$,

f_1, g_1 are Lipschitz continuous with common modulus κ

$$d\hat{l}_\rho\big(\operatorname{epi}(f_1 + f_2), \operatorname{epi}(g_1 + g_2)\big) \le \sup_{A_\rho} |f_1 - g_1| + \big(1 + \kappa\big) d\hat{l}_{\bar{\rho}}(\operatorname{epi} f_2, \operatorname{epi} g_2)$$

provided that $\operatorname{epi}(f_1 + f_2)$ and $\operatorname{epi}(g_1 + g_2)$ are nonempty,

$A_\rho = \{f_1 + f_2 \le \rho\} \cup \{g_1 + g_2 \le \rho\} \cap \mathbb{B}_X(\rho)$,

$\bar{\rho} \ge \rho + \max\{0, -\inf_{\mathbb{B}_X(\rho)} f_1, -\inf_{\mathbb{B}_X(\rho)} g_1\}$

Convergence of stationary points

First-order conditions for $\text{minimize}_{x \in X} f(x)$:

$$\begin{aligned} \text{Oresme Rule:} &\quad df(x; w) \geq 0 \ \forall w \in X \\ \text{Fermat Rule:} &\quad 0 \in \partial f(x) \end{aligned}$$

Convergence of stationary points

First-order conditions for $\text{minimize}_{x \in X} f(x)$:

$$\begin{aligned} \text{Oresme Rule:} \quad & df(x;w) \geq 0 \ \forall w \in X \\ \text{Fermat Rule:} \quad & 0 \in \partial f(x) \end{aligned}$$

More generally:

For set-valued mapping $S : X \rightrightarrows Y$ and point $y^\star \in Y$
Generalized equation $y^\star \in S(x)$ has solution set $S^{-1}(y^\star)$

Convergence of stationary points

First-order conditions for $\text{minimize}_{x \in X} f(x)$:

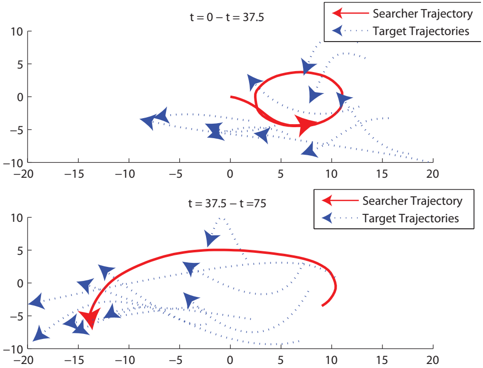
$$\begin{aligned} \text{Oresme Rule:} \quad & df(x; w) \geq 0 \ \forall w \in X \\ \text{Fermat Rule:} \quad & 0 \in \partial f(x) \end{aligned}$$

More generally:

For set-valued mapping $S : X \rightrightarrows Y$ and point $y^\star \in Y$
Generalized equation $y^\star \in S(x)$ has solution set $S^{-1}(y^\star)$

If $\operatorname{gph} S^\nu$ set-conv to $\operatorname{gph} S$, $y^\nu \to y^\star$, and $x^\nu \in (S^\nu)^{-1}(y^\nu) \to x^\star$, then $x^\star \in S^{-1}(y^\star)$

Convergence for Oresme Rule



$$\underset{u \in U}{\text{minimize}}\, \mathbb{E}\left[\varphi(x^u(1,\boldsymbol{\xi}),\boldsymbol{\xi})\right]$$

with $x^u(\cdot,\xi)$ solving $\dot{x}(t,\xi) = f\left(x(t,\xi),u(t),\xi\right);\; x(0,\xi) = x_0(\xi)$ a.s.

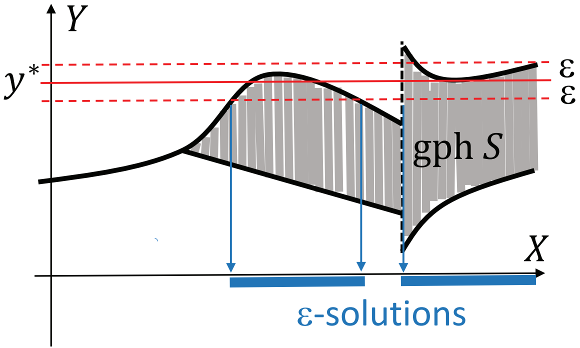
Sampling: Convergence of Oresme stationary points

Phelps, Royset & Gong, "Optimal Control of Uncertain Systems using Sample Average Approximations," SIAM J. Control and Optimization, 2016

Solutions of generalized equations

For $\varepsilon \geq 0$, the **set of ε-solutions** is defined as

$$S^{-1}\big(\mathbb{B}_Y(y^\star, \varepsilon)\big) = \bigcup_{y \in \mathbb{B}_Y(y^\star, \varepsilon)} S^{-1}(y)$$



Example

Optimality conditions for minimizing f over C

$$0 \in \partial f(x) + N_C(x)$$

With $S = \partial f + N_C$ and $y^\star = 0$, the set of ε-solutions becomes

$$S^{-1}\big(\mathbb{B}_{\mathbb{R}^n}(\varepsilon)\big) = \big\{x \in \mathbb{R}^n \mid 0 \in \partial f(x) + N_C(x) + \mathbb{B}_{\mathbb{R}^n}(\varepsilon)\big\}$$

Solution estimates for generalized equations

For metric spaces X and Y, suppose that $S, T : X \rightrightarrows Y$ have nonempty graphs, $0 \leq \varepsilon \leq \rho < \infty$, and $y^\star \in \mathbb{B}_Y(\rho - \varepsilon)$

