AN OVERVIEW OF VARIATIONAL ANALYSIS

3. SUBGRADIENTS AND OPTIMALITY

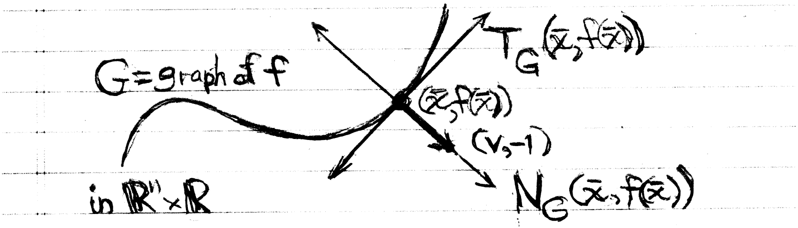
Terry Rockafellar
University of Washington, Seattle

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)

Classical Geometric Perspective on Differentiation



Tangent space: $T_G(\bar{x}, f(\bar{x}))$ = hyperplane giving the graph of the directional derivative function: $w \mapsto \nabla f(x) \cdot w$

Normal space: $N_G(\bar{x}, f(\bar{x}))$ = the one-dim. subspace $\perp$ to that

Local linearization: $f \approx l$, where $l(x) = f(\bar{x}) + \nabla f(\bar{x}) \cdot (x - \bar{x})$

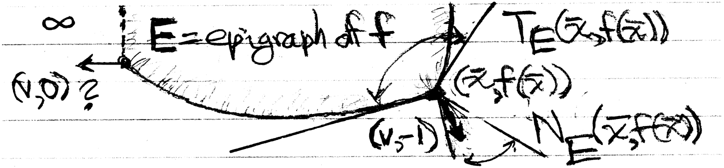
$$f(x) = f(\bar{x}) + \nabla f(\bar{x}) \cdot (x - \bar{x}) + o(||x - \bar{x}||)$$

Insight

$\nabla f(\bar{x})$ is the unique v such that $(v, -1) \in N_G(\bar{x}, f(\bar{x}))$

The Geometric Perspective in Convex Analysis

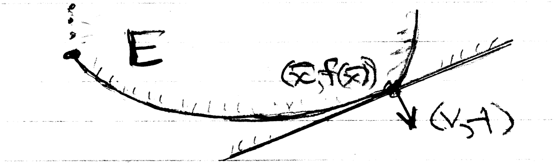
tangents and normals to epigraphs that are convex sets



Subderivatives: $T_E(\bar{x}, f(\bar{x}))$ = directional derivative epigraph

Subgradients: $(v, -1) \in N_E(\bar{x}, f(\bar{x})) \iff v \in \partial f(\bar{x})$

$f(x) \geq f(\bar{x}) + v \cdot (x - \bar{x})$ for all x



$(v, 0)$ normal to epigraph $\iff$ v normal to the convex domain

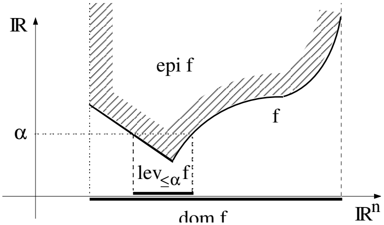
The Geometric Perspective of Variational Analysis

variational geometry of epigraphs that are just closed sets

Notation: for a function $f : R^n \to \overline{R} = [-\infty, \infty]$

$$\mathrm{epi}\, f = \{(x, \alpha \in R^n \times R \mid f(x) < \infty\} \quad \textit{epigraph}$$

$$\mathrm{dom}\, f = \{x \in R^n \mid f(x) < \infty\} \quad \textit{effective domain}$$



$\mathrm{epi}\, f$ is a closed set $\Longleftrightarrow$

f is lower semicontinuous (lsc)

Properness: f is called proper when

$$\forall x,\ f(x) > -\infty, \text{ and } \exists\, x \text{ with } f(x) < \infty$$

i.e., $\mathrm{epi}\, f$ and $\mathrm{dom}\, f$ are $\neq \emptyset$ and $\mathrm{epi}\, f$ contains no vertical lines

Main Features of This “Geometric Calculus” Approach

relying on tangent and normal vectors to epigraphs

- **subderivatives** are associated with **tangent** cones to $\operatorname{epi} f$
- **subgradients** are associated with **normal** cones to $\operatorname{epi} f$

