Nonsmooth Analysis and Applications

Francis Clarke

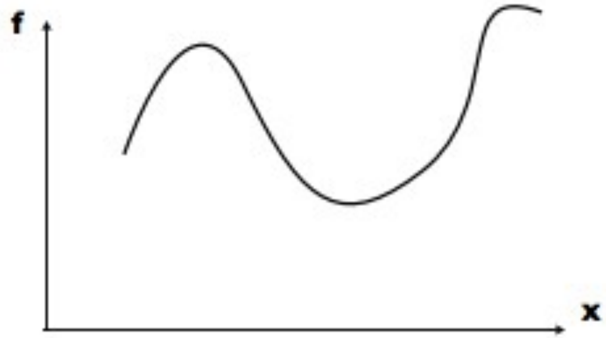
**Institut universitaire de France
et Université de Lyon**

Classical Calculus

A basic technique in mathematics is linearization

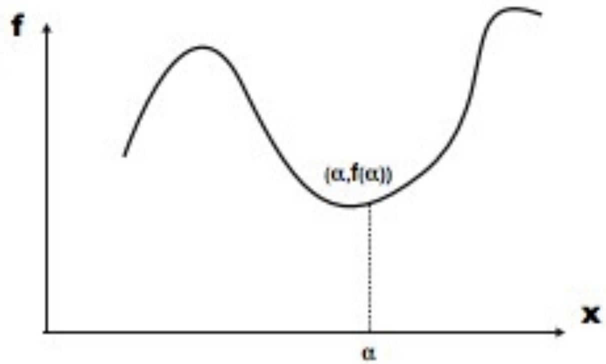
Classical Calculus

A basic technique in mathematics is linearization



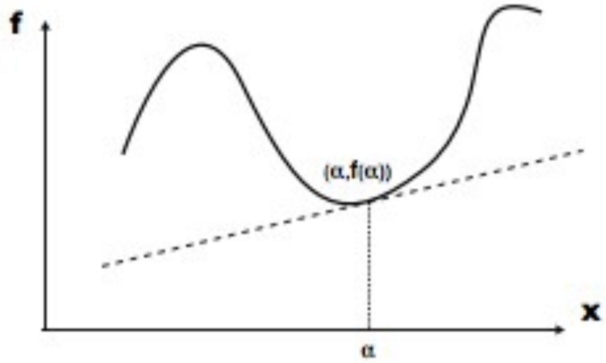
Classical Calculus

A basic technique in mathematics is linearization



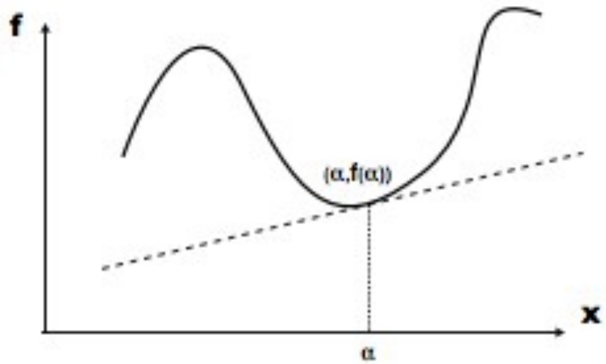
Classical Calculus

A basic technique in mathematics is linearization



Classical Calculus

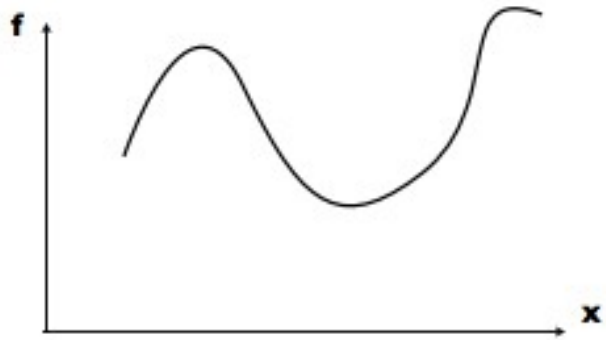
A basic technique in mathematics is linearization



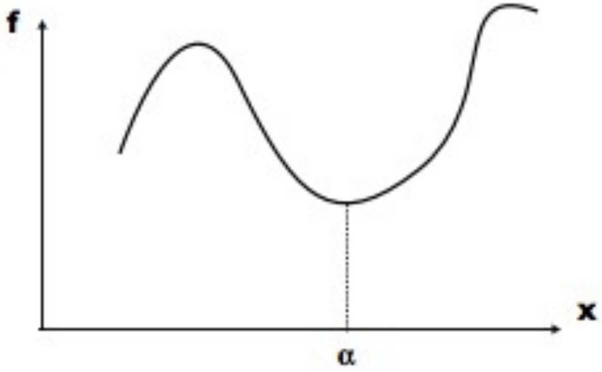
$f'(\alpha)$ = the slope of the tangent line to the graph of f through the point $(\alpha, f(\alpha))$

A. Minimizing a function f(x)

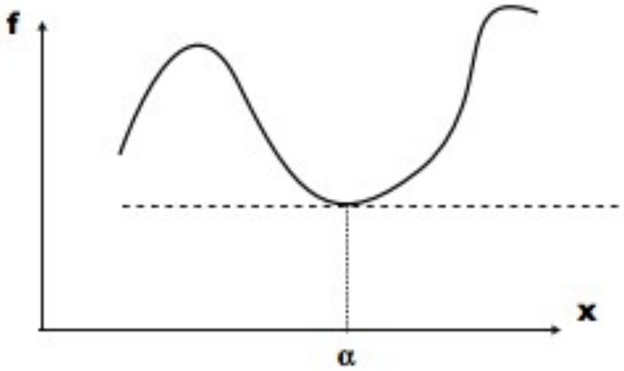
A. Minimizing a function f(x)