Then,

$$\text{exs}\left(S^{-1}\big(\mathbb{B}_Y(y^\star, \varepsilon)\big)\cap\mathbb{B}_X(\rho);\ T^{-1}\big(\mathbb{B}_Y(y^\star, \delta)\big)\right) \leq \hat{dl}_\rho(\text{gph}\, S, \text{gph}\, T)$$

provided that $\delta > \varepsilon + \hat{dl}_\rho(\text{gph}\, S, \text{gph}\, T)$

Solution estimates for generalized equations

For metric spaces X and Y, suppose that $S, T : X \rightrightarrows Y$ have nonempty graphs, $0 \leq \varepsilon \leq \rho < \infty$, and $y^\star \in \mathbb{B}_Y(\rho - \varepsilon)$

Then,

$$\text{exs}\left(S^{-1}\big(\mathbb{B}_Y(y^\star, \varepsilon)\big) \cap \mathbb{B}_X(\rho);\ T^{-1}\big(\mathbb{B}_Y(y^\star, \delta)\big)\right) \leq \hat{dl}_\rho(\text{gph}\, S, \text{gph}\, T)$$

provided that $\delta > \varepsilon + \hat{dl}_\rho(\text{gph}\, S, \text{gph}\, T)$

If X and Y have compact balls and gph T is closed, then the result also holds for $\delta = \varepsilon + \hat{dl}_\rho(\text{gph}\, S, \text{gph}\, T)$

Example: KKT solutions

minimize $f_0(x)$ subject to $f_i(x) \leq 0$ for $i = 1, \ldots, m$ (smooth)

$(x, y) \in \mathbb{R}^{n+m}$ KKT solution if and only if $0 \in S(x, y)$

Example: KKT solutions

minimize $f_0(x)$ subject to $f_i(x) \le 0$ for $i = 1, \ldots, m$ (smooth)

$(x, y) \in \mathbb{R}^{n+m}$ KKT solution if and only if $0 \in S(x, y)$

where $S : \mathbb{R}^{n+m} \rightrightarrows \mathbb{R}^{3m+n}$ has

$$S(x,y) = \begin{pmatrix} [f_1(x), \infty) \\ \vdots \\ [f_m(x), \infty) \\ (-\infty, y_1] \\ \vdots \\ (-\infty, y_m] \\ \{y_1 f_1(x)\} \\ \vdots \\ \{y_m f_m(x)\} \\ \{\nabla f_0(x) + \sum_{i=1}^{m} y_i \nabla f_i(x)\} \end{pmatrix}$$

Estimates of KKT solutions

Let $g_0, \ldots, g_m$ define $T : \mathbb{R}^{n+m} \rightrightarrows \mathbb{R}^{3m+n}$ similarly to S

Then,

$$d\hat{l}_\rho(\text{gph}\, S, \text{gph}\, T) \leq \max\left\{\delta, \rho\delta, (1 + m\rho)\eta\right\},$$

where

$$\delta = \max_{i=0,\ldots,m} \sup_{\|x\|_\infty \leq \rho} |f_i(x) - g_i(x)|$$

$$\eta = \max_{i=0,\ldots,m} \sup_{\|x\|_\infty \leq \rho} \|\nabla f_i(x) - \nabla g_i(x)\|_\infty$$

KKT system is stable while minimizers may not be

Optimality for composite functions

$\varphi : \mathbb{R}^m \to \overline{\mathbb{R}}$ proper lsc function

$F : \mathbb{R}^n \to \mathbb{R}^m$ smooth

$\text{minimize}_{x \in \mathbb{R}^n} \varphi\big(F(x)\big)$
with optimality condition $0 \in \nabla F(x)^\top \partial\varphi(F(x))$

Equivalently,

$$0 \in S(x, y, z) = \begin{pmatrix} \{F(x) - z\} \\ \partial\varphi(z) - \{y\} \\ \{\nabla F(x)^\top y\} \end{pmatrix}$$

Approximations

$\psi : \mathbb{R}^m \to \overline{\mathbb{R}}$ proper lsc function

$G : \mathbb{R}^n \to \mathbb{R}^m$ smooth

$\text{minimize}_{x \in \mathbb{R}^n} \psi\big(G(x)\big)$
with optimality condition $0 \in \nabla G(x)^\top \partial\psi(G(x))$

Equivalently,

$$0 \in T(x, y, z) = \begin{pmatrix} \{G(x) - z\} \\ \partial\psi(z) - \{y\} \\ \{\nabla G(x)^\top y\} \end{pmatrix}$$

Approximation error

$$\hat{dl}_\rho(\text{gph}\, S, \text{gph}\, T) \leq \sup_{\|x\|\leq\rho} \max\Big\{\rho\big\|\nabla G(x)^\top - \nabla F(x)^\top\big\|, \\ \big\|G(x) - F(x)\big\| + \hat{dl}_{2\rho}\big(\text{gph}\,\partial\varphi, \text{gph}\,\partial\psi\big)\Big\}$$

References

Stone, Royset & Washburn, Optimal Search for Moving Targets, Springer, 2016

Phelps, Royset & Gong, “Optimal Control of Uncertain Systems using Sample Average Approximations,” SIAM J. Control and Optimization, 2016

Walton, Lambrianides, Kaminer, Royset & Gong, “Optimal Motion Planning in Rapid-Fire Combat Situations with Attacker Uncertainty,” Naval Research Logistics, 2018

Royset, “Approximations and Solution Estimates in Optimization,” Mathematical Programming, 170(2):479-506, 2018

Royset, “Approximations of Semicontinuous Functions with Applications to Stochastic Optimization and Statistical Estimation,” Mathematical Programming, OnlineFirst, 2019

Royset, “Stability and Error Analysis for Optimization and Generalized Equations,” arXiv:1903.08754