Nonsmooth Analysis and Applications

Francis Clarke
Institut universitaire de France et Université de Lyon

Classical Calculus
A basic technique in mathematics is linearization

Classical Calculus
A basic technique in mathematics is linearization

Classical Calculus
A basic technique in mathematics is linearization

Classical Calculus
A basic technique in mathematics is linearization
Classical Calculus

A basic technique in mathematics is linearization

$$f' \alpha = \text{the slope of the tangent line to the graph of } f \text{ through the point } (\alpha, f(\alpha))$$

A. Minimizing a function $f(x)$
A. Minimizing a function $f(x)$

Fermat's rule: at a minimum $\alpha$, we have $f'(\alpha) = 0$

B. Solving an equation $f(x) = 0$
B. Solving an equation \( f(x) = 0 \)

Newton's Method:

\[ x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \]
C. Studying a system

\[ x'(t) = f(x(t), y(t)) \]
\[ y'(t) = g(x(t), y(t)) \]

around an equilibrium \((0,0)\)

Calculate

\[ A := \begin{bmatrix} f_x(0,0) & f_y(0,0) \\ g_x(0,0) & g_y(0,0) \end{bmatrix} \]

Then study

\[ \begin{bmatrix} x' \\ y' \end{bmatrix} = A \begin{bmatrix} x \\ y \end{bmatrix} \]

via eigenvalues...

Thus, in the space of almost precisely one century, infinitesimal calculus, or as we now call it in English, The Calculus, the calculating tool \textit{par excellence}, had been forged; and nearly three centuries of constant use have not dulled this incomparable instrument. \hfill \textit{Bourbaki}

\textit{But what if }f\textit{ is not differentiable?}
A. Minimizing a function $f(x)$

B. Solving an equation $f(x) = 0$

A. Minimizing a function $f(x)$

B. Solving an equation $f(x) = 0$

Natural context: lower semicontinuous functions

Natural context: continuous functions
C. Studying a system
\[ x'(t) = f(x(t), y(t)) \]
\[ y'(t) = g(x(t), y(t)) \]
around an equilibrium \((0, 0)\)

Natural context: locally Lipschitz functions

---

Je me détourne avec effroi et horreur de cette plaie lamentable des fonctions qui n'ont pas de dérivées.

Hermite

---

If Newton had thought that continuous functions do not necessarily have derivatives—and this is the general case—the differential calculus would never have been created.

Emil Picard
Je me détourne avec effroi et horreur de cette plaie lamentable des fonctions qui n'ont pas de dérivées.  

Hermite

If Newton had thought that continuous functions do not necessarily have derivatives—and this is the general case—the differential calculus would never have been created.  

Emil Picard

Nonsmooth Analysis began with "Dini Derivates":

Fondamenti per la teorica delle funzioni di variabili reali  
Ulyssse Dini 1878

The swing in your backyard
The nonlinear pendulum

The swing in your backyard
The nonlinear pendulum

\[ \ell \]

\[ m \]

\[ \theta \]

\[ \theta(t) \]
The swing in your backyard
The nonlinear pendulum

\[ m \ell \theta'' = -m g \sin \theta \implies \theta'' + \left(\frac{g}{\ell}\right) \sin \theta = 0 \]

**Newton(-Euler)**

Equilibrium \( \theta = 0 \)

\[ \theta'' + \left(\frac{g}{\ell}\right) \theta = 0 \]

What if there's a wind?
The swing in your backyard

The nonlinear pendulum

Newton(-Euler)

\[ m \ell \theta'' = -mg \sin \theta \implies \theta'' + \frac{(g/\ell)}{\theta} \sin \theta = 0 \]

equilibrium \( \theta = 0 \)

\[ \theta'' + \frac{(g/\ell)}{\theta} \theta = 0 \]

What if there’s a wind?

\[ m \ell \theta'' = -mg \sin \theta - f \cos \theta \]

---

The swing in your backyard

The nonlinear pendulum

Newton(-Euler)

\[ m \ell \theta'' = -mg \sin \theta \implies \theta'' + \frac{(g/\ell)}{\theta} \sin \theta = 0 \]

equilibrium \( \theta = 0 \)

\[ \theta'' + \frac{(g/\ell)}{\theta} \theta = 0 \]

What if there’s a wind?

\[ m \ell \theta'' = -mg \sin \theta - f \cos \theta \]

---

force \( f \)
Let the mass be a thin seat of width $w$

Reality is more complicated...

Let the mass be a thin seat of width $w$

$$f = cw|\sin \theta|$$

Reality is more complicated...

$$m \ddot{\theta}'' = -mg \sin \theta - f \cos \theta$$

$$\Rightarrow m \ddot{\theta}'' = -mg \sin \theta - cw|\sin \theta| \cos \theta$$
Let the mass be a thin seat of width $w$

$$w|\sin \theta|$$

$$f = cw|\sin \theta|$$

Reality is more complicated...

$$m \ell \dot{\theta}'' = -mg \sin \theta - f \cos \theta$$

$$\Rightarrow m \ell \dot{\theta}'' = -mg \sin \theta - cw|\sin \theta| \cos \theta$$

$0$ is now one of two equilibria; no linearization

---

Let the mass be a thin seat of width $w$

$$w|\sin \theta|$$

$$f = cw|\sin \theta|$$

Reality is more complicated...

$$m \ell \dot{\theta}'' = -mg \sin \theta - f \cos \theta$$

$$\Rightarrow m \ell \dot{\theta}'' = -mg \sin \theta - cw|\sin \theta| \cos \theta$$

$0$ is now one of two equilibria; no linearization
Directional or threshold phenomena are often nonsmooth. Also where there is presence of shapes... another example:

nonsmooth contact

\[
\frac{d}{dt} x(t) = \begin{cases} 
\alpha(u(t) - x(t)) & \text{if } x(t) \leq u(t) \\
-\beta(x(t) - u(t)) & \text{if } x(t) \geq u(t) 
\end{cases}
\]

f(x, u) has a corner at x = u.
Directional or threshold phenomena are often nonsmooth. Also where there is presence of shapes... another example:

nonsmooth contact

See Marsden et alii (generalized gradients + least action). Other nonsmooth mechanics and elasticity: Brogliato, Moreau, Panagiotopoulos, Paoli, Schatzman, Schuricht ...
Optimization
Example: eigenvalue design

Let $A(x) = [a_{ij}(x)]$ be an $n \times n$ symmetric matrix whose coefficients depend smoothly upon a parameter $x$.