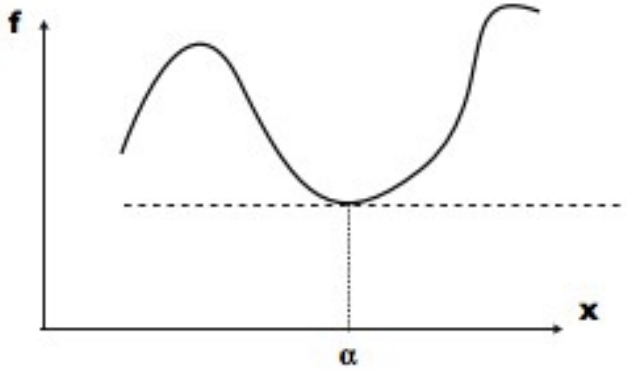
A. Minimizing a function f(x)



A. Minimizing a function f(x)

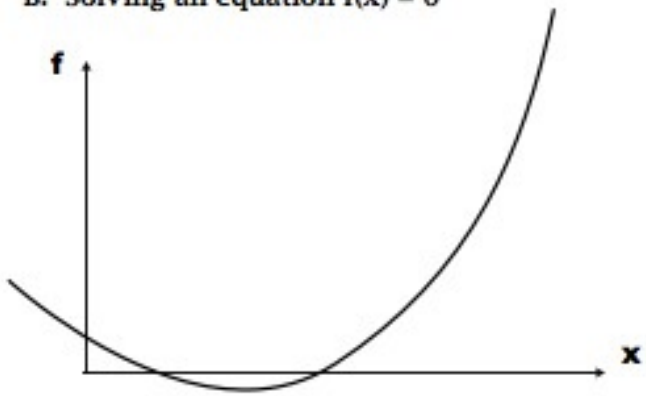


A. Minimizing a function f(x)

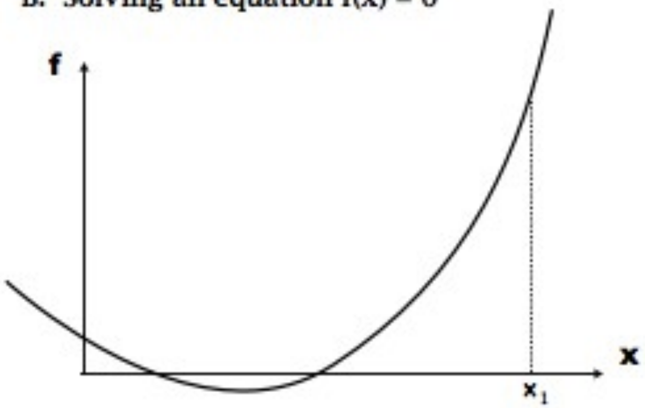


Fermat's rule : at a minimum α, we have $f'(\alpha) = 0$

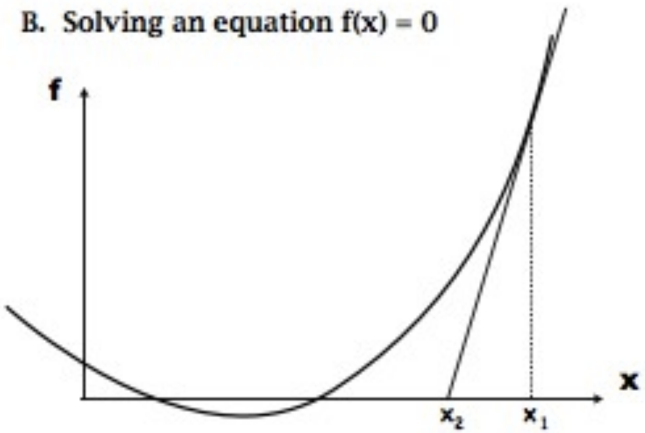
B. Solving an equation $f(x) = 0$



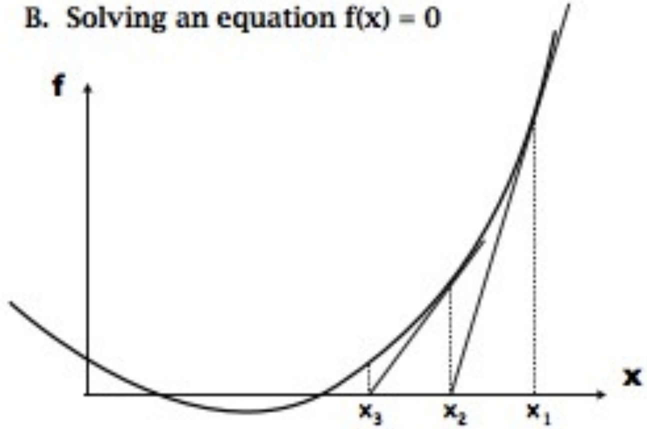
B. Solving an equation $f(x) = 0$



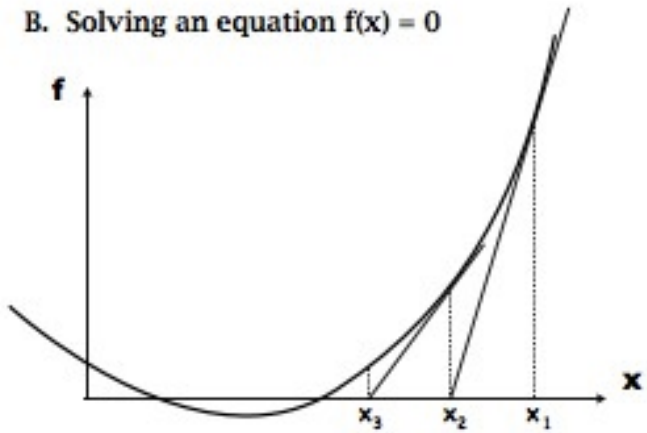
B. Solving an equation $f(x) = 0$



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 $\begin{array}{l} x'(t)=f(x(t),y(t)) \\ y'(t)=g(x(t),y(t)) \end{array}$

around an equilibrium (0,0)

C. Studying a system $\begin{array}{l} x'(t)=f(x(t),y(t)) \\ y'(t)=g(x(t),y(t)) \end{array}$

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

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = A \begin{bmatrix} x \\ y \end{bmatrix}$$

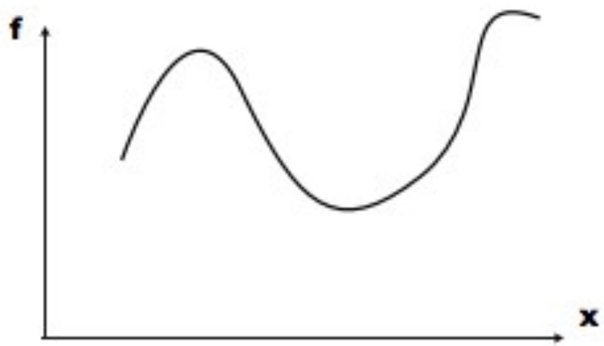
