AN OVERVIEW OF VARIATIONAL ANALYSIS

4. VARIATIONAL APPROXIMATION

Terry Rockafellar
University of Washington, Seattle

Forum on Developments and Origins of Operations Research

Organized by the *Mathematical Programming Branch of the Operations Research Society of China* and the *Southern University of Science and Technology* (SUSTech)

Shenzhen, China, 6–10 December 2021 (virtual)

Min and Argmin as "Operations"

Basic problem of optimization: minimize proper lsc f on $\mathbb{R}^n$

implicit feasible set: $\operatorname{dom} f = \{x \mid f(x) < \infty\}$

$\inf f$ = greatest lower bound to $\{f(x) \mid x \in \mathbb{R}^n\}$

$\operatorname{argmin} f$ = set of x, if any, such that $f(x) = \inf f < \infty$

$\min f$ = same as $\inf f$ but indicating that $\operatorname{argmin} f \neq \emptyset$

Traditional existence criterion: associated with $f = f_0 + \delta_C$

$\operatorname{argmin} f \neq \emptyset$ when f_0 is continuous and C is compact

too limited, doesn't cover unbounded sets, approximations

General existence criterion — utilizing level-boundedness of f

$\operatorname{argmin} f \neq \emptyset$ compact if $\exists\, c > \inf f$ with $\{x \mid f(x) \leq c\}$ bounded

Fundamental issue: behavior of $f \mapsto \operatorname{argmin} f$ and $f \mapsto \min f$

as "calculus" operations performed on functions f

Shortcomings of Classical Function Convergence on $\boldsymbol{R}^n$

Convergence to f of a sequence of functions f^ν, $\nu = 1, 2, \ldots$?

what conditions in approximating f by f^ν ensure that $\operatorname{argmin} f$, $\min f$, are approximated by $\operatorname{argmin} f^\nu$, $\min f^\nu$?

Pointwise convergence: traditional answer #1

$f^\nu(x) \to f(x)$ at every point $x \in \boldsymbol{R}^n$

Locally uniform convergence: traditional answer #2

$\max_{x \in B} |f^\nu(x) - f(x)| \to 0$ for every bounded set $B \subset \boldsymbol{R}^n$

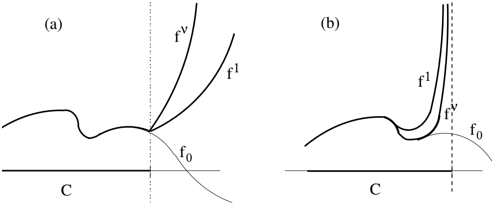
(these are equivalent under uniform local Lipschitz continuity)

But
- pointwise convergence requires $\operatorname{dom} f^\nu = \operatorname{dom} f$, $\forall \nu$!
- locally uniform convergence can't handle ∞ values at all!

what turns out to matter is whether $\operatorname{epi} f^\nu$ "converges" to $\operatorname{epi} f$

Illustration of Convergence Needs

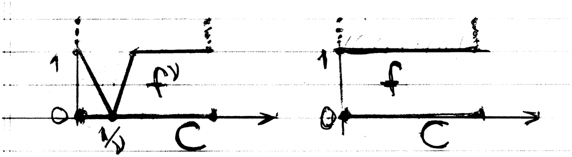
Problem approximations: in minimizing f_0 over C



$f = f_0 + \delta_C$ replaced by $f = f_0 + g$ for some $g \approx \delta_C$

penalty functions g, barrier functions g

Bad behavior to avoid: $\min f^\nu = 0$ always, but $\min f = 1$



continuous functions converging pointwise on a compact set

Set Convergence

Outer and inner limits: of closed sets C and C^ν for $\nu = 1, 2, \ldots$

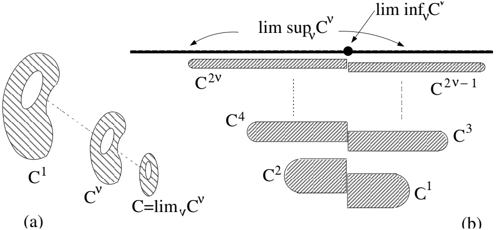
$C = \limsup_\nu C^\nu$ when C consists of all x such that