A function of interest:
\[ f(x) := \text{the greatest eigenvalue of } A(x). \]

FACT: $f$ is nonsmooth in general

\[ A(x) = \begin{bmatrix} 1 & x \\ x & 1 \end{bmatrix} \]

\[ \lambda = 1 \pm |x| \implies f(x) = 1 + |x| \]

Note that $f$ attains its min at a “corner”
Example: the distance function
\[ d_S(x) := \min_{s \in S} |x - s| \]
\( S \subset \mathbb{R}^n \) (closed, nonempty)
Example: the distance function \( d_S(x) := \min_{s \in S} |x - s| \) (closed, nonempty) \( S \subseteq \mathbb{R}^n \)

One use: exact penalization
\[
\min_{x \in S} g(x) \iff \min_{x} g(x) + k d_S(x)
\]

Then:
\[
0 \in \partial \{g + k d_S\}(x)
\]
\[
\Rightarrow 0 \in \partial g(x) + k \partial d_S(x)
\]
Example: Problem of Torricelli/Steiner, $n = 4$

We seek a central point $x$ relative to $h_1, h_2, h_3, h_4$: $x$ minimizes

$$|h_1 - x| + |h_2 - x| + |h_3 - x| + |h_4 - x|$$

A table can solve this problem...
Example: Problem of Torricelli/Steiner, n = 4

We seek a central point $x$ relative to $h_1, h_2, h_3, h_4$: $x$ minimizes $|h_1 - x| + |h_2 - x| + |h_3 - x| + |h_4 - x|$

A table can solve this problem...

At equilibrium, the point minimizes the potential energy of the system (d'Alembert)

Analytically, $x$ solves the problem iff

$$0 \in \sum_{i=1}^{4} \partial|x - h_i|$$

where $\partial|x - h_i| = \begin{cases} 
B & \text{if } x = h_i \\
\frac{x - h_i}{|x - h_i|} & \text{if } x \neq h_i 
\end{cases}$
Analytically, \( x \) solves the problem iff
\[
0 \in \sum_{i=1}^{4} \partial |x - h_i| \quad \text{where} \quad \partial |x - h_i| = \begin{cases} B & \text{if } x = h_i \\ \frac{x-h_i}{|x-h_i|} & \text{if } x \neq h_i \end{cases}
\]

Let us add one string going over the edge: Then
\[
\min x \sum_{i=1}^{4} |x - h_i| + cd_s(x), \quad S = \text{complement of the table}
\]

We have \( x''(t) = g(x(t)) \) with \( g = \nabla V \) discontinuous. And also local minima:
Analytically, $x$ solves the problem iff
\[ 0 \in \sum_{i=1}^{4} \vartheta|x-h_i| \text{ where } \vartheta|x-h_i| = \begin{cases} B & \text{if } x = h_i \\ \frac{x-h_i}{|x-h_i|} & \text{if } x \neq h_i \end{cases} \]

Let us add one string going over the edge: Then
\[ \min_x \sum_{i=1}^{4} |x-h_i| + cd_s(x), \quad S = \text{complement of the table} \]

We have $x''(t) = g(x(t))$ with $g = \nabla V$ discontinuous.
And also local minima:

```
  x
  
  e e e
```
Analytically, $x$ solves the problem iff

$$0 \in \sum_{i=1}^{4} \partial|x - h_i| \text{ where } \partial|x - h_i| = \begin{cases} B & \text{if } x = h_i \\ \frac{x - h_i}{|x - h_i|} & \text{if } x \neq h_i \end{cases}$$

Let us add one string going over the edge: Then

$$\min_{x} \sum_{i=1}^{4} |x - h_i| + cd_{0}(x), \quad S = \text{complement of the table}$$

We have $x''(t) = g(x(t))$ with $g = \nabla V$ discontinuous.
And also local minima:

We have $x''(t) = g(x(t))$ with $g = \nabla V$ discontinuous.
And also local minima:
Calculus of variations

The Basic Problem:

\[ \min_{x(\cdot)} \int_{a}^{b} L(t, x(t), x'(t)) \, dt, \quad x(a) = A, x(b) = B \]

Euler (1744) defined the problem, found the basic necessary condition, introduced multipliers for constrained problems, postulated the principle of least action, and gave 100 examples.

Example: soap bubble

Leonhard Euler

1707-1783
Example: soap bubble

Example: soap bubble

Example: soap bubble

Example: soap bubble

Surface area

Euler-Lagrange equation

Solutions with corners
Solutions with corners

\[ x(t) \]

\[ \rightarrow \]

\[ \rightarrow \]

Solutions with corners

\[ x(t) \]

\[ \rightarrow \]

\[ \rightarrow \]

Solutions with corners

\[ x(t) \]

\[ \rightarrow \]

\[ \rightarrow \]

Solutions with corners

\[ x(t) \]

\[ \rightarrow \]

\[ \rightarrow \]
Solutions with corners

A design problem

The Goldschmidt solution (1831)
A design problem
Ainsi c'est un problème de maximis et minimis de déterminer la courbe qui, par sa rotation autour de son axe formera une colonne capable de supporter la plus grande charge possible, la hauteur et la masse de la colonne étant données.

Lagrange (1770) Sur la figure des colonnes

To find the curve which by its revolution determines the column of greatest efficiency

Truesdell

Joseph Louis Lagrange
Born Turin 1736
- Writes to Euler in 1755, describes the method of variations
- Euler names the subject in his honor: calculus of variations
- Euler is his mentor until his death
• After 20 years in Berlin, he joins the Paris Academy in 1786
• During the revolution: metric system, Ecole Normale and Polytechnique
• Under Napoléon: senator, count of the Empire, grand officer of the Légion d'honneur
• After 20 years in Berlin, he joins the Paris Academy in 1786
• During the revolution: metric system, Ecole Normale and Polytechnique
• Under Napoléon: senator, count of the Empire, grand officer of the Légion d'honneur
• His ‘greatest treasure’: his (very) young wife, whom he marries at the age of 56
• Dies in Paris in 1813 at the age of 77

Designing an optimal column

1. Choose a profile $x$
Designing an optimal column

1. Choose a profile $x$
2. Rotate to generate a column $C(x)$
3. Respect the constraints on the volume, the height
4. Calculate (via Euler) the buckling strength $f(x)$ of the column $C(x)$
5. Maximize $f(x)$ over $x$
Three solutions

Lagrange 1770
Keller & Tadjbaksh 1960

67%
3%

Three solutions

Lagrange 1770
Keller & Tadjbaksh 1960
Cox & Overton 1992 (generalized gradients)

100%

Envelopes of smooth functions are nonsmooth
Envelopes of smooth functions are nonsmooth

\[ f \xrightarrow{\text{ }} x \]

\[ f \xrightarrow{\text{ }} g \]

Envelopes of smooth functions are nonsmooth

\[ f \xrightarrow{\text{ }} g \xrightarrow{\text{ }} x \]

the function \( \min(f,g) \) has a corner here
Envelopes of smooth functions are nonsmooth

\[ f \xrightarrow{x} g \]

the function \( \min(f,g) \) has a corner here

The function "maximal load supported by a column of profile \( x \)" is a nonsmooth function of \( x \) ... which is where the error was made

Generalized gradients and proximal normals

\[ f^0(x;v) = \limsup_{t \downarrow 0, y \to x} \frac{f(y + tv) - f(y)}{t} \]