via eigenvalues...

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

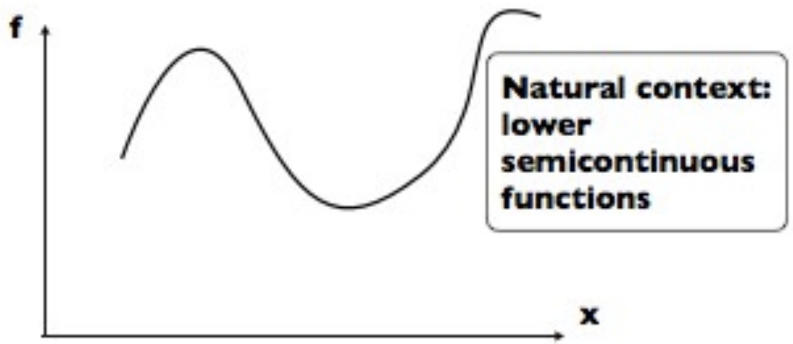
Thus, in the space of almost precisely one century, infinitesimal calculus, or as we now call it in English, The Calculus, the calculating tool *par excellence*, had been forged; and nearly three centuries of constant use have not dulled this incomparable instrument. Bourbaki

But what if f is not differentiable?

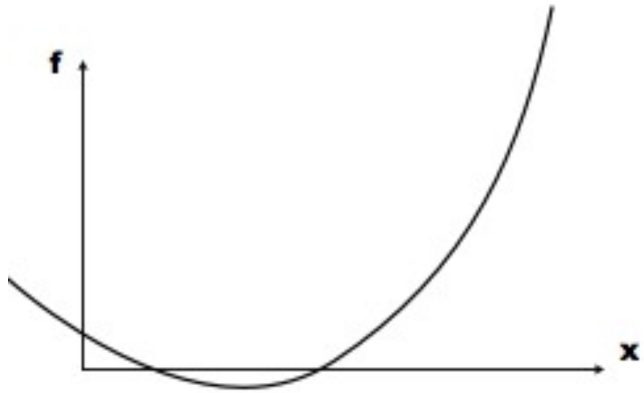
A. Minimizing a function f(x)



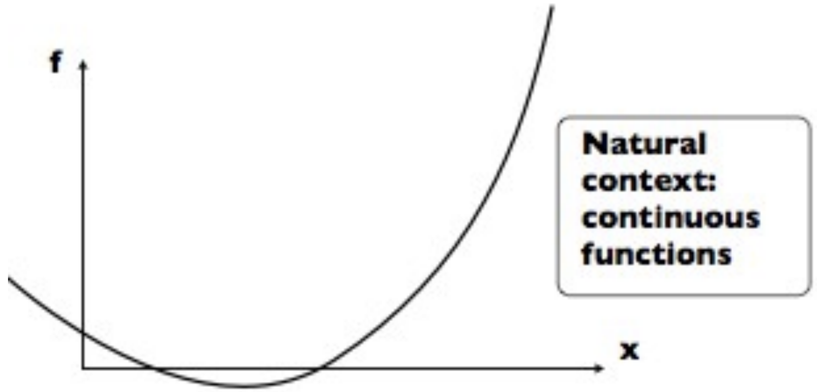
A. Minimizing a function f(x)



B. Solving an equation f(x) = 0



B. Solving an equation f(x) = 0



C. Studying a system $x'(t) = f(x(t), y(t))$
$y'(t) = g(x(t), y(t))$

around an equilibrium (0,0)

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

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

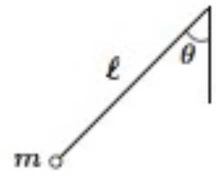
Nonsmooth Analysis began with "Dini Derivates" :

