Generalized Derivatives for Nonsmooth Optimization

Hajg Jasa

Trial Lecture
Ph.D. Defense @ IMF, NTNU, Trondheim

January 28th, 2026

Why *Nonsmooth* Optimization?

- Many optimization problems involve nondifferentiable terms:

$$\min_x |x|, \quad \min_x \|x\|_1, \quad \min_x f(x) + \iota_C(x)$$

Why *Nonsmooth* Optimization?

- Many optimization problems involve nondifferentiable terms:

$$\min_x |x|, \quad \min_x \|x\|_1, \quad \min_x f(x) + \iota_C(x)$$

- Examples:

  1. Regularization (ℓ_1, total variation)

Why *Nonsmooth* Optimization?

- Many optimization problems involve nondifferentiable terms:

$$\min_x |x|, \quad \min_x \|x\|_1, \quad \min_x f(x) + \iota_C(x)$$

- Examples:

  1. Regularization (ℓ_1, total variation)
  2. Max or indicator functions

Why *Nonsmooth* Optimization?

- Many optimization problems involve nondifferentiable terms:

$$\min_x |x|, \quad \min_x \|x\|_1, \quad \min_x f(x) + \iota_C(x)$$

- Examples:

  1. Regularization (ℓ_1, total variation)
  2. Max or indicator functions
  3. Constrained optimization

Why *Nonsmooth* Optimization?

- Many optimization problems involve nondifferentiable terms:

$$\min_x |x|, \quad \min_x \|x\|_1, \quad \min_x f(x) + \iota_C(x)$$

- Examples:

  1. Regularization (ℓ_1, total variation)
  2. Max or indicator functions
  3. Constrained optimization

Goal: Extend derivative concepts to nonsmooth settings.

Outline

First-order Concepts

Convex Subdifferential(s)

Clarke Generalization

Subgradient-based Algorithm

Second-order Concepts

General Subdifferential

Second-order Subderivative and Optimality

Newton Derivative

Semismooth Newton Method

Summary

Basic Notions

It is customary to study $f : \mathbb{R}^n \to \overline{\mathbb{R}} := [-\infty, +\infty]$

Basic Notions

It is customary to study $f : \mathbb{R}^n \to \overline{\mathbb{R}} := [-\infty, +\infty]$ with f being *proper*:

$$f(x) > -\infty \text{ for all } x \in \mathbb{R}^n, \quad \operatorname{dom} f := \{x \in \mathbb{R}^n \mid f(x) < +\infty\} \neq \emptyset$$

Basic Notions

It is customary to study $f : \mathbb{R}^n \to \overline{\mathbb{R}} := [-\infty, +\infty]$ with f being *proper*:

$$f(x) > -\infty \text{ for all } x \in \mathbb{R}^n, \quad \operatorname{dom} f := \{x \in \mathbb{R}^n \mid f(x) < +\infty\} \neq \emptyset$$

lower semicontinuous (l.s.c.) (at a point $x_0 \in \mathbb{R}^n$):

$$\text{for every } y < f(x_0), \text{ there is } \mathcal{U} \ni x_0 \text{ s.t. } f(x) > y \text{ for all } x \in \mathcal{U}$$

Basic Notions

It is customary to study $f : \mathbb{R}^n \to \overline{\mathbb{R}} := [-\infty, +\infty]$ with f being *proper*:

$$f(x) > -\infty \text{ for all } x \in \mathbb{R}^n, \quad \operatorname{dom} f := \{x \in \mathbb{R}^n \mid f(x) < +\infty\} \neq \emptyset$$

lower semicontinuous (l.s.c.) (at a point $x_0 \in \mathbb{R}^n$):

$$\text{for every } y < f(x_0), \text{ there is } \mathcal{U} \ni x_0 \text{ s.t. } f(x) > y \text{ for all } x \in \mathcal{U}$$

convex:

$$f((1-t)x + ty) \leq (1-t)f(x) + tf(y) \text{ for all } x, y \in \mathbb{R}^n, t \in [0,1]$$

Basic Notions

It is customary to study $f : \mathbb{R}^n \to \overline{\mathbb{R}} := [-\infty, +\infty]$ with f being *proper*:

$$f(x) > -\infty \text{ for all } x \in \mathbb{R}^n, \quad \operatorname{dom} f := \{x \in \mathbb{R}^n \mid f(x) < +\infty\} \neq \emptyset$$

lower semicontinuous (l.s.c.) (at a point $x_0 \in \mathbb{R}^n$):

$$\text{for every } y < f(x_0), \text{ there is } \mathcal{U} \ni x_0 \text{ s.t. } f(x) > y \text{ for all } x \in \mathcal{U}$$

convex:

$$f((1-t)x + ty) \leq (1-t)f(x) + tf(y) \text{ for all } x, y \in \mathbb{R}^n, t \in [0,1]$$

or *locally Lipschitz* (at a point $x_0 \in \mathbb{R}^n$):

$$\text{there is } \mathcal{U} \ni x_0,\ L > 0 \text{ s.t. } |f(x) - f(y)| \leq L\|x - y\| \text{ for all } x, y \in \mathcal{U}$$

NTNU

Convex Subdifferential

For a proper convex $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ and $x \in \operatorname{dom} f$

$$\partial f(x) := \{s \in \mathbb{R}^n \mid f(y) \geq f(x) + (s\,,\, y - x) \quad \forall y \in \mathbb{R}^n\}\,,$$

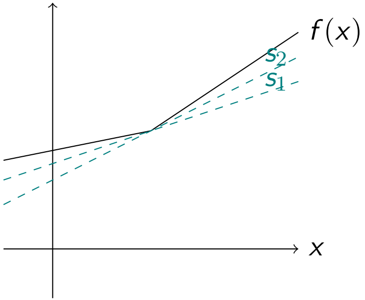
and $s \in \partial f(x)$ is called *subgradient* of f at x.

