Finding Gradients: parameterized functions

Monday, April 12, 2021 6:42 AM

Stephen Becker, 2021 [Caveat: check the original references for all technical assumptions]

Max $\boxed{f(x) = \max_{z \in Z} \varphi(x,z)}$, want ∇f or subdifferential ∂f

$x \in \mathbb{R}^n$, define $Z(x) = \operatorname{argmax}_z \varphi(x,z)$

Theorem "Danskin" (ref.: Prop. 4.5.1 Bertsekas "Convex Analysis + Optimization" '03)

Let Z be compact, $\varphi : \mathbb{R}^n \times Z \to \mathbb{R}$ continuous, and $\forall z \in Z$, $\varphi(\cdot, z) : \mathbb{R}^n \to \mathbb{R}$ is convex, then

1) f is convex and its directional derivative in direction d, D_d, is

$$D_d = \max_{\underbrace{z \in Z(x)}_{\text{over maximizers}}} D_d \varphi(x,z)$$

w.r.t. x only

and if $Z(x)$ is a singleton, then f is differentiable at x

2) if $\varphi(\cdot, z)$ is differentiable (in x) $\forall z \in Z$, and $\nabla_x \varphi(x, \cdot)$ is continuous in z $\forall x$, then the subdifferential is

$$\partial f(x) = \underbrace{\operatorname{conv}}_{\text{convex hull}} \{ \nabla_x \varphi(x,z) : z \in Z(x) \}$$

Ex $f(x) = \max\{x, -x\} = |x| = \max_{z \in \{-1,1\}} \{ \varphi(x,z) \}$ $\quad \varphi(x,z) = \begin{cases} x & z = 1 \\ -x & z = -1 \end{cases}$

Theorem doesn't apply since Z is discrete so φ can't be continuous in z

Then, use:

Theorem "Dubovitskii and Milyutin" (ref. Thm. 18.5 Bauschke + Combettes '17)

Let Z be a finite set and $\forall z$, $\varphi(\cdot, z)$ is convex and continuous (in x).

Then $\partial f(x) = \operatorname{conv} \{ \bigcup_{\underbrace{z \in Z(x)}_{\text{all maximizers}}} \partial \varphi(x,z) \}$

w.r.t. x

Min $\boxed{f(x) = \inf_z \varphi(x,z)}$ (allow a domain $z \in Z$ by allowing $\varphi(x,z) = +\infty$)

$x \in \mathbb{R}^n, z \in \mathbb{R}^m$ Analogously to before, define $Z(x) = \operatorname{argmin}_z \varphi(x,z)$

Theorem (ref. Thm 10.13 Rockafellar and Wets "Variational Analysis" '97)

Assume $\varphi \in \Gamma_0(\mathbb{R}^n \times \mathbb{R}^m)$ (ie, jointly convex, lsc, proper)

and φ is LBLU (see below), then

1) f is convex

2) $\partial f(x) = \partial \varphi(x, z_x)$ for any $z_x \in Z(x)$

w.r.t. x

LBLU = Level Bounded Locally Uniformly

(A sufficient condition is if $\varphi(x,z) = +\infty$ if $z \notin C$ for some bounded set C, hence this is similar to the assumption in Danskin's Thm)

Details:

$\text{lev}_{\leq\alpha} f := \{x : f(x) \leq \alpha\}$ are sub-level sets.

Rockafellar and Wets define a function f to be level bounded if all $\text{lev}_{\leq\alpha} f$ are bounded (i.e., $\forall \alpha \in \mathbb{R}$)

Note: via Bauschke, Combettes, $f : \mathcal{H} \to \overline{\mathbb{R}}$ coercive if $\lim_{\|x\| \to \infty} f(x) = \infty$, and this is equivalent to level-bounded (Prop. 11.12 Bauschke, Combettes)