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

The swing in your backyard

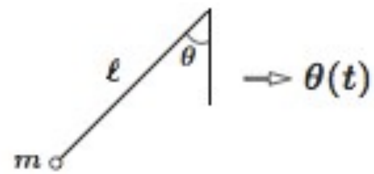
The swing in your backyard

The nonlinear pendulum



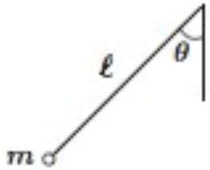
The swing in your backyard

The nonlinear pendulum



The swing in your backyard

The nonlinear pendulum



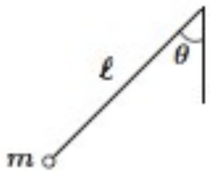
$\longrightarrow \theta(t)$

Newton(-Euler)

$$m\,\ell\,\theta'' = -m\,g\sin\theta \implies \theta'' + (g/\ell)\sin\theta = 0$$

The swing in your backyard

The nonlinear pendulum



$\longrightarrow \theta(t)$

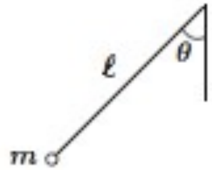
Newton(-Euler)

$$m\,\ell\,\theta'' = -m\,g\sin\theta \implies \theta'' + (g/\ell)\sin\theta = 0$$

equilibrium $\theta = 0$

The swing in your backyard

The nonlinear pendulum



$\longrightarrow \theta(t)$

Newton(-Euler)

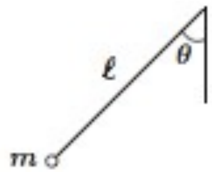
$$m\,\ell\,\theta'' = -m\,g\sin\theta \implies \theta'' + (g/\ell)\underbrace{\sin\theta}_{\downarrow} = 0$$

equilibrium $\theta = 0$

$$\theta'' + (g/\ell)\theta = 0$$

The swing in your backyard

The nonlinear pendulum



$\longrightarrow \theta(t)$

Newton(-Euler)

$$m\,\ell\,\theta'' = -m\,g\sin\theta \implies \theta'' + (g/\ell)\underbrace{\sin\theta}_{\downarrow} = 0$$

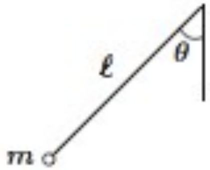
equilibrium $\theta = 0$

$$\theta'' + (g/\ell)\theta = 0$$

What if there's a wind?

The swing in your backyard

The nonlinear pendulum



$\rightarrow \theta(t)$

Newton(-Euler)

$$m\,\ell\,\theta'' = -m\,g\sin\theta \implies \theta'' + (g/\ell)\underbrace{\sin\theta}_{\downarrow} = 0$$

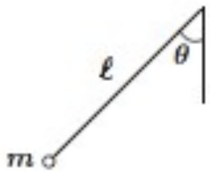
equilibrium $\theta = 0$

$$\theta'' + (g/\ell)\theta = 0$$

What if there's a wind? $\longleftarrow$ force f

The swing in your backyard

The nonlinear pendulum



$\rightarrow \theta(t)$

Newton(-Euler)

$$m\,\ell\,\theta'' = -m\,g\sin\theta \implies \theta'' + (g/\ell)\underbrace{\sin\theta}_{\downarrow} = 0$$

equilibrium $\theta = 0$

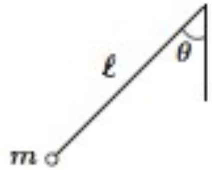
$$\theta'' + (g/\ell)\theta = 0$$

What if there's a wind? $\longleftarrow$ force f

$$m\,\ell\,\theta'' = -m\,g\sin\theta - f\cos\theta$$

The swing in your backyard

The nonlinear pendulum



$\rightarrow \theta(t)$

Newton(-Euler)

$$m\,\ell\,\theta'' = -m\,g\sin\theta \implies \theta'' + (g/\ell)\underbrace{\sin\theta}_{\downarrow} = 0$$

equilibrium $\theta = 0$

$$\theta'' + (g/\ell)\theta = 0$$

What if there's a wind? $\longleftarrow$ force f

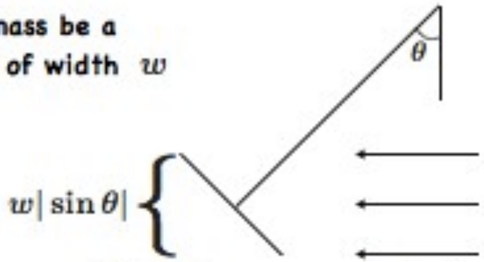
$$m\,\ell\,\theta'' = -m\,g\sin\theta - f\cos\theta$$

equilibrium $\theta = \theta_0$, $\tan\theta_0 = -f/(mg)$

Let the mass be a thin seat of width w

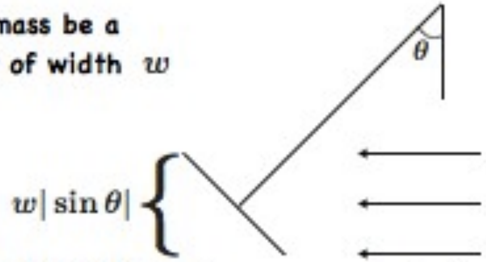
Reality is more complicated...

Let the mass be a thin seat of width w



Reality is more complicated...