$\exists$ subsequence C^{ν_κ} and $x^{\nu_\kappa} \in C^{\nu_\kappa}$ with $x^{\nu_\kappa} \to x$

$C = \liminf_\nu C^\nu$ when C consists of all x such that

for high enough ν, $\exists\, x^\nu \in C^\nu$ with $x^\nu \to x$

$C = \lim_\nu C^\nu$ when $C = \limsup_\nu C^\nu = \liminf_\nu C^\nu$



Distance characterization, with $d_C(x) := \min\{|x' - x| \,|\, x' \in C\}$

$$C = \lim_\nu C^\nu \iff d_{C^\nu}(x) \to d_C(x) \text{ for all } x$$

Epi-convergence of Functions on R^n

let f and f^ν for $\nu = 1, 2, \ldots$ be lsc proper functions on R^n

The meaning of epi-convergence of f^ν to f, notation $f^\nu \xrightarrow{epi} f$

the sequence of sets $E^\nu = \mathrm{epi}\, f^\nu$ converges to $E = \mathrm{epi}\, f$

Characterization in limits of function values

f^ν **epi-converges** to f $\iff$ at **all** points x,

- $\liminf_\nu f^\nu(x^\nu) \geq f(x)$ for **every** sequence $x^\nu \to x$,
- $\limsup_\nu f^\nu(x^\nu) \leq f(x)$ for **some** sequence $x^\nu \to x$

Application to min and argmin — key theorem

Assume $\exists c > \inf f > \infty$ with $\{x \mid f^\nu(x) \leq c\} \subset$ bounded B, $\forall \nu$.

Then $\mathrm{argmin}\, f$ is nonempty and compact, and

$$\min f^\nu \to \min f, \qquad \limsup_\nu [\mathrm{argmin}\, f^\nu] \subset \mathrm{argmin}\, f$$

Some Examples of Epi-convergence

Approximation of $f = f_0 + \delta_C$ for f_0 continuous, C closed

Suppose $f^\nu = f_0^\nu + \delta_{C^\nu}$ with f_0^ν continuous. If $f_0^\nu \to f_0$ locally uniformly and $C^\nu \to C$, then f^ν epi-converges to f

note: $C^\nu \to C \iff \delta_{C^\nu} \overset{epi}{\to} \delta_C$

Application to duality: recall conjugacy of convex functions

$$f^*(v) = \sup_x \{v{\cdot}x - f(x)\}, \quad f(x) = f^{**}(x) = \sup_v \{v{\cdot}x - f^*(v)\}$$

Wijsman's Theorem: on epi-continuity of the transform $f \mapsto f^*$

f^ν epi-converges to $f \iff f^{\nu *}$ epi-converges to f^*

Polarity of convex cones as a special case:

$$K^\nu \to K \iff K^{\nu *} \to K^* \qquad \text{because } \delta_{K^*} = \delta_K^*$$

Graphical Convergence of Mappings

Classical setting: single-valued mappings F, F^ν, from $\mathbb{R}^n$ to $\mathbb{R}^m$

pointwise convergence, locally uniform convergence

not good concepts for treating **set-valued** mappings M, M^ν

$M : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$, graphically identified with a subset of $\mathbb{R}^n \times \mathbb{R}^m$:

$$\operatorname{gph} M = \{(x,u) \mid \in M(x)\}$$

$\operatorname{dom} M = \{x \mid M(x) \neq \emptyset\}$, inverse: $x \in M^{-1}(u) \Leftrightarrow u \in M(x)$

pointwise convergence? $M^\nu(x) \to M(x)$ as sets?

locally uniform convergence? $|M^\nu(x) - M(x)|$ means what?

trouble for both in particular when $\operatorname{dom} M^\nu \neq \operatorname{dom} M$