Generalized gradients and proximal normals

Four definitions
Clarke 1973
Generalized gradients and proximal normals

\[ f^0(x; v) = \limsup_{t \downarrow 0, y \to x} \frac{f(y + tv) - f(y)}{t} \]

Four definitions
Clarke 1973

\[ \partial_C f(x) = \left\{ \zeta \in X^* : f^0(x; v) \geq \langle \zeta, v \rangle \forall v \right\} \]

\[ \zeta \in N^p_S(x) \iff \exists \sigma \geq 0 \text{ s.t. } \langle \zeta, x' - x \rangle \leq \sigma |x' - x|^2 \forall x' \in S \]

\[ \zeta \in N^p_S(x) \iff \exists \sigma \geq 0 \text{ s.t. } \langle \zeta, x' - x \rangle \leq \sigma |x' - x|^2 \forall x' \in S \]
Generalized gradients and proximal normals

\[ f^\circ(x; v) = \limsup_{t \downarrow 0, y \rightarrow x} \frac{f(y + tv) - f(y)}{t} \]

Four definitions
Clarke 1973

\[ \partial_C f(x) = \{ \zeta \in X^* : f^\circ(x; v) \geq \langle \zeta, v \rangle \ \forall v \} \]

\[ \zeta \in N_S^\circ(x) \iff \exists \sigma \geq 0 \text{ s.t.} \quad \langle \zeta, x' - x \rangle \leq \sigma |x' - x|^2 \ \forall x' \in S \]

Then

- 
- 
- 
-
Then
- $\partial_C f(x)$ is compact convex nonempty

Then
- $\partial_C f(x)$ is compact convex nonempty
- $\partial_C (-f)(x) = -\partial_C f(x)$
- $\partial_C (f + g)(x) \subset \partial_C f(x) + \partial_C g(x)$
- $\partial_C \max_{1 \leq i \leq n} f_i(x) \subset \ldots$
Then

- $\partial_C f(x)$ is compact convex nonempty
- $\partial_C (-f)(x) = -\partial_C f(x)$
- $\partial_C (f + g)(x) \subset \partial_C f(x) + \partial_C g(x)$
- $\partial_C \max_{1 \leq i \leq n} f_i(x) \subset \ldots$
- Mean value theorem
- Tangent vectors and normals to closed sets

These generalized gradients (1972) apply on any Banach space.

The classical derivative corresponds to a two-sided local approximation by an affine function.
The classical derivative corresponds to a two-sided local approximation by an affine function.

\[ f(\cdot) \]

\[ f'(\alpha) = \text{slope} \]
The classical derivative corresponds to a two-sided local approximation by an affine function.

\[ f'(\alpha) = \text{slope} \]

\[ f(\cdot) \]

\[ \alpha \]

We may also approximate just from below, using nonlinear functions: proximal analysis.
We can apply the 'local lower-approximation by parabolas' idea to nonsmooth (lsc) functions.

The set of all 'contact slopes' of lower locally supporting parabolas is the proximal subdifferential $\partial_p f(\alpha)$.
The set of all ‘contact slopes’ of lower locally supporting parabolas is the proximal subdifferential $\partial_p f(\alpha)$.

$\partial_p f(\alpha) = [-2, 1]$
The set of all `contact slopes` of lower locally supporting parabolas is the proximal subdifferential $\partial_p f(\alpha)$

$\partial_p f(\alpha) = [-2, 1]$  

$\partial_p f(\alpha) = \emptyset$

$\zeta \in \partial_p f(\alpha) \iff f(x) \geq \langle \zeta, x - \alpha \rangle + f(\alpha) - \sigma |x - \alpha|^2$ locally

$\partial_p f$ has a very complete (but fuzzy!) theory and calculus... Borwein, Ioffe, Ledyaev, Loewen, Rockafellar, Vinter, Zeidan...
The Hamilton–Jacobi equation: Various solution concepts
\[ \phi_t(t, x) + H(t, x, \phi_x(t, x)) = 0 \quad \text{(and bdry cdns)} \]

- Classical
  (\( \psi \) smooth, pointwise equality)

- Almost everywhere solutions
  (\( \psi \) Lipschitz)
The Hamilton–Jacobi equation: Various solution concepts

\( \phi_t(t, x) + H(t, x, \phi_u(t, x)) = 0 \) (and bdry cdns)

- Classical
  \( (\psi \text{ smooth, pointwise equality}) \)
- Almost everywhere solutions
  \( (\psi \text{ Lipschitz}) \)
- Using generalized gradients
  \( (\text{Clarke 1977}) \)
- Using Dini derivates bilaterally
  \( (\text{Subbotin 1980}) \)
- Using sub- and superdifferentials [viscosity] \( (\text{Fleming, Crandall & Lions 1982, Evans}...) \)

The Hamilton–Jacobi equation: Various solution concepts

\( \phi_t(t, x) + H(t, x, \phi_u(t, x)) = 0 \) (and bdry cdns)

- Classical
  \( (\psi \text{ smooth, pointwise equality}) \)
- Almost everywhere solutions
  \( (\psi \text{ Lipschitz}) \)
- Using generalized gradients
  \( (\text{Clarke 1977}) \)
- Using Dini derivates bilaterally
  \( (\text{Subbotin 1980}) \)
- Using sub- and superdifferentials [viscosity] \( (\text{Fleming, Crandall & Lions 1982, Evans}...) \)
- Unilateral/proximal/KAM...
The Hamilton–Jacobi equation: Various solution concepts
\[ \phi_t(t, x) + H(t, x, \phi_x(t, x)) = 0 \] (and bdry cdns)

- Classical (Ψ smooth, pointwise equality)
- Almost everywhere solutions (Ψ Lipschitz)
- Using generalized gradients (Clarke 1977)
- Using Dini derivatives bilaterally (Subbotin 1980)
- Using sub- and superdifferentials [viscosity] (Fleming, Crandall & Lions 1982, Evans...)
- Unilateral/proximal/KAM...

Example (n = 1)
\[ [\varphi'(x)]^2 - 1 = 0, \quad \varphi(0) = \varphi(1) = 0 \]
- No smooth solutions

For linear pde's one can circumvent nonsmoothness by distributions... but in the nonlinear case, a careful analysis of the points of nondifferentiability is required.