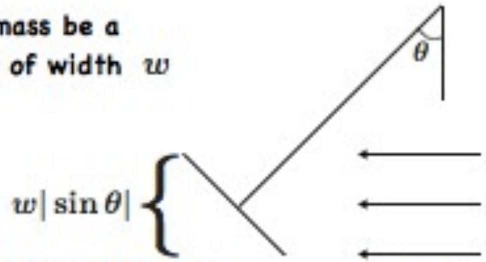
Let the mass be a thin seat of width w



$f = cw|\sin\theta|$

Reality is more complicated...

Let the mass be a thin seat of width w



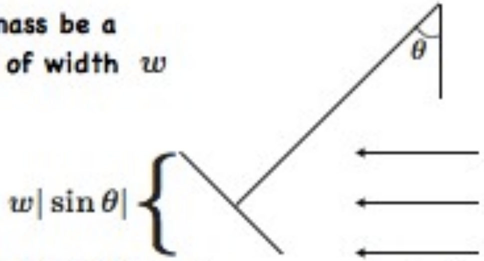
$f = cw|\sin\theta|$

Reality is more complicated...

$$m\,\ell\,\theta'' = -m\,g\sin\theta - f\cos\theta$$

$$\Rightarrow \quad m\,\ell\,\theta'' = -m\,g\sin\theta - cw|\sin\theta|\cos\theta$$

Let the mass be a
thin seat of width w



$f = cw|\sin\theta|$

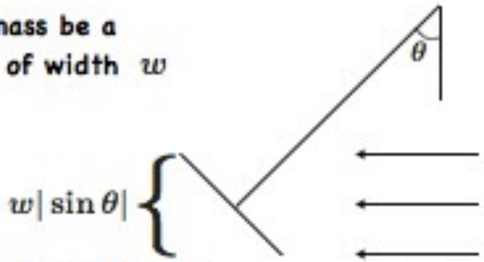
Reality is more complicated...

$$m\,\ell\,\theta'' = -m\,g\sin\theta - f\cos\theta$$

$$\Rightarrow\quad m\,\ell\,\theta'' = -m\,g\sin\theta - cw|\sin\theta|\cos\theta$$

0 is now one of two equilibria;
no linearization

Let the mass be a
thin seat of width w

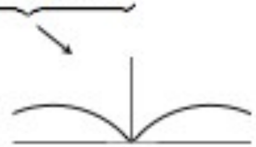


$f = cw|\sin\theta|$

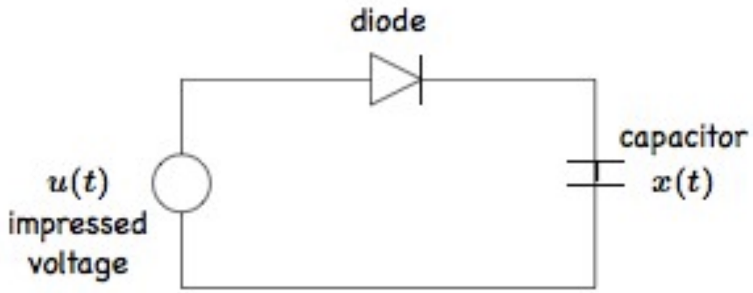
Reality is more complicated...

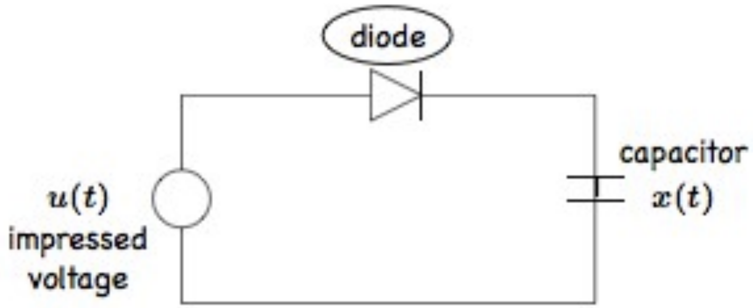
$$m\,\ell\,\theta'' = -m\,g\sin\theta - f\cos\theta$$

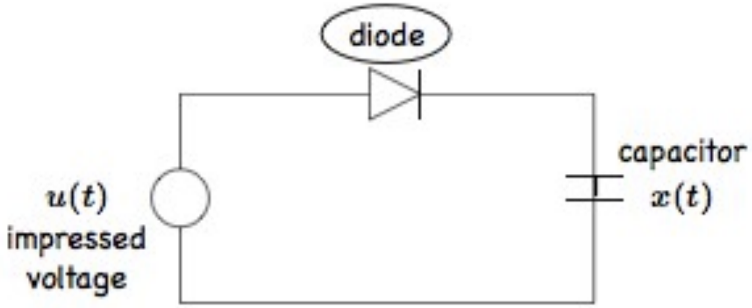
$$\Rightarrow\quad m\,\ell\,\theta'' = -m\,g\sin\theta - \underbrace{cw|\sin\theta|\cos\theta}$$