Convex Subdifferential

For a proper convex $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ and $x \in \operatorname{dom} f$

$$\partial f(x) := \{s \in \mathbb{R}^n \mid f(y) \geq f(x) + (s\,,\, y - x) \quad \forall y \in \mathbb{R}^n\}\,,$$

and $s \in \partial f(x)$ is called *subgradient* of f at x.



Convex Subdifferential

For a proper convex $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ and $x \in \operatorname{dom} f$

$$\partial f(x) := \{s \in \mathbb{R}^n \mid f(y) \geq f(x) + (s\,,\, y - x) \quad \forall y \in \mathbb{R}^n\}\,,$$

and $s \in \partial f(x)$ is called *subgradient* of f at x.

- Extends gradient to nondifferentiable convex functions. If f is differentiable
$$\partial f(x) = \{\nabla f(x)\}$$

Convex Subdifferential

For a proper convex $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ and $x \in \operatorname{dom} f$

$$\partial f(x) := \{s \in \mathbb{R}^n \mid f(y) \geq f(x) + (s\,,\, y - x) \quad \forall y \in \mathbb{R}^n\}\,,$$

and $s \in \partial f(x)$ is called *subgradient* of f at x.

- Extends gradient to nondifferentiable convex functions. If f is differentiable
$$\partial f(x) = \{\nabla f(x)\}$$
- $0 \in \partial f(x^*) \iff x^*$ is a minimizer.

Convex Subdifferential

For a proper convex $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ and $x \in \operatorname{dom} f$

$$\partial f(x) := \{s \in \mathbb{R}^n \mid f(y) \geq f(x) + (s\,,\, y - x) \quad \forall y \in \mathbb{R}^n\}\,,$$

and $s \in \partial f(x)$ is called *subgradient* of f at x.

- Extends gradient to nondifferentiable convex functions. If f is differentiable
$$\partial f(x) = \{\nabla f(x)\}$$
- $0 \in \partial f(x^*) \iff x^*$ is a minimizer.
- $\partial f(x)$ always nonempty if f is convex and $x \in \operatorname{int} \operatorname{dom} f$ or f is l.s.c.

Convex Subdifferential: Example

Consider the absolute value function on $\mathbb{R}$, $f(x) = |x|$.

Convex Subdifferential: Example

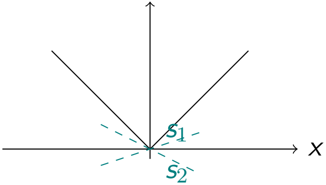
Consider the absolute value function on $\mathbb{R}$, $f(x) = |x|$.

$$\partial f(x) = \begin{cases} \{1\} & \text{if } x > 0, \\ \{-1\} & \text{if } x < 0, \\ ? & \text{if } x = 0. \end{cases}$$

Convex Subdifferential: Example

Consider the absolute value function on $\mathbb{R}$, $f(x) = |x|$.

$$\partial f(x) = \begin{cases} \{1\} & \text{if } x > 0, \\ \{-1\} & \text{if } x < 0, \\ [-1, 1] & \text{if } x = 0. \end{cases}$$



Convex Subdifferential: Example II

Consider the indicator function of a convex set $\mathcal{C} \subseteq \mathbb{R}^n$,

$$\iota_{\mathcal{C}}(x) = \begin{cases} 0 & \text{if } x \in \mathcal{C}, \\ +\infty & \text{if } x \notin \mathcal{C}. \end{cases}$$

NTNU

Convex Subdifferential: Example II

Consider the indicator function of a convex set $\mathcal{C} \subseteq \mathbb{R}^n$,

$$\iota_{\mathcal{C}}(x) = \begin{cases} 0 & \text{if } x \in \mathcal{C}, \\ +\infty & \text{if } x \notin \mathcal{C}. \end{cases}$$

$$\partial \iota_{\mathcal{C}}(x) = \{ v \in \mathbb{R}^n \,|\, \langle v \,,\, y - x \rangle \leq 0, \ \forall y \in \mathcal{C} \}$$

Convex Subdifferential: Example II

Consider the indicator function of a convex set $\mathcal{C} \subseteq \mathbb{R}^n$,

$$\iota_{\mathcal{C}}(x) = \begin{cases} 0 & \text{if } x \in \mathcal{C}, \\ +\infty & \text{if } x \notin \mathcal{C}. \end{cases}$$

$$\partial \iota_{\mathcal{C}}(x) = \{v \in \mathbb{R}^n \,|\, (v\,,\, y - x) \leq 0,\ \forall y \in \mathcal{C}\}$$

Also known as the (convex) *normal cone* $\mathcal{N}_{\mathcal{C}}(x)$ of $\mathcal{C}$ at $x \in \mathbb{R}^n$.