Example (n = 1)
\[ [\varphi'(x)]^2 - 1 = 0, \quad \varphi(0) = \varphi(1) = 0 \]
- Many "almost everywhere" solutions:
Example (n = 1)
- No smooth solutions
- Many "almost everywhere" solutions:

\[ [\phi'(x)]^2 - 1 = 0, \quad \phi(0) = \phi(1) = 0 \]

A unique continuous \( \phi \) satisfies

\[ [\partial x \phi(x)]^2 - 1 = 0 \]

(i.e. \( [x^2 - 1 = 0 \forall x \in \phi(x), x \in (0,1)] \))

Hint: it is one of these two functions:

\[ \phi_1, \phi_2 \]

Optimal control: an example in bioeconomics

\[ x'(t) = G(x(t)) - u(t) x(t) \]

Maximize

\[ \int_0^\infty e^{-\delta t} \{ \pi x(t) - k \} u(t) dt \]

Subject to

\[ 0 \leq u(t) \leq \bar{u} \]

\( x = \text{biomass} \)
\( u = \text{fishing effort} \)
\( G = \text{natural growth} \)
\( \bar{u} = \text{maximum} \)
\( \delta = \text{fishing effort} \)
\( \pi = \text{discount rate} \)
\( k = \text{effort cost} \)
Optimal control: an example in bioeconomics

\[ x'(t) = G(x(t)) - u(t) x(t) \]
\[ \text{MIN} \int_0^\infty e^{-\delta t} \{x - c(t) - k\} u(t) \, dt \]
\[ 0 \leq u(t) \leq E \]

\( x \) = biomass
\( u \) = fishing effort
\( G \) = natural growth
\( E \) = maximum fishing effort
\( \delta \) = discount rate
\( c \) = resource price
\( k \) = effort cost

If \( \delta \) is sufficiently large, we have \( x_c = 0 \) (extinction)
Example: Optimal fishing strategy in the presence of both investment and depreciation in boats
(Clark, Clarke, Munro / Econometrica)

\[
\max \int_0^\infty e^{-\delta t} \left\{ (\pi x(t) - k)u(t) - cI(t) \right\} \, dt + \sum_i e^{-\delta t_i} \Delta E(t_i)
\]

\[
x'(t) = g(x(t)) - u(t)x(t), \quad 0 \leq u(t) \leq E(t)
\]

\[
E'(t) = -\gamma E(t) + I(t), \quad 0 \leq I(t) \leq +\infty
\]
Proof: non-smooth verification functions

There is an optimal feedback $u(y,E)$... discontinuous

Verification functions

\[
\min J(x, u) := \int_a^b L(t, x(t), u(t)) \, dt \quad x(a) = A \\
x'(t) = f(t, x(t), u(t)), \ u(t) \in U(t) \text{ a.e.} \\
x(b) = B
\]

Goal: verify that a candidate $(x_*, u_*)$ is optimal
\[ \min J(x, u) := \int_a^b L(t, x(t), u(t)) \, dt \quad x(a) = A \]
\[ x'(t) = f(t, x(t), u(t)), \ u(t) \in U(t) \text{ a.e.} \quad x(b) = B \]

Goal: verify that a candidate \((x_*, u_*)\) is optimal
Method: exhibit a function \(\phi\) satisfying
\[ L(t, x, u) \geq \phi(t, x) + \langle \phi_x(t, x), f(t, x, u) \rangle \quad \forall (t, x), u \in U(t) \]
\[ (= \text{at } (t, x_*(t), u_*(t))) \]

Proof: For any admissible \((x, u)\) we have
\[ L(t, x(t), u(t)) \geq \phi(t, x(t)) + \langle \phi_x(t, x(t)), f(t, x(t), u(t)) \rangle \]
\[ = d/dt \{ \phi(t, x(t)) \} \]
\[ \Rightarrow J(x, u) \geq \phi(b, B) - \phi(a, A) \]

\[ \min J(x, u) := \int_a^b L(t, x(t), u(t)) \, dt \quad x(a) = A \]
\[ x'(t) = f(t, x(t), u(t)), \ u(t) \in U(t) \text{ a.e.} \quad x(b) = B \]

Goal: verify that a candidate \((x_*, u_*)\) is optimal
Method: exhibit a function \(\phi\) satisfying
\[ L(t, x, u) \geq \phi(t, x) + \langle \phi_x(t, x), f(t, x, u) \rangle \quad \forall (t, x), u \in U(t) \]
\[ (= \text{at } (t, x_*(t), u_*(t))) \]

Proof: For any admissible \((x, u)\) we have
\[ L(t, x(t), u(t)) \geq \phi(t, x(t)) + \langle \phi_x(t, x(t)), f(t, x(t), u(t)) \rangle \]
\[ = d/dt \{ \phi(t, x(t)) \} \]
\[ \Rightarrow J(x, u) \geq \phi(b, B) - \phi(a, A) \]

\[ (= \text{for } (x, u) = (x_*, u_*)) \] QED
\[
\min J(x, u) := \int_a^b L(t, x(t), u(t)) \, dt \quad x(a) = A
\]
\[
x'(t) = f(t, x(t), u(t)), \ u(t) \in U(t) \text{ a.e.} \quad x(b) = B
\]

Goal: verify that a candidate \((x_*, u_*)\) is optimal
Method: exhibit a function \(\phi\) satisfying
\[
L(t, x, u) \geq \phi(t, x(t)) + \langle \phi_x(t, x(t)), f(t, x(t), u(t)) \rangle \quad \forall (t, x, u) \in U(t)
\]
\(= \text{ at } (t, x_*(t), u_*(t)))

Proof: For any admissible \((x, u)\) we have
\[
L(t, x(t), u(t)) \geq \phi(t, x(t)) + \langle \phi_x(t, x(t)), f(t, x(t), u(t)) \rangle
\]
\[
= d/dt \{ \phi(t, x(t)) \}
\]
\[
\implies J(x, u) \geq \phi(b, B) - \phi(a, A)
\]
\(= \text{ for } (x, u) = (x_*, u_*)) \quad \text{QED}

Fact: smooth verification functions may not exist, but nonsmooth ones do (Clarke & Vinter, 1980's)

---

A reference

Nonsmooth Analysis and Control Theory
by
F. Clarke, Yu. Ledyaev, R. Stern, P. Wolenski
Graduate Texts in Mathematics
Springer-Verlag 1998

There are two kinds of mathematics books: the kind you can't read past the first sentence, and the kind you can't read past the first page.

Richard Feynman
clarke@math.univ-lyon1.fr

---

A reference

Generalized Gradients and Proximal analysis

Francis Clarke
Institut universitaire de France et Université de Lyon

clarke@math.univ-lyon1.fr
Yesterday, we motivated the need for nonsmooth analysis. It appears that nonsmoothness is more common than one might have thought, and that the opposite of "linear" is often "nonsmooth".

Today, we examine the basic constructs and some elements of the calculus. We stress that difficult nonsmooth problems remain difficult even if one has mastered this theory! (But it can help...)