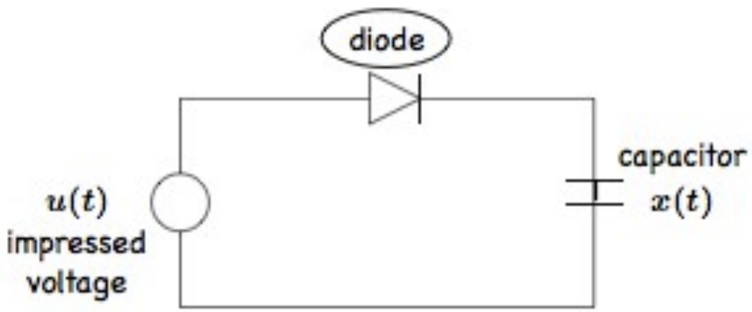
0 is now one of two equilibria;
no linearization







$$\frac{d}{dt}x(t) = \begin{cases} \alpha\big(u(t) - x(t)\big) & \text{if } x(t) \le u(t) \\ -\beta\big(x(t) - u(t)\big) & \text{if } x(t) \ge u(t) \end{cases}$$



$$\frac{d}{dt}x(t) = \begin{cases} \alpha\big(u(t) - x(t)\big) & \text{if } x(t) \le u(t) \\ -\beta\big(x(t) - u(t)\big) & \text{if } x(t) \ge u(t) \end{cases}$$

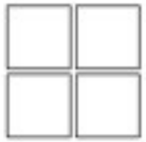
$f(x,u)$ has a corner at $x = u$

Directional or threshold phenomena are often nonsmooth.
Also where there is presence of shapes... another example:

nonsmooth contact

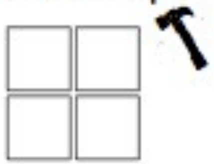
Directional or threshold phenomena are often nonsmooth.
Also where there is presence of shapes... another example:

nonsmooth contact



Directional or threshold phenomena are often nonsmooth.
Also where there is presence of shapes... another example:

nonsmooth contact

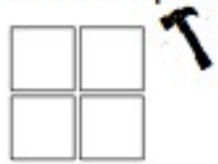


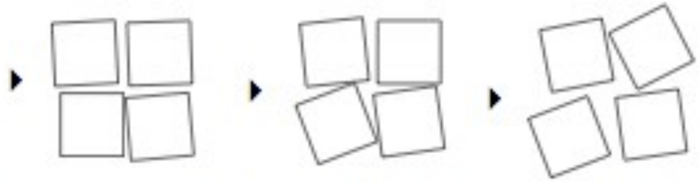
Directional or threshold phenomena are often nonsmooth.
Also where there is presence of shapes... another example:

nonsmooth contact

Directional or threshold phenomena are often nonsmooth.
Also where there is presence of shapes... another example:

nonsmooth contact

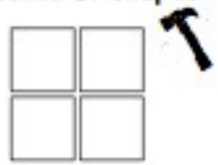


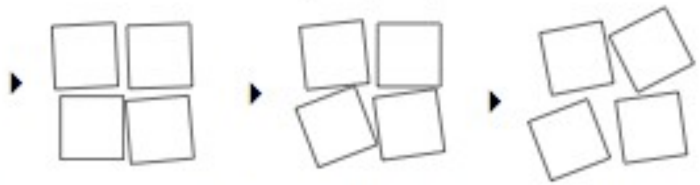


See Marsden et alii (generalized gradients + least action).
Other nonsmooth mechanics and elasticity: Brogliato, Moreau, Panagiotopoulos, Paoli, Schatzman, Schuricht ...

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 ...

hard-working dots

Optimization

Example: eigenvalue design

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

Optimization

Example: eigenvalue design

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

A function of interest:

$f(x)$:= the greatest eigenvalue of $A(x)$.

Optimization

Example: eigenvalue design

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

A function of interest:

$f(x)$:= the greatest eigenvalue of $A(x)$.

FACT: f is nonsmooth in general

Optimization

Example: eigenvalue design

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

A function of interest:

$f(x)$:= 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

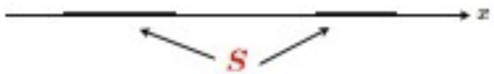
$$d_S(x) := \min_{s \in S} |x - s|$$

$S \subset \mathbb{R}^n$ (closed, nonempty)

Example: the distance function

$$d_S(x) := \min_{s \in S} |x - s|$$

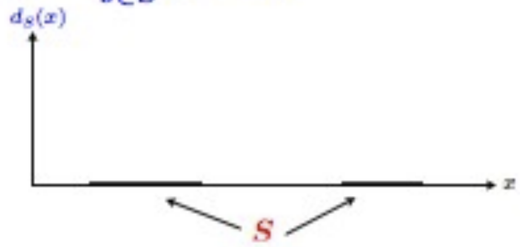
$S \subset \mathbb{R}^n$ (closed, nonempty)



Example: the distance function

$$d_S(x) := \min_{s \in S} |x - s|$$

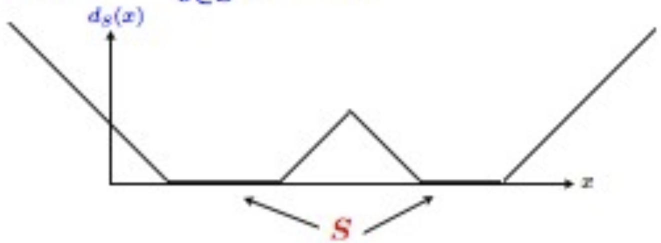
$S \subset \mathbb{R}^n$ (closed, nonempty)



Example: the distance function

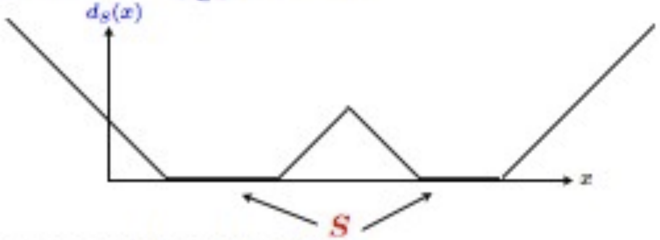
$$d_S(x) := \min_{s \in S} |x - s|$$

$S \subset \mathbb{R}^n$ (closed, nonempty)