- **regular** and **general** tangent and normal cones yield different kinds of subderivatives and subgradients
- the regular and general kinds coincide wherever the $\operatorname{epi} f$ is **variational regular** (= the variational regularity of f there)

- horizontal normals to $\operatorname{epi} f$ furnish **horizon** subgradients, which might not just be normals to $\operatorname{dom} f$
- horizon subgradients support rules of **subdifferential calculus**

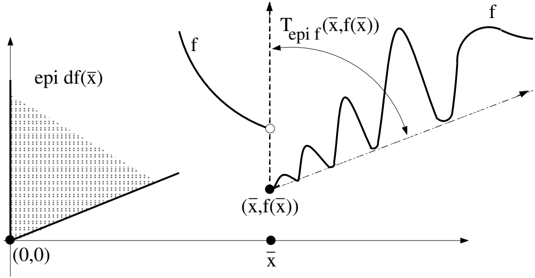
Subderivatives in General

consider proper lsc f on $\boldsymbol{R}^n$ and $\bar{x} \in \operatorname{dom} f$

Definition: [general] subderivative of f at $\bar{x}$ for a vector $w \in \boldsymbol{R}^n$

$$df(\bar{x})(w) = \liminf_{\substack{w' \to w \\ \tau \searrow 0}} \frac{f(\bar{x} + \tau w') - f(\bar{x})}{\tau} \qquad df(\bar{x}) : w \mapsto df(\bar{x})(w)$$

epigraph of $df(\bar{x})$ = the tangent cone to $\operatorname{epi} f$ at $(\bar{x}, f(\bar{x}))$



Regular subderivatives: $\widehat{d}f(\bar{x})$, epigraph = regular tangent cone

Subgradient Definitions

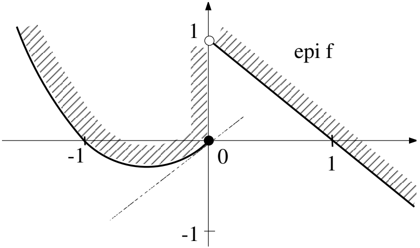
Regular subgradients: $v \in \widehat{\partial} f(\bar{x})$

$f(x) \geq f(\bar{x}) + v{\cdot}(x - \bar{x}) + o(||x - \bar{x}||)$ classical error term

General subgradients: $v \in \partial f(\bar{x})$

for some $x^\nu \to \bar{x}$ with $f(x^\nu) \to f(\bar{x})$, $\exists\, v^\nu \in \widehat{\partial} f(x^\nu)$ with $v^\nu \to v$

the need for $f(x^\nu) \to f(\bar{x})$:



Horizon subgradients: $v \in \partial^\infty f(\bar{x})$

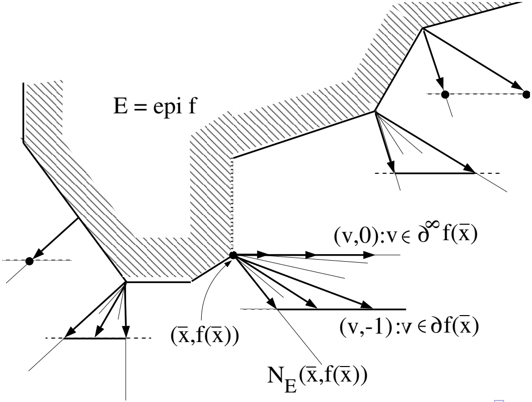
for some $x^\nu \to \bar{x}$ with $f(x^\nu) \to f(\bar{x})$,

$\exists\, v^\nu \in \widehat{\partial} f(x^\nu)$ and $\lambda^\nu \searrow 0$, with $\lambda^\nu v^\nu \to v$

Normal Cone Characterizations of Subgradients

Descriptions in terms of $E = \text{epi}\, f$

$$v \in \widehat{\partial} f(\bar{x}) \iff (v, -1) \in \widehat{N}_E(\bar{x}, f(\bar{x}))$$
$$v \in \partial f(\bar{x}) \iff (v, -1) \in N_E(\bar{x}, f(\bar{x}))$$
$$v \in \partial^\infty f(\bar{x}) \iff (v, 0) \in N_E(\bar{x}, f(\bar{x}))$$