**Generalized gradients and associated geometry**

In an arbitrary Banach space, the starting point for functions is the generalized directional derivative:

$$f^o(x; v) = \limsup_{t \to 0, y \to x} \frac{f(y + tv) - f(y)}{t} \quad x, v \in X$$

**Generalized gradients and associated geometry**

In an arbitrary Banach space, the starting point for functions is the generalized directional derivative:

$$f^o(x; v) = \limsup_{t \to 0, y \to x} \frac{f(y + tv) - f(y)}{t} \quad x, v \in X$$

When $f$ is locally Lipschitz, this is finite, and we find:

$$f^o(x; v + w) \leq f^o(x; v) + f^o(x; w) \quad \forall v, w$$

$$f^o(x; tv) = tf^o(x; v) \quad \forall t \geq 0$$
Generalized gradients and associated geometry

In an arbitrary Banach space, the starting point for functions is the generalized directional derivative:

\[ f^0(x; v) = \limsup_{t \downarrow 0, y \to x} \frac{f(y + tv) - f(y)}{t} \quad x, v \in X \]

When \( f \) is locally Lipschitz, this is finite, and we find:

\[ f^0(x; v + w) \leq f^0(x; v) + f^0(x; w) \quad \forall v, w \]
\[ f^0(x; tv) = tf^0(x; v) \quad \forall t \geq 0 \]

These are properties of support functions.

If \( Z \) is a nonempty bounded set in \( X^* \), then the support function has these properties:

\[ H_Z(v) := \sup_{\zeta \in Z} \langle \zeta, v \rangle \]

If \( Z \) is a nonempty bounded set in \( X^* \), then the support function has these properties:

\[ H_Z(v) := \sup_{\zeta \in Z} \langle \zeta, v \rangle \]

When restricted to \( w^* \)-closed convex sets, the support function characterizes \( Z \). The Hahn-Banach theorem implies the existence of a unique \( w^* \)-closed, convex, bounded set \( Z \) such that

\[ f^0(x; v) = H_Z(v) \quad \forall v \in X \]

We denote this set by \( \partial_{C} f(x) \), the generalized gradient. The following duality holds:

\[ \partial_{C} f(x) = \{ \zeta \in X^* : f^0(x; v) \geq \langle \zeta, v \rangle \quad \forall v \} \]
\[ f^0(x; v) = \max_{\zeta \in \partial_{C} f(x)} \langle \zeta, v \rangle \]
\( \partial_C f(x) \) is convex, compact, and closed, which may explain the subscript \( C \).

It is often referred to as the Clarke generalized gradient.

Other constructs will include:
\( \partial_P f(x) \) (proximal subdifferential) and
\( \partial_L f(x) \) (limiting subdifferential)

Let \( S \) be a nonempty closed subset of \( X \).
Its **distance function** (Lipschitz) is given by

\[
d_S(x) := \inf_{y \in S} \| x - y \|
\]

We define the generalized **normal and tangent cones** by

\[
N_S^C(x) := \text{cl} \{ t \partial_C d_S(x) : t \geq 0 \}
\]

\[
T_S^C(x) = [N_S^C(x)]^\circ
\]

\[
= \{ v : \langle \zeta, v \rangle \leq 0 \; \forall \zeta \in N_S^C(x) \}
\]

\[
= \{ v : d_S^0(x; v) = 0 \}
\]

If we wish to start with geometry, the last shall be first:

\[
T_S^C(x) := \{ v : \forall \, x_i \to_S x, \; \forall \, t_i \downarrow 0, \exists \, v_i \to v / x_i + t_i v_i \in S \}
\]
If we wish to start with geometry, the last shall be first:

$$T^C_S(x) := \{v : \forall x_i \rightarrow_S x, \forall t_i \downarrow 0, \exists v_i \rightarrow v / x_i + t_iv_i \in S\}$$

$$N^C_S(x) := [T^C_S(x)]^\circ$$

$$= \{\zeta : \langle \zeta, v \rangle \leq 0 \forall v \in T^C_S(x)\}$$

How do we recover the functional constructs?

---

*Classical Calculus*

$f'$ is the slope of the tangent line to the graph of $f$ through the point $(\alpha, f(\alpha))$.

$f'(\alpha)$ is the slope of the tangent line to the graph of $f$ through the point $(\alpha, f(\alpha))$.

Dually, the value $\zeta$ such that $(\zeta, -1)$ is normal to the graph of $f$. 
If we wish to start with geometry, the last shall be first:

\[ T^C_S(x) := \{ v : \forall x_i \rightarrow_S x, \forall t_i \downarrow 0, \exists v_i \rightarrow v / x_i + t_i v_i \in S \} \]
\[ N^C_S(x) := [T^C_S(x)]^\circ \]
\[ = \{ \xi : \langle \xi, v \rangle \leq 0 \ \forall v \in T^C_S(x) \} \]

How do we recover the functional constructs?

If we wish to start with geometry, the last shall be first:

\[ T^C_S(x) := \{ v : \forall x_i \rightarrow_S x, \forall t_i \downarrow 0, \exists v_i \rightarrow v / x_i + t_i v_i \in S \} \]
\[ N^C_S(x) := [T^C_S(x)]^\circ \]
\[ = \{ \xi : \langle \xi, v \rangle \leq 0 \ \forall v \in T^C_S(x) \} \]

How do we recover the functional constructs?

\[ \partial_C f(x) := \{ \xi : (\xi, -1) \in N^C_{\text{ opi } f}(x, f(x)) \} \]

(and then \( f^\circ(x; \cdot) \) is the support function of \( \partial_C f(x) \))
The smooth case

If \( f \) is smooth, then \( \partial_C f(x) = \{ f'(x) \} \), since
\[
\langle f'(x), v \rangle = f'(x; v) = f^c(x; v) = \max_{\zeta \in \partial_C f(x)} \langle \zeta, v \rangle
\]

If \( S \) is a smooth manifold, or manifold with boundary:

\[
N^C_S = \text{a ray}
\]
The smooth case

If $f$ is smooth, then $\partial_C f(x) = \{ f'(x) \}$, since

$$\langle f'(x), v \rangle = f'(x; v) = f^c(x; v) = \max_{\zeta \in \partial_C f(x)} \langle \zeta, v \rangle$$

If $S$ is a smooth manifold, or manifold with boundary:

$N^C_S = \text{a ray}$

$T^C_S = \text{a halfspace}$

The convex case

If $f$ is convex, then

$$\partial_C f(x) = \partial f(x)$$

the subdifferential

$$= \{ \zeta : f(y) - f(x) \geq \langle \zeta, y - x \rangle \ \forall \ y \in X \}$$

If $S$ is convex, then

$S$
The convex case

If $f$ is convex, then

$$\partial_C f(x) = \partial f(x)$$

the subdifferential

$$= \{ \zeta : f(y) - f(x) \geq \langle \zeta, y - x \rangle \ \forall \ y \in X \}$$

If $S$ is convex, then

If $S$ is convex, then

$$\zeta \in N_S(x) \iff \langle \zeta, y - x \rangle \leq 0 \ \forall \ y \in S$$

The convex case

If $f$ is convex, then

$$\partial_C f(x) = \partial f(x)$$

the subdifferential

$$= \{ \zeta : f(y) - f(x) \geq \langle \zeta, y - x \rangle \ \forall \ y \in X \}$$

If $S$ is convex, then

$$\zeta \in N_S(x) \iff \langle \zeta, y - x \rangle \leq 0 \ \forall \ y \in S$$