Example: the distance function $S \subset \mathbb{R}^n$ (closed, nonempty)

$$d_S(x) := \min_{s \in S} |x - s|$$

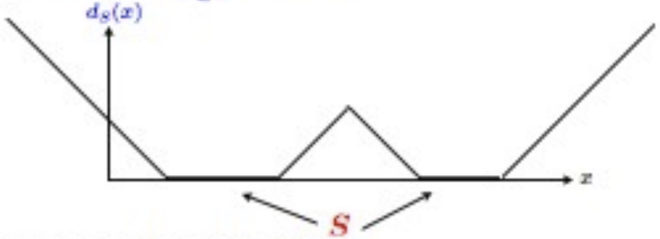


One use: exact penalization

$$\min_{x \in S} g(x) \quad \underset{?}{\Leftrightarrow} \quad \min_x g(x) + k\, d_S(x)$$

Example: the distance function $S \subset \mathbb{R}^n$ (closed, nonempty)

$$d_S(x) := \min_{s \in S} |x - s|$$



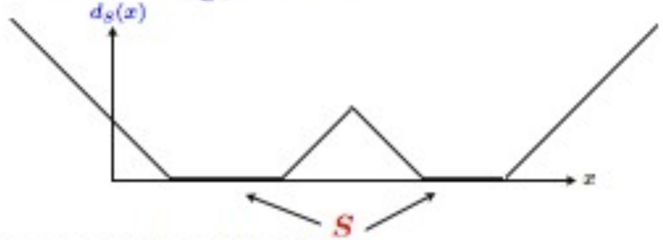
One use: exact penalization

$$\min_{x \in S} g(x) \quad \underset{?}{\Leftrightarrow} \quad \min_x g(x) + k\, d_S(x)$$

Then: $0 \in \partial\{g + k d_S\}(x)$

Example: the distance function $S \subset \mathbb{R}^n$ (closed, nonempty)

$$d_S(x) := \min_{s \in S} |x - s|$$



One use: exact penalization

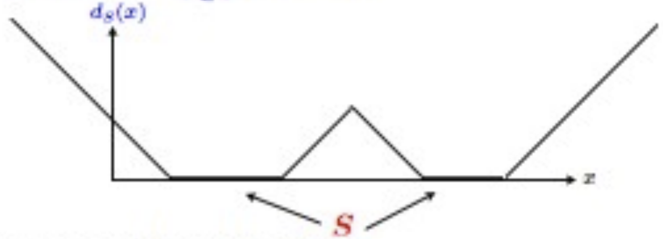
$$\min_{x \in S} g(x) \quad \underset{?}{\Leftrightarrow} \quad \min_x g(x) + k\, d_S(x)$$

Then: $0 \in \partial\{g + k d_S\}(x)$

$\Longrightarrow 0 \in \partial g(x) + k \partial d_S(x)$

Example: the distance function $S \subset \mathbb{R}^n$ (closed, nonempty)

$$d_S(x) := \min_{s \in S} |x - s|$$



One use: exact penalization

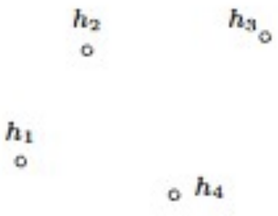
$$\min_{x \in S} g(x) \quad \underset{?}{\Leftrightarrow} \quad \min_x g(x) + k\, d_S(x)$$

Then: $0 \in \partial\{g + k d_S\}(x)$

$\Longrightarrow 0 \in \partial g(x) + k \boxed{\partial d_S(x)}$

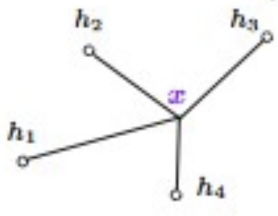
geometric interpretation via normals or (Euler-) Lagrange multipliers ?

Example: Problem of Torricelli/Steiner, n = 4



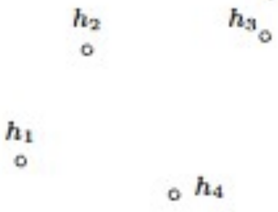
Example: Problem of Torricelli/Steiner, n = 4

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



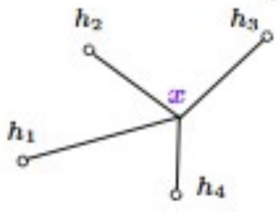
Example: Problem of Torricelli/Steiner, n = 4

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



Example: Problem of Torricelli/Steiner, n = 4

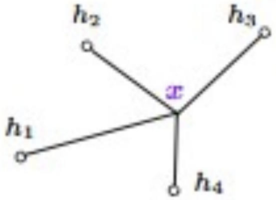
We seek a central point x relative to h_1, h_2, h_3, h_4: x minimizes $|h_1 - x| + |h_2 - x| + |h_3 - x| + |h_4 - x|$



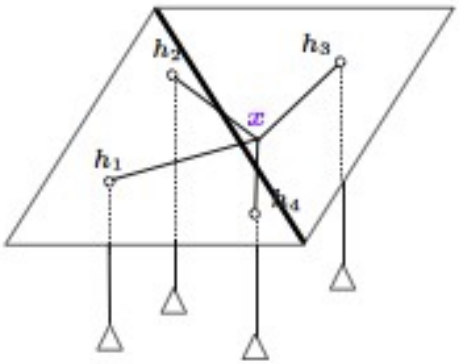
A table can solve this problem...