Connections With Lipschitz Continuity

Recall: f is locally Lipschitz continuous around $\bar{x}$ $\iff$

$\exists\, \kappa \geq 0$ such that $|f(x') - f(x)| \leq \kappa|x' - x|$ for x, x' near $\bar{x}$

Horizon subgradient criterion for Lipschitz continuity

f is (finite and) Lipschitz continuous on a neighborhood of $\bar{x}$

$\iff$ the only horizon subgradient $v \in \partial^\infty f(\bar{x})$ is $v = 0$

$\iff$ the subgradient set $\partial f(\bar{x})$ is nonempty and bounded

Regular subderivatives then: for f Lip. continuous around $\bar{x}$

$$\widehat{d}f(\bar{x})(w) = \limsup_{x' \to \bar{x},\, \tau \searrow 0} \frac{f(x' + \tau w) - f(x')}{\tau}$$

= the Clarke directional derivative function, convex!

Contrast then to: $$df(\bar{x})(w) = \liminf_{\tau \searrow 0} \frac{f(\bar{x} + \tau w) - f(\bar{x})}{\tau}$$

Benefits of Variational Regularity

Equivalent properties when f is locally Lipschitz continuous

f is **variationally regular** at $\bar{x}$ i.e., $E = \operatorname{epi} f$ is, at $(\bar{x}, f(\bar{x}))$

$\Longleftrightarrow$ every subgradient $v \in \partial f(\bar{x})$ is regular: $v \in \widehat{\partial} f(\bar{x})$

$$\Longleftrightarrow \quad df(\bar{x})(w) = \widehat{d} f(\bar{x})(w) = \lim_{\tau \searrow 0} \frac{f(\bar{x} + \tau w) - f(\bar{x})}{\tau} = f'(\bar{x}; w)$$

$$\Longrightarrow \quad f(\bar{x} + w) = f(\bar{x}) + df(\bar{x})(w) + o(|w|)$$

Subderivative-subgradient duality — in this especially nice case

$$df(\bar{x})(w) = \max\left\{ v{\cdot}w \,\middle|\, v \in \partial f(\bar{x}) \right\} \quad \text{finite convex function}$$

$$\partial f(\bar{x}) = \left\{ v \,\middle|\, df(\bar{x})(w) \geq v{\cdot}w, \; \forall w \right\} \quad \text{compact convex set}$$

Comparison with classical formula: $f'(\bar{x}; w) = \nabla f(\bar{x}){\cdot}w$

$\partial f(\bar{x}) = \text{singleton } \{\nabla f(\bar{x})\} \longleftrightarrow f$ "strictly" differentiable at $\bar{x}$

Reconnection With Sets Through Indicator Functions

Specialization to indicators: $\delta_C(x) = \begin{cases} 0 \text{ if } x \in C, \\ \infty \text{ if } x \notin C \end{cases}$

C closed $\iff$ δ_C is lsc, C convex $\iff$ δ_C is convex

Indicator characterization of tangents and normals

$$\partial\delta_C(\bar{x}) = \partial^\infty\delta_C(\bar{x}) = N_C(\bar{x}), \qquad \widehat{\partial}\delta_C(\bar{x}) = \widehat{N}_C(\bar{x})$$
$$d\delta_C(\bar{x}) = \text{indicator of } T_C(\bar{x}), \qquad \widehat{d}\delta_C(\bar{x}) = \text{indicator of } \widehat{T}_C(\bar{x})$$

regularity of δ_C at $\bar{x}$ $\longleftrightarrow$ regularity of C at $\bar{x}$

Note: up to now, $N_C(x)$ has only had meaning for $x \in C$, but by by taking $N_C(x) = \emptyset$ for $x \notin C$ to define $N_C : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$, we get

$$N_C(x) = \partial\delta_C(x) \text{ for all } x \in \mathbb{R}^n$$

Important example: C = closed convex cone K with polar K^*

$$v \in N_K(x) \iff x \in K,\ v \in K^*,\ x \cdot v = 0 \iff x \in N_{K^*}(v)$$

$$N_{K^*} = N_K^{-1}$$

Application to First-Order Optimality

Generalization of Fermat's rule:

local minimum of f at $\bar{x} \implies 0 \in \widehat{\partial} f(\bar{x}) \implies 0 \in \partial f(\bar{x})$