An example which is neither smooth nor convex

If $f$ is convex, then

$$\partial_C f(x) = \partial f(x)$$

the subdifferential

$$= \{ \zeta : f(y) - f(x) \geq \langle \zeta, y - x \rangle \ \forall \ y \in X \}$$

If $S$ is convex, then

$$\zeta \in N_S(x) \iff \langle \zeta, y - x \rangle \leq 0 \ \forall \ y \in S$$
An example which is neither smooth nor convex

S

An example which is neither smooth nor convex

S

An example which is neither smooth nor convex

S

An example which is neither smooth nor convex

S
An example which is neither smooth nor convex

S

An example which is neither smooth nor convex

S
An example which is neither smooth nor convex

\[ S \]

\[ N_S^C \]

An example which is neither smooth nor convex

\[ S \]

\[ N_S^C \]

An example which is neither smooth nor convex

\[ S \]

\[ N_S^C \]

An example which is neither smooth nor convex

\[ S \]

\[ N_S^C \]

\[ T_S^C \]

Bouligand contingent cone

\[ T_S(x) := \left\{ \lim_{i \to \infty} \frac{x_i - x}{t_i} : x_i \to x, t_i \downarrow 0 \right\} \]
Sums:
\[ \partial_C (f_1 + f_2)(x) \subseteq \partial_C f_1(x) + \partial_C f_2(x) \]
(equality when \( f_1, f_2 \) regular)
Some calculus

Sums:
\[ \partial_C (f_1 + f_2)(x) \subseteq \partial_C f_1(x) + \partial_C f_2(x) \]
(equality when \( f_1, f_2 \) regular)

Mean value theorem:
\[ \exists z \in (x,y) / f(y) - f(x) \in (\zeta, y - x) \]

Maximum functions:
\[ f(x) = \max_{1 \leq i \leq n} f_i(x) \] (each \( f_i \) smooth)
\[ I(x) = \{ i \in \{1, 2, \ldots, n\} : f_i(x) = f(x) \} \]
Then \( \partial_C f(x) = \text{co} \{ f'_i(x) : i \in I(x) \} \)

Some calculus

Sums:
\[ \partial_C (f_1 + f_2)(x) \subseteq \partial_C f_1(x) + \partial_C f_2(x) \]
(equality when \( f_1, f_2 \) regular)

Mean value theorem:
\[ \exists z \in (x,y) / f(y) - f(x) \in (\zeta, y - x) \]

Maximum functions:
\[ f(x) = \max_{1 \leq i \leq n} f_i(x) \] (each \( f_i \) smooth)
\[ I(x) = \{ i \in \{1, 2, \ldots, n\} : f_i(x) = f(x) \} \]
Then \( \partial_C f(x) = \text{co} \{ f'_i(x) : i \in I(x) \} \)

Optimization:
\[ \min_S f \text{ at } x \implies 0 \in \partial_C f(x) + N^C_S(x) \]
(more generally, Lagrange multipliers)
Some calculus

Mean value theorem:
\[ \exists \xi \in (a,b) : f'(\xi) = \frac{f(b) - f(a)}{b-a} \]

Sum:
\[ \partial C f(x_1 + x_2) \subset \partial C f(x_1) + \partial C f(x_2) \]
(equality when \( f_1, f_2 \) regular)

Maximum function:
\[ f(x) = \max_{x \in \Omega} f_i(x) \quad (\text{each } f_i \text{ smooth}) \]
\[ f(x) = \max_{i \in \{1,2,\ldots,n\}} f_i(x) \quad (\text{each } f_i \text{ smooth}) \]
Then \( \partial C f(x) = \{ f_i(x) : i \in n(x) \} \)

Graph-closed:
\[ \zeta \in \partial_C f(x_i) \]
\[ \downarrow \quad \downarrow \]
\[ \zeta \quad x \]

Gradient formula

When \( f \) is locally Lipschitz on \( \mathbb{R}^n \), then \( f \) is differentiable a.e. (Rademacher). Let \( \Omega \) be any set of measure 0 including the nondifferentiability points. Then
\[ \partial_C f(x) = \text{co} \{ \lim_{i \to \infty} \nabla f(x_i) : x_i \to x, x_i \notin \Omega \} \]

("blind to sets of measure 0")

This is a useful tool for calculation.

Example
\[ f(x, y) = \max \{ \min [x, -y], y - x \} \]
Example \[ f(x, y) = \max \{ \min [x, -y], y - x \} \]

Example \[ f(x, y) = \max \{ \min [x, -y], y - x \} \]

Example \[ f(x, y) = \max \{ \min [x, -y], y - x \} \]

Example \[ f(x, y) = \max \{ \min [x, -y], y - x \} \]
Example \[ f(x, y) = \max \{ \min [x, -y], y - x \} \]

\[
\begin{align*}
\n\n\n\n\end{align*}
\]

\[
\begin{align*}
\n\n\n\n\end{align*}
\]

\[
\begin{align*}
\n\n\n\n\end{align*}
\]

\[
\begin{align*}
\n\n\n\n\end{align*}
\]

Example \[ f(x, y) = \max \{ \min [x, -y], y - x \} \]

\[
\begin{align*}
\n\n\n\n\end{align*}
\]

\[
\begin{align*}
\n\n\n\n\end{align*}
\]

\[
\begin{align*}
\n\n\n\n\end{align*}
\]

Example \[ f(x, y) = \max \{ \min [x, -y], y - x \} \]

\[
\begin{align*}
\n\n\n\n\end{align*}
\]

\[
\begin{align*}
\n\n\n\n\end{align*}
\]

\[
\begin{align*}
\n\n\n\n\end{align*}
\]

Example \[ f(x, y) = \max \{ \min [x, -y], y - x \} \]

\[
\begin{align*}
\n\n\n\n\end{align*}
\]

\[
\begin{align*}
\n\n\n\n\end{align*}
\]

\[
\begin{align*}
\n\n\n\n\end{align*}
\]
**Example**  
$f(x, y) = \max \{ \min [x, -y], y - x \}$

\[ \nabla f = (-1, 1) \]
\[ f(x, y) = y - x \]
\[ \nabla f = (0, -1) \]
\[ f(x, y) = -y \]
\[ \nabla f = (1, 0) \]
\[ f(x, y) = x \]
\[ y = 2x \]
\[ y = x/2 \]

\[ \partial_C f(0, 0) = \text{co} \{ (1, 0), (0, -1), (-1, 1) \} \]

---

**Gradient formula**

When $f$ is locally Lipschitz on $\mathbb{R}^n$, then $f$ is differentiable a.e. (Rademacher). Let $\Omega$ be any set of measure 0 including the nondifferentiability points. Then

\[ \partial_C f(x) = \{ \lim_{i \to \infty} \nabla f(x_i) : x_i \to x, x_i \notin \Omega \} \]

This is a useful tool for calculation.