ε-subdifferential

If $x \in \partial \operatorname{dom} f$, it may happen that $\partial f(x) = \emptyset$.

ε-subdifferential

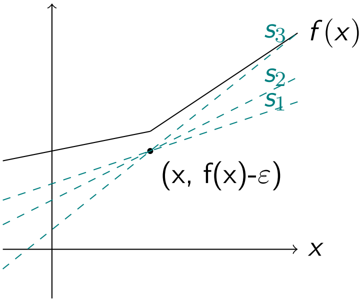
If $x \in \partial \operatorname{dom} f$, it may happen that $\partial f(x) = \emptyset$. We can *enlarge* this set to

$$\partial_\varepsilon f(x) := \{s \in \mathbb{R}^n \mid f(y) \geq f(x) + (s\,,\ y - x) - \varepsilon \quad \forall y \in \mathbb{R}^n\}\,,$$

ε-subdifferential

If $x \in \partial \operatorname{dom} f$, it may happen that $\partial f(x) = \emptyset$. We can *enlarge* this set to

$$\partial_\varepsilon f(x) := \{s \in \mathbb{R}^n \mid f(y) \geq f(x) + (s, y - x) - \varepsilon \quad \forall y \in \mathbb{R}^n\},$$



ε-subdifferential

If $x \in \partial \operatorname{dom} f$, it may happen that $\partial f(x) = \emptyset$. We can *enlarge* this set to

$$\partial_\varepsilon f(x) := \{s \in \mathbb{R}^n \,|\, f(y) \geq f(x) + (s\,,\, y - x) - \varepsilon \quad \forall y \in \mathbb{R}^n\}\,,$$

- $\partial_0 f(x) = \partial f(x) \subseteq \partial_\varepsilon f(x)$

ε-subdifferential

If $x \in \partial \operatorname{dom} f$, it may happen that $\partial f(x) = \emptyset$. We can *enlarge* this set to

$$\partial_\varepsilon f(x) := \{s \in \mathbb{R}^n \mid f(y) \geq f(x) + (s\,,\, y - x) - \varepsilon \quad \forall y \in \mathbb{R}^n\}\,,$$

- $\partial_0 f(x) = \partial f(x) \subseteq \partial_\varepsilon f(x)$
- $\partial_\varepsilon f(x) \neq \emptyset$ for all $x \in \operatorname{dom} f$

ε-subdifferential

If $x \in \partial \operatorname{dom} f$, it may happen that $\partial f(x) = \emptyset$. We can *enlarge* this set to

$$\partial_\varepsilon f(x) := \{s \in \mathbb{R}^n \mid f(y) \geq f(x) + (s\,,\, y - x) - \varepsilon \quad \forall y \in \mathbb{R}^n\}\,,$$

- $\partial_0 f(x) = \partial f(x) \subseteq \partial_\varepsilon f(x)$
- $\partial_\varepsilon f(x) \neq \emptyset$ for all $x \in \operatorname{dom} f$
- x^* is an ε-minimizer of f, i. e. $f(x^*) \leq f(x) + \varepsilon$ for all $x \in \mathbb{R}^n$, if and only if $0 \in \partial_\varepsilon f(x^*)$

ε-subdifferential

If $x \in \partial \operatorname{dom} f$, it may happen that $\partial f(x) = \emptyset$. We can *enlarge* this set to

$$\partial_\varepsilon f(x) := \{s \in \mathbb{R}^n \mid f(y) \geq f(x) + (s\,,\, y - x) - \varepsilon \quad \forall y \in \mathbb{R}^n\}\,,$$

- $\partial_0 f(x) = \partial f(x) \subseteq \partial_\varepsilon f(x)$
- $\partial_\varepsilon f(x) \neq \emptyset$ for all $x \in \operatorname{dom} f$
- x^* is an ε-minimizer of f, i. e. $f(x^*) \leq f(x) + \varepsilon$ for all $x \in \mathbb{R}^n$, if and only if $0 \in \partial_\varepsilon f(x^*)$
- can be easily approximated by polyhedral sets

NTNU

Calculus Rules

Both subdifferentials admit good rules for studying optimality conditions of general problems.

Calculus Rules

Both subdifferentials admit good rules for studying optimality conditions of general problems. Assume $f, g \colon \mathbb{R}^n \to \overline{\mathbb{R}}$ to be proper, convex, and lower semicontinuous.

- Sum rule: for all $x \in \operatorname{dom} f \cap \operatorname{dom} g$

$$\partial f(x) + \partial g(x) \subseteq \partial(f+g)(x),$$

  with equality if, e. g., there is a $x_0 \in \operatorname{int} \operatorname{dom} f \cap \operatorname{int} \operatorname{dom} g$

Calculus Rules

Both subdifferentials admit good rules for studying optimality conditions of general problems. Assume $f, g \colon \mathbb{R}^n \to \overline{\mathbb{R}}$ to be proper, convex, and lower semicontinuous.