Definition of $M^\nu \xrightarrow{gph} M$

M^ν converges **graphically** to M if $\operatorname{gph} M^\nu$ converges to $\operatorname{gph} M$

inner and outer limits can be useful as well

Specialization to Subgradient Mappings

Monotonicity of a set-valued mapping: $M : \mathbb{R}^n \to \mathbb{R}^n$

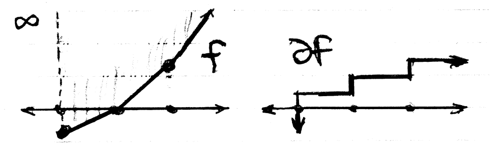
$$(x_1 - x_0)\cdot(v_1 - v_0) \geq 0 \text{ for all } (x_0, v_0), (x_1, v_1) \in \operatorname{gph} M$$

maximal: when $\nexists$ monotone M' with $\operatorname{gph} M' \supset \operatorname{gph} M$, $\neq$

Poliquin's Theorem: on monotonicity of subgradient mappings

For f proper lsc on $\mathbb{R}^n$ and the mapping $\partial f : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$,

∂f monotone $\iff$ ∂f max monotone $\iff$ f convex



Attouch's Theorem: on convergence of subgradient mappings

For proper lsc **convex** functions f and f^ν on $\mathbb{R}^n$,

$f^\nu \xrightarrow{epi} f \iff \partial f^\nu \xrightarrow{gph} \partial f$ and $\exists\, x^\nu \to x$ with $f^\nu(x^\nu) \to f(x) < \infty$

Graphical Differentiation of Mappings

Aim: develop derivatives of set-valued mappings via variational geometry of graphs, then apply that to subgradient mappings

consider $M : \mathbb{R}^n \to \mathbb{R}^d$ and $(\bar{x}, \bar{u}) \in \operatorname{gph} M$, closed in $\mathbb{R}^n \times \mathbb{R}^d$

Graphical derivative mapping: $DM(\bar{x} \,|\, \bar{u}) : \mathbb{R}^n \rightrightarrows \mathbb{R}^d$

graph of $DM(\bar{x} \,|\, \bar{u})$:= tangent cone to $\operatorname{gph} M$ at $(\bar{x}, \bar{u})$

Expression through difference quotient mappings

$$\Delta_\tau M(\bar{x} \,|\, \bar{u}) : \mathbb{R}^n \rightrightarrows \mathbb{R}^d, \qquad \Delta_\tau M(\bar{x} \,|\, \bar{u})(w) = \tfrac{1}{\tau}\big[M(\bar{x} + \tau w) - \bar{u}\big]$$

$$\operatorname{gph} DM(\bar{x} \,|\, \bar{u}) = \limsup_{\tau \searrow 0} \Big[\operatorname{gph} \Delta_\tau M(\bar{x}, \bar{u})\Big]$$

Proto-differentiability: the case of "**lim**," not just "**limsup**"

Second-Order Variational Analysis?

Recall the classical framework: for a function $f : R^n \to R$

- $\nabla f(x) = v \in R^n$ with $f(x+w) = f(x) = v{\cdot}w + o(|w|)$
- $\nabla^2 f(x) = H \in R^{n\times n}$ with $\nabla f(x+w) = \nabla f(x) + H{\cdot}w + o(|w|)$
- then also $f(x+w) = f(x) + \nabla f(x){\cdot}w + \frac{1}{2}w{\cdot}\nabla^2 f(x)w + o(|w|^2)$

Connection with directional derivatives:

$$\nabla f(x){\cdot}w = \lim_{\tau \searrow 0} [f(x+\tau w) - f(x)]/\tau$$

$$\tfrac{1}{2}w{\cdot}\nabla^2 f(x)\, w = \lim_{\tau \searrow 0} [f(x+\tau w) - f(x) - \tau w{\cdot}\nabla f(x)]/\tau^2$$

Challenges of generalization to a function $f : R^n \to \overline{I\!R}$

- single-valued mapping ∇f replaced by set-valued mapping ∂f
- articulation of directional derivatives in ∞-valued context

Generalized Second Derivatives Using Epi-convergence

let f be proper lsc on R^n and let $(\bar{x}, \bar{v}) \in \operatorname{gph} \partial f$

Second-order difference quotients:

$$\tfrac{1}{2}\Delta_\tau^2 f(\bar{x} \,|\, \bar{v})(w) = [f(\bar{x} + \tau w) - f(\bar{x}) - \bar{v}{\cdot}\tau w]/\tau^2 \text{ for } \tau > 0$$

$$\partial\left[\tfrac{1}{2}\Delta_\tau^2 f(\bar{x} \,|\, \bar{v})\right] = \Delta_\tau[\partial f](\bar{x} \,|\, \bar{v})$$