---

**Example**  
$f(x, y) = \max \{ \min [x, -y], y - x \}$

\[ \nabla f = (-1, 1) \]
\[ f(x, y) = y - x \]
\[ \nabla f = (0, -1) \]
\[ f(x, y) = -y \]
\[ \nabla f = (1, 0) \]
\[ f(x, y) = x \]
\[ y = 2x \]
\[ y = x/2 \]

\[ \partial_C f(0, 0) = \text{co} \{ (1, 0), (0, -1), (-1, 1) \} \]

---

**Gradient formula**

When $f$ is locally Lipschitz on $\mathbb{R}^n$, then $f$ is differentiable a.e. (Rademacher). Let $\Omega$ be any set of measure 0 including the nondifferentiability points. Then

\[ \partial_C f(x) = \{ \lim_{i \to \infty} \nabla f(x_i) : x_i \to x, x_i \notin \Omega \} \]

This is a useful tool for calculation.

When $f : \mathbb{R}^n \to \mathbb{R}^m$ is locally Lipschitz, we can define the **generalized Jacobian** this way:

\[ \partial_C f(x) := \{ \lim_{i \to \infty} Df(x_i) : x_i \to x, x_i \notin \Omega \}, \]

A convex set of $m \times n$ matrices. Then: inverse function theorem, Sard, etc. 

[General case $f : X \to Y$; Pales/Zeidler]
Theorem (1973)

Let $\partial^c F(x_0)$ be of maximal rank, where $F : \mathbb{R}^n \to \mathbb{R}^n$ is Lipschitz near $x_0$. Then there exist neighborhoods $U$ of $x_0$ and $V$ of $F(x_0)$ and a Lipschitz function $G : V \to \mathbb{R}^n$ such that

$$G(F(u)) = u \quad \forall \ u \in U,$$
$$F(G(v)) = v \quad \forall \ v \in V.$$ 

Example

$$F(x,y) = |x| + y, 2x + |y|$$

$$\partial^c F(0,0) = \left\{ \begin{bmatrix} s \\ 2 \\ t \end{bmatrix} : -1 \leq s \leq 1, \ -1 \leq t \leq 1 \right\}$$

$$\det \begin{bmatrix} s & 1 \\ 2 & t \end{bmatrix} = st - 2 \neq 0$$

Calculus of sets

Let $x_0 \in S := \{ x : f(x) \leq 0 \}$. If $0 \notin \partial^c f(x_0)$, then

$$T^c_S(x_0) \supset \{ v \in X : f'(x_0; v) \leq 0 \}.$$ 

If in addition $f$ is regular at $x_0$, then

$$T^c_S(x_0) = \{ v \in X : f'(x_0; v) \leq 0 \} \text{ and }$$
$$N^c_S(x_0) = \{ \lambda \zeta : \lambda \geq 0, \zeta \in \partial^c f(x_0) \}.$$ 

Calculus of sets

Let $x_0 \in S := \{ x : f(x) \leq 0 \}$. If $0 \notin \partial^c f(x_0)$, then

$$T^c_S(x_0) \supset \{ v \in X : f'(x_0; v) \leq 0 \}.$$ 

If in addition $f$ is regular at $x_0$, then

$$T^c_S(x_0) = \{ v \in X : f'(x_0; v) \leq 0 \} \text{ and }$$
$$N^c_S(x_0) = \{ \lambda \zeta : \lambda \geq 0, \zeta \in \partial^c f(x_0) \}.$$ 

Let $Y$ be another Banach space, and $F : X \to Y$ a continuously differentiable function. Set

$$S := \{ x \in X : F(x) = 0 \}.$$ 

If $F'(x_0)$ is surjective, then

$$T^c_S(x_0) = \{ v \in X : \langle F'(x_0), v \rangle = 0 \} \text{ and }$$
$$N^c_S(x_0) = \{ \theta F'(x_0) : \theta \in Y^* \}.$$
Calculus of sets
Let $Y$ be another Hausdorff space, and $F: X \rightarrow Y$ a continuously differentiable function. Set

$S = \{x \in X : F(x) = y\}.$

If $F'(a)$ is surjective, then

$T^C_F(a) = \{x \in X : \langle F'(a)x, a \rangle = 0\}$

with equality when $S$ is regular.

If $N^C_{S_1}(x) \cap -N^C_{S_2}(x) = \{0\}$, then

$N^C_{S_1 \cap S_2}(x) \subset N^C_{S_1}(x) + N^C_{S_2}(x)$

and

$T^C_{S_1 \cap S_2}(x) \supset T^C_{S_1}(x) \cap T^C_{S_2}(x),$

with equality when $S_1$ and $S_2$ are regular.

Wedged (or epi-Lipschitz) sets
A set $S$ is said to be \textbf{wedged} if:

$\text{int } T^C_S(x) \neq \emptyset \forall x \in S.$

Let $S \subset \mathbb{R}^n$ be wedged. Then

- $\text{int } S \neq \emptyset$
- $S = \text{cl} \{\text{int } S\}$
- $T^C_S(x) = \mathbb{R}^n$ iff $x \in \text{int } S$
- $S$ is locally the epigraph of a Lipschitz function.
Let $\phi : \mathbb{R}^n \to \mathbb{R}^n$ be a Lipschitz function satisfying
\[ \phi(x) \in T_{B(0,1)}(x) \quad \forall x \in B(0,1). \]
Then there exists $x_0 \in B(0,1)$ such that $\phi(x_0) = 0$.

($\iff$ Brouwer's Fixed Point Theorem)

**Boundary analysis:** Inner and outer sphere conditions, lower $C^2$ property, reach, semiconcavity, $\phi$-convexity, packing, etc. (Federer, Stern, Colombo, Nour, Cannarsa...)
Boundary analysis: Inner and outer sphere conditions, lower $C^2$ property, reach, semiconcavity, $\phi$-convexity, packing, etc. (Federer, Stern, Colombo, Nour, Cannarsa...)

Example: The union of uniform closed balls conjecture

$S \subset \mathbb{R}^n$ has the R-inner ball property if:

Is $S$ then the union of balls of radius $R$?

No in general
**Boundary analysis:** Inner and outer sphere conditions, lower $C^2$ property, reach, semiconcavity, $\phi$-convexity, packing, etc. (Federer, Stern, Colombo, Nour, Cannarsa...)

**Example:** The union of uniform closed balls conjecture

$S \subset \mathbb{R}^n$ has the R-inner ball property if:

- Is $S$ then the union of balls of radius $R$? No in general
- Is $S$ the union of balls of radius $r$, $0 < r < R$? Yes, if $S$ is wedged and bounded

**Boundary analysis:** Inner and outer sphere conditions, lower $C^2$ property, reach, semiconcavity, $\phi$-convexity, packing, etc. (Federer, Stern, Colombo, Nour, Cannarsa...)