- Sum rule: for all $x \in \operatorname{dom} f \cap \operatorname{dom} g$

  $$\partial f(x) + \partial g(x) \subseteq \partial (f+g)(x),$$

  with equality if, e. g., there is a $x_0 \in \operatorname{int} \operatorname{dom} f \cap \operatorname{int} \operatorname{dom} g$
- Additionally: $\partial(f+g)(x) = \bigcap_{\varepsilon>0} \overline{\partial_\varepsilon f(x) + \partial_\varepsilon g(x)}$,

NTNU

Calculus Rules

Both subdifferentials admit good rules for studying optimality conditions of general problems. Assume $f, g\colon \mathbb{R}^n \to \overline{\mathbb{R}}$ to be proper, convex, and lower semicontinuous.

- Sum rule: for all $x \in \operatorname{dom} f \cap \operatorname{dom} g$

  $$\partial f(x) + \partial g(x) \subseteq \partial(f+g)(x),$$

  with equality if, e. g., there is a $x_0 \in \operatorname{int}\operatorname{dom} f \cap \operatorname{int}\operatorname{dom} g$
- Additionally: $\partial(f+g)(x) = \bigcap_{\varepsilon>0} \overline{\partial_\varepsilon f(x) + \partial_\varepsilon g(x)}$,
- Chain rule: for a convex, increasing map $h\colon \mathbb{R} \to \mathbb{R}$ and $x \in \mathbb{R}^n$

  $$\partial(h \circ f) = \partial h(f(x))\partial f(x) = \{sv \mid s \in \partial h(f(x)),\ v \in \partial f(x)\}$$

NTNU

Clarke Subdifferential

What if f is not convex?

Clarke Subdifferential

What if f is not convex? For a locally Lipschitz $f : \mathbb{R}^n \to \overline{\mathbb{R}}$, the *Clarke generalized derivative* is:

$$f^\circ(x; v) = \limsup_{\substack{y \to x \\ t \downarrow 0}} \frac{f(y + tv) - f(y)}{t}.$$

Clarke Subdifferential

What if f is not convex? For a locally Lipschitz $f : \mathbb{R}^n \to \overline{\mathbb{R}}$, the *Clarke generalized derivative* is:

$$f^\circ(x; v) = \limsup_{\substack{y \to x \\ t \downarrow 0}} \frac{f(y + tv) - f(y)}{t}.$$

This can be used to define the *Clarke subdifferential*

$$\partial^\circ f(x) = \{s \in \mathbb{R}^n \,|\, (s\,,\ v) \leq f^\circ(x; v),\ \forall v \in \mathbb{R}^n\}.$$

Clarke Subdifferential

What if f is not convex? For a locally Lipschitz $f : \mathbb{R}^n \to \overline{\mathbb{R}}$, the *Clarke generalized derivative* is:

$$f^\circ(x; v) = \limsup_{\substack{y \to x \\ t \downarrow 0}} \frac{f(y + tv) - f(y)}{t}.$$

This can be used to define the *Clarke subdifferential*

$$\partial^\circ f(x) = \{s \in \mathbb{R}^n \,|\, (s\,,\, v) \leq f^\circ(x; v),\ \forall v \in \mathbb{R}^n\}.$$

- Nonempty, convex, compact set of subgradients for all $x \in \mathbb{R}^n$.

Clarke Subdifferential

What if f is not convex? For a locally Lipschitz $f : \mathbb{R}^n \to \overline{\mathbb{R}}$, the *Clarke generalized derivative* is:

$$f^{\circ}(x; v) = \limsup_{\substack{y \to x \\ t \downarrow 0}} \frac{f(y + tv) - f(y)}{t}.$$

This can be used to define the *Clarke subdifferential*

$$\partial^{\circ} f(x) = \{s \in \mathbb{R}^n \,|\, (s\,,\, v) \leq f^{\circ}(x; v),\ \forall v \in \mathbb{R}^n\}.$$

- Nonempty, convex, compact set of subgradients for all $x \in \mathbb{R}^n$.
- Also reduces to the classical gradient when f is continuously differentiable.

Clarke Subdifferential

What if f is not convex? For a locally Lipschitz $f : \mathbb{R}^n \to \overline{\mathbb{R}}$, the *Clarke generalized derivative* is:

$$f^\circ(x; v) = \limsup_{\substack{y \to x \\ t \downarrow 0}} \frac{f(y + tv) - f(y)}{t}.$$

This can be used to define the *Clarke subdifferential*

$$\partial^\circ f(x) = \{s \in \mathbb{R}^n \,|\, (s\,,\, v) \leq f^\circ(x; v),\ \forall v \in \mathbb{R}^n\}.$$

- Nonempty, convex, compact set of subgradients for all $x \in \mathbb{R}^n$.
- Also reduces to the classical gradient when f is continuously differentiable.
- If x^* is a *local* minimizer of f, then $0 \in \partial^\circ f(x^*)$.

Calculus Rules II

Assume $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is locally Lipschitz.

- Sum rule and chain rule are analogous to ∂f

Calculus Rules II

Assume $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is locally Lipschitz.