Illustration: minimizing $f_0(x)$ over $x \in C$ for a $\mathcal{C}^1$ function f_0
equivalent to minimizing $f = f_0 + \delta_C$ over $\boldsymbol{R}^n$

Rule of subdifferential calculus for sums $f = f_1 + f_2$

Suppose at $\bar{x}$ that $\nexists$ nonzero $v \in \partial^\infty f_1(\bar{x})$ with $-v \in \partial^\infty f_2(\bar{x})$ (constraint qual.). Then $\partial f(\bar{x}) \subset \partial f_1(\bar{x}) + \partial f_2(\bar{x})$ and more...

Application to the case at hand: where $f = f_0 + \delta_C$

$\partial f_0(\bar{x}) = \{\nabla f_0(\bar{x})\}$, $\partial^\infty f_0(\bar{x}) = \{0\}$, $\partial \delta_C(\bar{x}) = \partial^\infty \delta_C(\bar{x}) = N_C(\bar{x})$

$\implies \partial f(\bar{x}) \subset \{\nabla f_0(\bar{x})\} + N_C(\bar{x})$, so that

$0 \in \partial f(\bar{x}) \implies -\nabla f_0(\bar{x}) \in N_C(\bar{x})$ as a *necessary* condition

Bringing in Lagrange Multipliers

Example: building on the condition $-\nabla f_0(\bar{x}) \in N_C(\bar{x})$ when

$\bar{x} \in C = \{x \in R^n \mid (f_1(x), \dots, f_m(x)) \in D\}$, $f_i \in \mathcal{C}^1$, for $D = R^m_-$

constraint system: $f_1(x) \leq 0, \dots, f_m(x) \leq 0$

aim: apply formula for $N_C(\bar{x})$ via $N_D(f_1(\bar{x}), \dots, f_m(\bar{x}))$

Recall complementary slackness condition: on multipliers y_i

(S) $y_i \geq 0$ if $f_i(\bar{x}) = 0$, but $y_i = 0$ if $f_i(\bar{x}) < 0$

$\iff$ $y = (y_1, \dots, y_m)$ belongs to $N_D(f_1(\bar{x}), \dots, f_m(\bar{x}))$!

Specializing the normal cone formula given earlier:

If (Q) only $y = 0$ satisfies (S) with $\sum_{i=1}^m y_i \nabla f_i(\bar{x}) = 0$, then

$v \in N_C(\bar{x}) \iff \exists\, \bar{y}$ satisfying (S) with $\sum_{i=1}^m \bar{y}_i \nabla f_i(\bar{x}) = v$

Resulting necessary condition – under constraint qualification (Q)

Local optimality of $\bar{x}$ in minimizing of f_0 over C entails

$\exists\, \bar{y}$ satisfying (S) such that $\nabla f_0(\bar{x}) + \sum_{i=1}^m \bar{y}_i \nabla f_i(\bar{x}) = 0$

Second-Order Conditions Linked to Local Optimality?

General context: minimizing f over R^n, $0 \in \partial f(\bar{x})$

what more about f at $\bar{x}$ may help to characterize a local min?

What purpose is to be served? what kind of **help**?

- traditional outlook: trying to eliminate as far as possible any "false candidates" among points $\bar{x}$ with $0 \in \partial f(\bar{x})$
- modern outlook: identifying typical properties of a local minimizer that promote numerical methods for finding it

after all, even finding $\bar{x}$ with $0 \in \partial f(\bar{x})$ may require computation

Second-order variational analysis: still to be explained

- second-order subderivatives of f
- graphical derivatives of ∂f
- local monotonicity properties of ∂f

Further Study

- Details about the topic of this lecture can be learned from the first half of Chapter 8 of *Variational Analysis*. It's possible to get into that more or less directly from Chapter 6, without studying Chapter 7. (The basics of Chapter 7 will anyway be in Lecture 4.)

- Second-order variational analysis and its role in sufficient conditions for local optimality is an active research subject which is full of interesting developments, beyond those in Chapter 13. More will be said about it in the next two lectures, but there is much that can't be covered.