**Example:** The union of uniform closed balls conjecture

$S \subset \mathbb{R}^n$ has the R-inner ball property if:

- Is $S$ then the union of balls of radius $R$? No in general
- Is $S$ the union of balls of radius $r$, $0 < r < R$? Yes, if $S$ is wedged and bounded

---

**Proximal theory**

The classical derivative corresponds to a two-sided local approximation by an affine function.

$$f'(\alpha) = \text{slope}$$
Proximal theory

The classical derivative corresponds to a two-sided local approximation by an affine function.

\[ f'(\alpha) = \text{slope} \]

We may also approximate just from below, using nonlinear functions: proximal analysis

The set of all `contact slopes' of lower locally supporting parabolas is the proximal subdifferential \( \partial_P f(\alpha) \)

(P = proximal) subdifferential \( \partial_P f(\alpha) \)
The set of all ‘contact slopes’ of lower locally supporting parabolas is the proximal subdifferential $\partial_p f(\alpha)$.

$\partial_p f(\alpha) = [-2, 1]$
The set of all `contact slopes` of lower locally supporting parabolas is the proximal subdifferential $\partial_P f(\alpha)$

\[ \partial_P f(\alpha) = [-2, 1] \]

We cannot expect to have, in general:

\[ \partial_P (f_1 + f_2)(x) \subset \partial_P f_1(x) + \partial_P f_2(x) \]

In fact, the opposite is true (but not useful)!

\[ \zeta \in \partial_P (f_1 + f_2)(x) \implies \]

We cannot expect to have, in general:

\[ \partial_P (f_1 + f_2)(x) \subset \partial_P f_1(x) + \partial_P f_2(x) \]

In fact, the opposite is true (but not useful)!

\[ \zeta \in \partial_P (f_1 + f_2)(x) \implies \]

\[ \zeta \in \partial_P f(\alpha) \iff \quad f(x) \geq \langle \zeta, x - \alpha \rangle + f(\alpha) - \sigma |x - \alpha|^2 \text{ locally} \]

\[ \partial_P f \text{ has a very complete (but fuzzy!) theory and calculus...} \]

Borwein, Ioffe, Ledyaev, Loewen, Rockafellar, Vinter, Zeidan...
We cannot expect to have, in general:
\[ \partial_P(f_1 + f_2)(x) \subset \partial_P f_1(x) + \partial_P f_2(x) \]

In fact, the **opposite** is true (but not useful)!

\[ \zeta \in \partial_P(f_1 + f_2)(x) \implies \forall \epsilon > 0 \exists x_1, x_2, \eta : \]
\[ |x_i - x| < \epsilon, \quad |\eta| < \epsilon, \quad |f_i(x_i) - f_i(x)| < \epsilon \quad (i = 1, 2) \]

and
\[ \zeta \in \partial_P f_1(x_1) + \partial_P f_2(x_2) + \eta \]
The geometry of proximal normals

\[ S \]

\[ x \]

\[ y \]

\[ \zeta \text{ is a proximal normal to } S \text{ at } x \]
The geometry of proximal normals

\( \zeta \in N^P_S(x) \iff \exists \sigma \geq 0 : \langle \zeta, y - x \rangle \leq \sigma |y - x|^2 \ \forall y \in S \)

\( \zeta \) is a proximal normal to \( S \) at \( x \)

Such normals don’t exist at every boundary point of \( S \), but in finite dimensions they exist “often”, and generate the cone \( N^P_S(x) \)

\( \zeta \in N^P_S(x) \iff \exists \sigma \geq 0 : \langle \zeta, y - x \rangle \leq \sigma |y - x|^2 \ \forall y \in S \)

In infinite dimensions, closest points may not exist, and proximal normals may be scarce. But they exist “densely” in a Hilbert space (Lau’s Theorem), or, more generally, in smooth Banach spaces

Fact:

if \( f \) is lower semicontinuous, finite at \( x \), then

\[ \zeta \in \partial_pf(x) \iff (\zeta, -1) \in N^P_{ept f}(x, f(x)) \]

In infinite dimensions, closest points may not exist, and proximal normals may be scarce. But they exist “densely” in a Hilbert space (Lau’s Theorem), or, more generally, in smooth Banach spaces
**Limiting constructs**  
When proximal normals exist densely, as in a Hilbert space, we define

$$N_S^L(x) = \left\{ \lim_{i \to \infty} \zeta_i : \zeta_i \in N_S^P(x_i), x_i \to_S x \right\}$$

$$\partial_L f(x) = \left\{ \lim_i \zeta_i : \zeta_i \in \partial_P f(x_i), x_i \to x, f(x_i) \to f(x) \right\}$$

$L = \text{Limiting}$

**Limiting constructs**  
When proximal normals exist densely, as in a Hilbert space, we define

$$N_S^L(x) = \left\{ \lim_{i \to \infty} \zeta_i : \zeta_i \in N_S^P(x_i), x_i \to_S x \right\}$$

$$\partial_L f(x) = \left\{ \lim_i \zeta_i : \zeta_i \in \partial_P f(x_i), x_i \to x, f(x_i) \to f(x) \right\}$$

These constructs inherit a calculus that is "less fuzzy".

For example:

In finite dimensions, if $f$ and $g$ are lower semicontinuous, and if one of them is Lipschitz near $x$, then

$$\partial_L (f + g)(x) \subset \partial_L f(x) + \partial_L g(x)$$

$\partial_C f \ \text{vis-à-vis} \ \partial_P f / \partial_L f$

- All of these reduce to the subdifferential if $f$ is convex, to the derivative if $f$ is smooth
\( \partial_C f \) vis-à-vis \( \partial_P f / \partial_L f \)
- All of these reduce to the subdifferential if \( f \) is convex, to the derivative if \( f \) is smooth.
- \( \partial_C f \) can be defined on any Banach space, along with its geometry; most useful for Lipschitz functions; can be estimated by \( f^\circ \), or by the gradient formula ('blind to sets of measure 0'); gives directions of decrease and tangency; has a vector-valued extension ('generalized Jacobian'); used in all the numerical implementations.

\( \partial_C f \) vis-à-vis \( \partial_P f / \partial_L f \)
- All of these reduce to the subdifferential if \( f \) is convex, to the derivative if \( f \) is smooth.
- \( \partial_C f \) can be defined on any Banach space, along with its geometry; most useful for Lipschitz functions; can be estimated by \( f^\circ \), or by the gradient formula ('blind to sets of measure 0'); gives directions of decrease and tangency; has a vector-valued extension ('generalized Jacobian'); used in all the numerical implementations.
- \( \partial_P f \) can be defined on 'smooth spaces'; applies to lsc functions; smaller but difficult to calculate; its emptiness can be a plus in the theory (as in viscosity solutions); has links to 'variational principles'.
- For a Lipschitz function on a Hilbert space we have
\[
\partial_C f = \text{co} \partial_L f
\]

Two references chosen at random:

Optimization and Nonsmooth Analysis
Clarke, 1983

Nonsmooth Analysis and Control Theory
Clarke, Ledyaev, Stern and Wolenski, Graduate Texts in Mathematics 1998

clarke@math.univ-lyon1.fr
(or web site)
THE END