- Sum rule and chain rule are analogous to ∂f

- Lebourg's mean value theorem: for any $x, y \in \mathbb{R}^n$ there exist $z \in (x, y)$ and $v \in \partial^\circ f(z)$ such that

$$f(y) - f(x) = (v\,,\ y - x)$$

Calculus Rules II

Assume $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is locally Lipschitz.

- Sum rule and chain rule are analogous to ∂f
- Lebourg's mean value theorem: for any $x, y \in \mathbb{R}^n$ there exist $z \in (x, y)$ and $v \in \partial^\circ f(z)$ such that

$$f(y) - f(x) = (v\,,\, y - x)$$

- Convex hull characterization:

$$\partial^\circ f(x) = \mathrm{conv}\{v \in \mathbb{R}^n \mid v = \lim_{k\to\infty} \nabla f(x^{(k)}),\ x^{(k)} \to x,\ x^{(k)} \in S\}$$

where S is the set of points where f is differentiable.

Subgradient-based Algorithm

Goal: Minimize a proper, l.s.c. function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$

Subgradient-based Algorithm

Goal: Minimize a proper, l.s.c. function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$

Data: $f, \partial f, x_0 \in \mathbb{R}^n$, step-size sequence $\{\alpha^{(k)}\}_{k \geq 0}$, $k = 0$.

while *convergence criterion is not fulfilled* **do**

  Compute $g^{(k)} \in \partial f(x^{(k)})$

  Set $x^{(k+1)} = x^{(k)} - \alpha^{(k)} g^{(k)}$

  Set $k := k + 1$

end

Result: $x^{(k_*)}$

Subgradient-based Algorithm

Goal: Minimize a proper, l.s.c. function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$

Data: $f, \partial f, x_0 \in \mathbb{R}^n$, step-size sequence $\{\alpha^{(k)}\}_{k \geq 0}$, $k = 0$.

while *convergence criterion is not fulfilled* **do**

  Compute $g^{(k)} \in \partial f(x^{(k)})$

  Set $x^{(k+1)} = x^{(k)} - \alpha^{(k)} g^{(k)}$

  Set $k := k + 1$

end

Result: $x^{(k_*)}$

Step-size choices:

- Diminishing: $\alpha^{(k)} \downarrow 0$ with $\sum \alpha^{(k)} = \infty$ (e. g. $a/\sqrt{k}$, with $a > 0$)
- Polyak: $\alpha^{(k)} = \dfrac{f(x^{(k)}) - f^*}{\|g^{(k)}\|^2}$ (requires f^* known or estimated).

Subgradient-based Algorithm

Goal: Minimize a proper, l.s.c. function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$

Data: $f, \partial f, x_0 \in \mathbb{R}^n$, step-size sequence $\{\alpha^{(k)}\}_{k \geq 0}$, $k = 0$.

while *convergence criterion is not fulfilled* **do**
  Compute $g^{(k)} \in \partial f(x^{(k)})$
  Set $x^{(k+1)} = x^{(k)} - \alpha^{(k)} g^{(k)}$
  Set $k := k + 1$
end

Result: $x^{(k_*)}$

Step-size choices:

- Diminishing: $\alpha^{(k)} \downarrow 0$ with $\sum \alpha^{(k)} = \infty$ (e. g. $a/\sqrt{k}$, with $a > 0$)
- Polyak: $\alpha^{(k)} = \dfrac{f(x^{(k)}) - f^*}{\|g^{(k)}\|^2}$ (requires f^* known or estimated).

- Sublinear convergence to (approximate) minimizer if f is convex
- Convergence to stationary points if f is locally Lipschitz

Second-Order Motivation

- First-order subgradients describe local slopes.
  - Necessary optimality conditions, only sufficient for convexity.

Second-Order Motivation

- First-order subgradients describe local slopes.
  - Necessary optimality conditions, only sufficient for convexity.
- **Second-order generalized derivatives** provide curvature information.
  - This enables second-order necessary and sufficient optimality conditions for nonconvex functions.

General Subdifferential

For a proper function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ and a point $x \in \mathbb{R}^n$ where f is finite, a vector $v \in \mathbb{R}^n$ is a

General Subdifferential

For a proper function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ and a point $x \in \mathbb{R}^n$ where f is finite, a vector $v \in \mathbb{R}^n$ is a *regular subgradient* of f at x, $v \in \hat{\partial} f(x)$ if

$$f(y) \geq f(x) + (v\,,\, y - x) + o(\|y - x\|),$$

i. e.

$$\liminf_{y \to x, y \neq x} \frac{f(y) - f(x) - (v\,,\, y - x)}{\|y - x\|} \geq 0,$$

NTNU

General Subdifferential

For a proper function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ and a point $x \in \mathbb{R}^n$ where f is finite, a vector $v \in \mathbb{R}^n$ is a *regular subgradient* of f at x, $v \in \hat{\partial} f(x)$ if

$$f(y) \geq f(x) + (v\,,\, y - x) + o(\|y - x\|),$$

i. e.

$$\liminf_{y \to x, y \neq x} \frac{f(y) - f(x) - (v\,,\, y - x)}{\|y - x\|} \geq 0,$$

or a *general subgradient* of f at x, $v \in \partial f(x)$, if there are sequences $x^{(k)} \to_f x$ and $v^{(k)} \in \hat{\partial} f(x^{(k)})$ such that $v^{(k)} \to v$.