In fact, if $\mathcal{H}$ finite-dimensional and $f \in \Gamma_0(\mathcal{H})$, it's sufficient to show $\exists\, \alpha \in \mathbb{R}$ s.t. $\text{lev}_{\leq\alpha} f \neq \emptyset$ is bounded (Prop. 11.13)

Then

Def LBLU (def. 1.16 Rockafellar + Wets)

$\varphi : \mathbb{R}^n \times \mathbb{R}^m \to \overline{\mathbb{R}}$ is Level Bounded (in z) Locally Uniformly (in x)

if $\forall\, x_0 \in \mathbb{R}^n$ $\forall\, \alpha \in \mathbb{R}$, $\exists$ a neighborhood V of x_0 and a bounded set $B \subseteq \mathbb{R}^m$ s.t. $\forall\, x \in V$, $\{z : \varphi(x,z) \leq \alpha\} \underset{\text{bounded}}{\underline{\subseteq B}}$

uniformly locally

(special case: Fenchel-Legendre conjugates)

$f^*(y) = \sup_x \underbrace{\langle x,y\rangle - f(x)}_{\varphi(y,x)}$ ... but think of as negative infimum since want to exploit convexity of f

Unique minimizer guaranteed if f is strictly convex (Prop. 18.9 Bauschke + Combettes '17)

* fundamental theorem

Theorem (Thm. 18.15 Bauschke + C.)

f is differentiable and has a L-Lipschitz gradient

iff

f^* is $\mu = 1/L$ strongly convex

(and can swap f, f^*)

cf. Goebel + Rockafellar '07 "Local strong convexity of..." for local results

Integrals

$$\boxed{f(x) = \int_\Omega \varphi(x,z)\,dz}$$

cf. Kim Border's notes for Ma 3 '20 at Caltech "Supplement 4: Differentiating under an integral sign"

Theorem (informal) (refs: Border, or Aliprantis + Burkinshaw "Principles of Real Analysis" '98 3rd ed.)

Assume $\forall x$, $\underbrace{\varphi(x,\cdot) \in L^1(\Omega)}_{\text{i.e., } \int_\Omega |\varphi(x,z)|\,dz < \infty}$, $\frac{\partial}{\partial x}\varphi(x,\cdot)$ exists and is also in L^1

and assume a uniform local integrability condition (refs: Aliprantis + Burkinshaw p. 193–194 or Dudley, Appendix A p. 417–423 "uniform CLTs" 2nd ed.)

(sufficient: Ω is bounded, $\frac{\partial}{\partial x}\varphi$ is jointly continuous) (refs: Dieudonné p. 177 Thm 8.11.2, 1969 "Foundations of Modern Analysis")

then f is differentiable and

$$\frac{d}{dx} f(x) = \int \frac{\partial}{\partial x}\varphi(x,z)\,dz$$

Counterexamples (when assumptions not met) in Gelbaum + Olmsted, p. 123 Ex. 9.15 "Counterexamples in Analysis" 2003, 1965

variant

$$\boxed{f(x) = \int_{a(x)}^{b(x)} \varphi(x,z)\,dz}$$

Theorem Leibniz Integral Rule (refs: wikipedia) $x \in \mathbb{R}^1$, $z \in \mathbb{R}^1$

Assume φ and $\frac{\partial}{\partial x}\varphi$ are jointly continuous in (x,z) and $a(x)$, $b(x)$ continuously differentiable, then

$$\frac{d}{dx} f(x) = \int_{a(x)}^{b(x)} \frac{\partial}{\partial x}\varphi(x,z)\,dz + f(x, b(x))\cdot b'(x) - f(x,a(x))\cdot a'(x)$$

i.e., Fundamental Theorem of Calculus

Generally, use Lebesgue's Dominated Convergence Theorem to prove these results

Kim Border's notes are cached here:
https://healy.econ.ohio-state.edu/kcb/Ma103/Notes/DifferentiatingAnIntegral.pdf

Bauschke and Combettes '17 is this book:
"Convex Analysis and Monotone Operator Theory in Hilbert Spaces" 2017, Springer