Finding Gradients: parameterized functions

Monday, April 28, 2025 3:04 PM

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

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

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

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 $Z(x) = \operatorname{argmin}_z \varphi(x,z)$

see also their Thm. 10.58 for similar results

Theorem (ref. Thm 10.13 Rockafellar and Wets "Variational Analysis" '97) or '09 3rd printing

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

Parameterized fcn: p. 2

Monday, April 28, 2025 3:09 PM

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:

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

Rockafeller and Wets define a function f to be level bounded if all $\mathrm{lev}_{\le\alpha} f$ 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. $\mathrm{lev}_{\le\alpha} f \ne \emptyset$ is bounded (Prop. 11.13)

Then

Def LBLU (def. 1.16 Rockafeller + 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. $\forall\, x \in V$, $\{z : \varphi(x,z) \le \alpha\} \subseteq \underbrace{B}_{\text{bounded}}$

uniformly locally

(special case: Fenchel-Legendre conjugates)

$f^*(x) = \sup_z \underbrace{\langle z, x\rangle - f(z)}_{\varphi(x,z)}$ ... 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 + Rockafeller '07 "Local strong convexity of..." for local results

Theorem Prop. 12.30 Bauschke + Combettes

Define ${}^{\gamma}f(x) := \inf_z f(z) + \frac{1}{2\gamma}\|z-x\|^2$ "Moreau envelope of f"

If $f \in \Gamma_0(\mathbb{R}^n)$ and $\gamma > 0$ then ${}^{\gamma}f$ is Fréchet differentiable and its gradient, $\nabla({}^{\gamma}f) = \frac{1}{\gamma}(\mathrm{Id} - \mathrm{prox}_{\gamma f})$, is $\frac{1}{\gamma}$ Lipschitz continuous

See also §18.3 "Differentiability of Infimal Convolutions", also §18.4

Parameterized fcn: p. 3

Monday, April 28, 2025 3:09 PM

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

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

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

Parameterized fcn: p. 4

Monday, August 4, 2025 4:27 PM

Bi-level optimization

$$\min_{x \in X,\, y^* \in Y} f(x, y^*) \quad \text{s.t.} \quad y^* \in \operatorname{argmin}_{y \in Y} g(x,y)$$

Regularizing ERM

Example: $f(x,y^*) = f(y^*)$

$$\underbrace{\text{loss}(y^*;\ \text{validation-data})}$$

$$\min_{\lambda > 0,\, y} \quad \text{"}x\text{"} \to \lambda$$

$$\text{s.t. } y^* \in \operatorname{argmin}_y \overbrace{\left[\text{loss}(y, \text{training data}) + \lambda \cdot \|y\|^2 \right]}^{g(\lambda, y)}$$

If y^* is unique, then $y^* = y^*(x)$ is an (implicit) function of x, so define

$$\phi(x) = f(x, y^*(x)).$$

Then (under some conditions),

basically, total derivative version of chain rule

$$\nabla_x \phi(x) = \nabla_x f(x, y^*(x)) - \nabla^2_{xy} g(x, y^*(x)) \cdot \left[\nabla^2_{yy} g(x, y^*(x)) \right]^{-1} \cdot \nabla_y f(x, y^*(x))$$

Note: not all bi-level algorithms use $\nabla \phi$, as there are workarounds

Refs: lemma 2.1 (and eq'n 2.4) in "Approximation Methods for Bilevel Programming" by Ghadimi and Wang, arXiv 1802.02246, 2018
(a well-cited preprint, never published!)

for some more recent (2024) refs, see:

See "On Penalty Methods for Nonconvex Bilevel Optimization and First-Order Stochastic Approximation" by Kwon, Kwon, Wright and Nowak, 2024, arXiv.org/pdf/2309.01753 as one example of a reference on these methods (and they have links to more references, they certainly are not the first to derive this formula)