General Subdifferential

For a proper function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ and a point $x \in \mathbb{R}^n$ where f is finite, a vector $v \in \mathbb{R}^n$ is a *regular subgradient* of f at x, $v \in \hat{\partial} f(x)$ if

$$f(y) \geq f(x) + (v\,,\, y - x) + o(\|y - x\|),$$

i. e.

$$\liminf_{y \to x, y \neq x} \frac{f(y) - f(x) - (v\,,\, y - x)}{\|y - x\|} \geq 0,$$

or a *general subgradient* of f at x, $v \in \partial f(x)$, if there are sequences $x^{(k)} \to_f x$ and $v^{(k)} \in \hat{\partial} f(x^{(k)})$ such that $v^{(k)} \to v$.

- both sets are closed, with $\hat{\partial} f(x)$ being convex and $\hat{\partial} f(x) \subseteq \partial f(x)$

General Subdifferential

For a proper function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ and a point $x \in \mathbb{R}^n$ where f is finite, a vector $v \in \mathbb{R}^n$ is a *regular subgradient* of f at x, $v \in \hat{\partial} f(x)$ if

$$f(y) \geq f(x) + (v\,,\, y - x) + o(\|y - x\|),$$

i. e.

$$\liminf_{y \to x, y \neq x} \frac{f(y) - f(x) - (v\,,\, y - x)}{\|y - x\|} \geq 0,$$

or a *general subgradient* of f at x, $v \in \partial f(x)$, if there are sequences $x^{(k)} \to_f x$ and $v^{(k)} \in \hat{\partial} f(x^{(k)})$ such that $v^{(k)} \to v$.

- both sets are closed, with $\hat{\partial} f(x)$ being convex and $\hat{\partial} f(x) \subseteq \partial f(x)$
- if f is differentiable at x, $\hat{\partial} f(x) = \{\nabla f(x)\}$

General Subdifferential

For a proper function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ and a point $x \in \mathbb{R}^n$ where f is finite, a vector $v \in \mathbb{R}^n$ is a *regular subgradient* of f at x, $v \in \hat{\partial} f(x)$ if

$$f(y) \geq f(x) + (v\,,\, y - x) + o(\|y - x\|),$$

i. e.

$$\liminf_{y \to x, y \neq x} \frac{f(y) - f(x) - (v\,,\, y - x)}{\|y - x\|} \geq 0,$$

or a *general subgradient* of f at x, $v \in \partial f(x)$, if there are sequences $x^{(k)} \to_f x$ and $v^{(k)} \in \hat{\partial} f(x^{(k)})$ such that $v^{(k)} \to v$.

- both sets are closed, with $\hat{\partial} f(x)$ being convex and $\hat{\partial} f(x) \subseteq \partial f(x)$
- if f is differentiable at x, $\hat{\partial} f(x) = \{\nabla f(x)\}$
- if f is convex, both sets coincide with the convex subdifferential

General Subdifferential

For a proper function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ and a point $x \in \mathbb{R}^n$ where f is finite, a vector $v \in \mathbb{R}^n$ is a *regular subgradient* of f at x, $v \in \hat{\partial} f(x)$ if

$$f(y) \geq f(x) + (v\,,\, y - x) + o(\|y - x\|),$$

i. e.

$$\liminf_{y \to x, y \neq x} \frac{f(y) - f(x) - (v\,,\, y - x)}{\|y - x\|} \geq 0,$$

or a *general subgradient* of f at x, $v \in \partial f(x)$, if there are sequences $x^{(k)} \to_f x$ and $v^{(k)} \in \hat{\partial} f(x^{(k)})$ such that $v^{(k)} \to v$.

- both sets are closed, with $\hat{\partial} f(x)$ being convex and $\hat{\partial} f(x) \subseteq \partial f(x)$
- if f is differentiable at x, $\hat{\partial} f(x) = \{\nabla f(x)\}$
- if f is convex, both sets coincide with the convex subdifferential
- $\partial^\circ f(x) = \operatorname{conv} \partial f(x)$.

General Subdifferential: Example

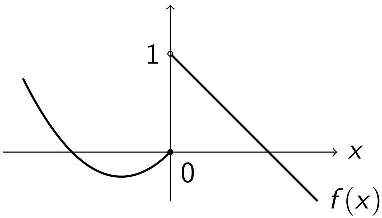
Consider the function

$$f(x) = \begin{cases} x^2 + x & \text{if } x \leq 0, \\ 1 - x & \text{if } x > 0. \end{cases}$$

General Subdifferential: Example

Consider the function

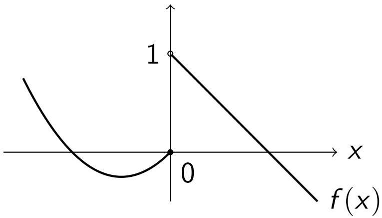
$$f(x) = \begin{cases} x^2 + x & \text{if } x \leq 0, \\ 1 - x & \text{if } x > 0. \end{cases}$$