Second-order epi-derivatives:

$$\tfrac{1}{2} d^2 f(\bar{x} \,|\, \bar{v})(w) = \liminf_{w' \to w, \tau \searrow 0} \tfrac{1}{2}\Delta_\tau^2 f(\bar{x} \,|\, \bar{v})(w')$$

$$\operatorname{epi}\left[\tfrac{1}{2} d^2 f(\bar{x} \,|\, \bar{v})\right] = \limsup_{\tau \searrow 0} \operatorname{epi}\left[\tfrac{1}{2}\Delta_\tau^2 f(\bar{x} \,|\, \bar{v})\right]$$

Twice epi-differentiability: when the "limsup" is 'lim"

Connection with graphical derivatives of ∂f for convex f

f twice epi-differentiable $\iff$ ∂f proto-differentiable, and then

$$D[\partial f](\bar{x} \,|\, \bar{v}) = \partial[\tfrac{1}{2} d^2 f(\bar{x} \,|\, \bar{v})]$$

a consequence of Attouch's theorem

Geometric Connection with "Curvature"

making use of the correspondence $C \longleftrightarrow \delta_C$

Second-order epi-derivatives of an indicator function:

although $d\delta_C(\bar{x})$ is an indicator, $d^2\delta_C(\bar{x}\,|\,\bar{v})$ usually isn't!

Example: $C = \{x \,|\, g(x) \le 0\}$, $g(\bar{x}) = 0$, $\nabla g(\bar{x}) \ne 0$

for $\bar{v} = \nabla g(\bar{x})$,

$$d^2\delta_C(\bar{x}\,|\,\bar{v})(w) = \begin{cases} w \cdot \nabla^2 g(\bar{x}) w & \text{if } \nabla g(\bar{x}) \cdot w = 0, \\ \infty & \text{otherwise} \end{cases}$$

the **curvature** of C at $\bar{x}$ is reflected in this

Example without curvature: C polyhedral convex, $\bar{v} \in N_C(\bar{x})$

$d^2\delta_C(\bar{x}\,|\,\bar{v}) = \delta_{K(\bar{x}\,|\,\bar{v})}$ for $K(\bar{x}\,|\,\bar{v}) = \{w \in T_C(\bar{x}) \,|\, \bar{v} \cdot w = 0\}$

"critical cone"

Application to Optimality Conditions

Problem: minimize a proper lsc function f on R^n

Second-order conditions for a local minimum at $\bar{x}$

necessary: $0 \in \partial f(\bar{x}),\quad d^2 f(\bar{x}\,|\,0)(w) \geq 0$ for all $w \neq 0$

sufficient: $0 \in \partial f(\bar{x}),\quad d^2 f(\bar{x}\,|\,0)(w) > 0$ for all $w \neq 0$

$\iff\ \exists\, \varepsilon > 0$ such that $f(x) \geq f(\bar{x}) + \varepsilon |x - \bar{x}|^2$ for x near $\bar{x}$

Follow-up: calculus rules for determining the second-order epi-derivatives of various functions f from their structure

Additional theory: "parabolic" derivatives, in a kind of duality

---

But: this isn't the only approach in variational analysis to second-order sufficient conditions for a local minimum

Recent developments: using "augmented Lagrangian functions"
a primal-dual saddle point approach aimed at numerical methods

Further Study

- Set convergence is covered in Chapter 4 of Variational Analysis, and its extension to epi-convergence is in Chapter 7. Graphical convergence of set-valued mappings is in Chapter 5.

- The special results on convex functions and their subgradient mappings come later in Variational Analysis. Wijsman's theorem is in Chapter 11. The theorems of Poliquin and Attouch are covered as part of the theory of monotone mappings in Chapter 12.

- The second-order theory at the end of this lecture draws on Chapter 13 of Variational Analysis. Parabolic derivatives and connections with second-order tangent cones are there as well.

- The Lagrangian approach to second-order optimality is more recent. For that, see my article #256 as a start.

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