General Subdifferential: Example

Consider the function

$$f(x) = \begin{cases} x^2 + x & \text{if } x \leq 0, \\ 1 - x & \text{if } x > 0. \end{cases}$$



We use the definition with $x^{(k)} = -1/k \to 0$ and find

$$\hat{\partial} f(0) = \partial f(0) = [1, +\infty).$$

General Subdifferential: Example II

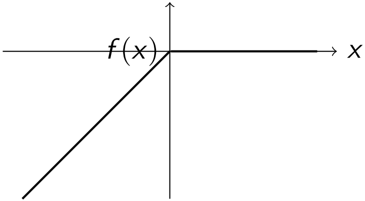
Let $f : \mathbb{R} \to \mathbb{R}$ be the function

$$f(x) = \min\{0, x\}.$$

General Subdifferential: Example II

Let $f : \mathbb{R} \to \mathbb{R}$ be the function

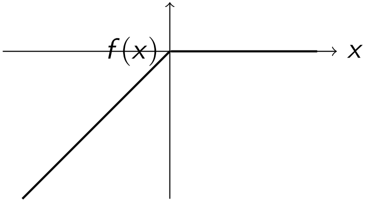
$$f(x) = \min\{0, x\}.$$



General Subdifferential: Example II

Let $f : \mathbb{R} \to \mathbb{R}$ be the function

$$f(x) = \min\{0, x\}.$$



At $x = 0$, using the definition $\hat{\partial} f(0) = \emptyset$, $\partial f(0) = \{0, 1\}$, so that $\partial^\circ f(0) = [0, 1]$.

Second-order Subderivative

Different generalizations are possible.

Second-order Subderivative

Different generalizations are possible. For $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ finite at $x \in \mathbb{R}^n$ and two directions $v, w \in \mathbb{R}^n$:

$$d^2 f(x|v)(w) = \liminf_{t \downarrow 0,\ w' \to w} \frac{f(x + tw') - f(x) - t\langle v\,,\ w'\rangle}{\frac{1}{2}t^2}.$$

Second-order Subderivative

Different generalizations are possible. For $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ finite at $x \in \mathbb{R}^n$ and two directions $v, w \in \mathbb{R}^n$:

$$d^2 f(x|v)(w) = \liminf_{t \downarrow 0,\ w' \to w} \frac{f(x + tw') - f(x) - t(v\,,\ w')}{\frac{1}{2}t^2}.$$

If f is twice differentiable at x, then for $v = \nabla f(x)$

$$d^2 f(x|v)(w) = (w\,,\ \nabla^2 f(x) w).$$

Second-order Subderivative

Different generalizations are possible. For $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ finite at $x \in \mathbb{R}^n$ and two directions $v, w \in \mathbb{R}^n$:

$$d^2 f(x|v)(w) = \liminf_{t\downarrow 0,\ w'\to w} \frac{f(x+tw') - f(x) - t\langle v\,,\, w'\rangle}{\frac{1}{2}t^2}.$$

If f is twice differentiable at x, then for $v = \nabla f(x)$

$$d^2 f(x|v)(w) = \langle w\,,\, \nabla^2 f(x) w\rangle.$$

Necessary and sufficient optimality conditions

- If x^* is a local minimizer of f, then $0 \in \partial f(x^*)$ and $d^2 f(x^*|0)(w) \geq 0$ for all $w \in \mathbb{R}^n$.

Second-order Subderivative

Different generalizations are possible. For $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ finite at $x \in \mathbb{R}^n$ and two directions $v, w \in \mathbb{R}^n$:

$$d^2 f(x|v)(w) = \liminf_{t \downarrow 0,\ w' \to w} \frac{f(x + tw') - f(x) - t\langle v\,,\ w' \rangle}{\frac{1}{2} t^2}.$$

If f is twice differentiable at x, then for $v = \nabla f(x)$

$$d^2 f(x|v)(w) = \langle w\,,\ \nabla^2 f(x) w \rangle.$$

Necessary and sufficient optimality conditions

- If x^* is a local minimizer of f, then $0 \in \partial f(x^*)$ and $d^2 f(x^*|0)(w) \geq 0$ for all $w \in \mathbb{R}^n$.
- If $0 \in \partial f(x^*)$ and $d^2 f(x^*|0)(w) > 0$ for all $w \neq 0$, then x^* is a local minimizer of f.

Newton Derivative

A function $f : \mathbb{R}^m \to \mathbb{R}^n$ is *Newton differentiable* at $x \in \mathbb{R}^m$ with Newton derivative $D_N f(x)$ if

$$\lim_{\|h\| \to 0} \frac{\|f(x+h) - f(x) - D_N f(x+h)h\|}{\|h\|} = 0$$

Newton Derivative

A function $f : \mathbb{R}^m \to \mathbb{R}^n$ is *Newton differentiable* at $x \in \mathbb{R}^m$ with Newton derivative $D_N f(x)$ if

$$\lim_{\|h\| \to 0} \frac{\|f(x+h) - f(x) - D_N f(x+h)h\|}{\|h\|} = 0$$

- If f is continuously differentiable at x, then $D_N f(x) = \nabla f(x)$

Newton Derivative

A function $f : \mathbb{R}^m \to \mathbb{R}^n$ is *Newton differentiable* at $x \in \mathbb{R}^m$ with Newton derivative $D_N f(x)$ if

$$\lim_{\|h\|\to 0} \frac{\|f(x+h) - f(x) - D_N f(x+h)h\|}{\|h\|} = 0$$

- If f is continuously differentiable at x, then $D_N f(x) = \nabla f(x)$
- The chain rule applies

Newton Derivative

A function $f : \mathbb{R}^m \to \mathbb{R}^n$ is *Newton differentiable* at $x \in \mathbb{R}^m$ with Newton derivative $D_N f(x)$ if

$$\lim_{\|h\| \to 0} \frac{\|f(x+h) - f(x) - D_N f(x+h)h\|}{\|h\|} = 0$$

- If f is continuously differentiable at x, then $D_N f(x) = \nabla f(x)$
- The chain rule applies
- If f is piecewise differentiable, then it is Newton differentiable everywhere and every $D_N f(x) \in \partial^\circ f(x)$ is a Newton derivative.

Newton Derivative: Example

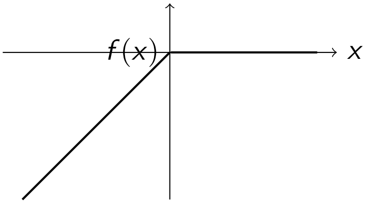
Consider again $f : \mathbb{R} \to \mathbb{R}$ given by

$$f(x) = \min\{0, x\}.$$

Newton Derivative: Example

Consider again $f : \mathbb{R} \to \mathbb{R}$ given by

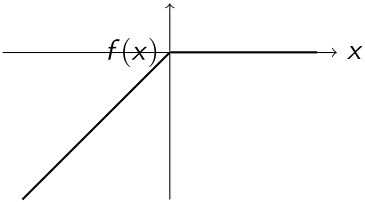
$$f(x) = \min\{0, x\}.$$



Newton Derivative: Example

Consider again $f : \mathbb{R} \to \mathbb{R}$ given by

$$f(x) = \min\{0, x\}.$$

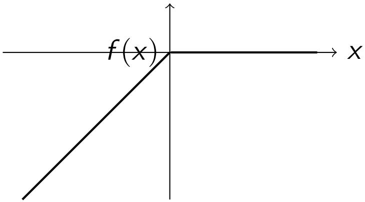


At $x = 0$, we've seen $\hat{\partial} f(0) = \emptyset$, $\partial f(0) = \{0, 1\}$, and $\partial^\circ f(0) = [0, 1]$.

Newton Derivative: Example

Consider again $f : \mathbb{R} \to \mathbb{R}$ given by

$$f(x) = \min\{0, x\}.$$



At $x = 0$, we've seen $\hat{\partial} f(0) = \emptyset$, $\partial f(0) = \{0, 1\}$, and $\partial^\circ f(0) = [0, 1]$. Then its Newton derivative is, for $\delta \in [0, 1]$

$$D_N f(x)h = \begin{cases} h & \text{if } x < 0, \\ \delta h & \text{if } x = 0, \\ 0 & \text{if } x > 0. \end{cases} .$$

Semismooth Newton Method

Goal: Find a zero of a function $f : \mathbb{R}^m \to \mathbb{R}^n$

Semismooth Newton Method

Goal: Find a zero of a function $f : \mathbb{R}^m \to \mathbb{R}^n$

Data: $f, D_N f, x_0 \in \mathbb{R}^n, k = 0.$

```
while convergence criterion is not fulfilled do
    Solve D_N f(x^(k)) g^(k) = -f(x^(k)) for g^(k)
    Set x^(k+1) = x^(k) + g^(k)
    Set k := k + 1
end
```

while *convergence criterion is not fulfilled* **do**

Solve $D_N f(x^{(k)}) g^{(k)} = -f(x^{(k)})$ for $g^{(k)}$

Set $x^{(k+1)} = x^{(k)} + g^{(k)}$

Set $k := k + 1$

end

Result: $x^{(k_*)}$

Semismooth Newton Method

Goal: Find a zero of a function $f : \mathbb{R}^m \to \mathbb{R}^n$

Data: f, $D_N f$, $x_0 \in \mathbb{R}^n$, $k = 0$.

while *convergence criterion is not fulfilled* **do**

  Solve $D_N f(x^{(k)}) g^{(k)} = -f(x^{(k)})$ for $g^{(k)}$

  Set $x^{(k+1)} = x^{(k)} + g^{(k)}$

  Set $k := k + 1$

end

Result: $x^{(k_*)}$

- Local superlinear convergence, i. e. $\lim_{k \to +\infty} \frac{x^{(k+1)} - x^*}{x^{(k)} - x^*} = 0$, to a x^* such that $f(x^*) = 0$.

Summary

- **Convex subdifferential**
- **Generalized gradient for locally Lipschitz functions**
- **Subgradient-based algorithm**
- **Second-order generalized derivatives and optimality**
- **Semismooth Newton algorithm**

See you later for more on manifolds!