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...

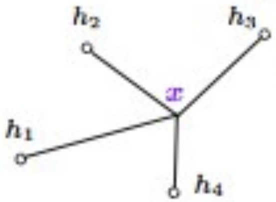


Analytically, x solves the problem iff

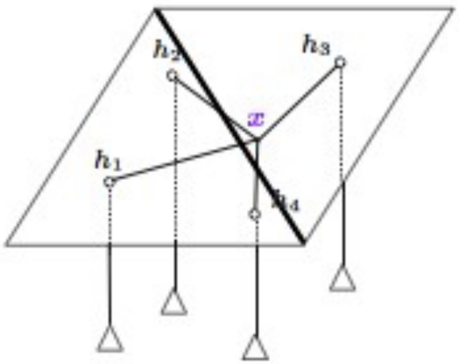
$$0 \in \sum_{i=1}^{4} \partial|x - h_i| \text{ where } \partial|x - h_i| = \begin{cases} B & \text{if } x = h_i \\ \frac{x-h_i}{|x-h_i|} & \text{if } x \neq h_i \end{cases}$$

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

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

Let us add one string going over the edge: Then

$$\min_x \sum_{i=1}^{4} |x - h_i| + c d_S(x),$$ S = complement of the table

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

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

We have $x''(t) = g(x(t))$ with $g = \nabla V$ discontinuous.

And also local minima:

Analytically, x solves the problem iff

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

Let us add one string going over the edge: Then

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

We have $x''(t) = g(x(t))$ with $g = \nabla V$ discontinuous.

And also local minima:

Analytically, x solves the problem iff

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

Let us add one string going over the edge: Then

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

We have $x''(t) = g(x(t))$ with $g = \nabla V$ discontinuous.

And also local minima:

Analytically, x solves the problem iff

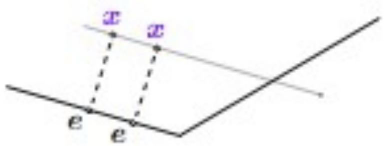
$$0 \in \sum_{i=1}^{4} \partial|x-h_i| \text{ where } \partial|x-h_i| = \begin{cases} B & \text{if } x = h_i \\ \frac{x-h_i}{|x-h_i|} & \text{if } x \neq h_i \end{cases}$$

Let us add one string going over the edge: Then

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

We have $x''(t) = g(x(t))$ with $g = \nabla V$ discontinuous.

And also local minima:



Analytically, x solves the problem iff

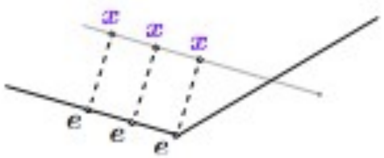
$$0 \in \sum_{i=1}^{4} \partial|x-h_i| \text{ where } \partial|x-h_i| = \begin{cases} B & \text{if } x = h_i \\ \frac{x-h_i}{|x-h_i|} & \text{if } x \neq h_i \end{cases}$$

Let us add one string going over the edge: Then

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

We have $x''(t) = g(x(t))$ with $g = \nabla V$ discontinuous.

And also local minima:



Analytically, x solves the problem iff

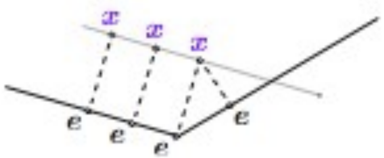
$$0 \in \sum_{i=1}^{4} \partial|x-h_i| \text{ where } \partial|x-h_i| = \begin{cases} B & \text{if } x = h_i \\ \frac{x-h_i}{|x-h_i|} & \text{if } x \neq h_i \end{cases}$$

Let us add one string going over the edge: Then

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

We have $x''(t) = g(x(t))$ with $g = \nabla V$ discontinuous.

And also local minima:



Analytically, x solves the problem iff

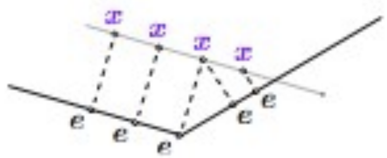
$$0 \in \sum_{i=1}^{4} \partial|x-h_i| \text{ where } \partial|x-h_i| = \begin{cases} B & \text{if } x = h_i \\ \frac{x-h_i}{|x-h_i|} & \text{if } x \neq h_i \end{cases}$$

Let us add one string going over the edge: Then

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

We have $x''(t) = g(x(t))$ with $g = \nabla V$ discontinuous.

And also local minima:



Analytically, x solves the problem iff

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

Let us add one string going over the edge: Then

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

We have $x''(t) = g(x(t))$ with $g = \nabla V$ discontinuous.

And also local minima:

Analytically, x solves the problem iff

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

Let us add one string going over the edge: Then

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

We have $x''(t) = g(x(t))$ with $g = \nabla V$ discontinuous.

And also local minima:

Analytically, x solves the problem iff

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

Let us add one string going over the edge: Then

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

We have $x''(t) = g(x(t))$ with $g = \nabla V$ discontinuous.

And also local minima:

Analytically, x solves the problem iff

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

Let us add one string going over the edge: Then

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

We have $x''(t) = g(x(t))$ with $g = \nabla V$ discontinuous.

And also local minima:

Analytically, x solves the problem iff

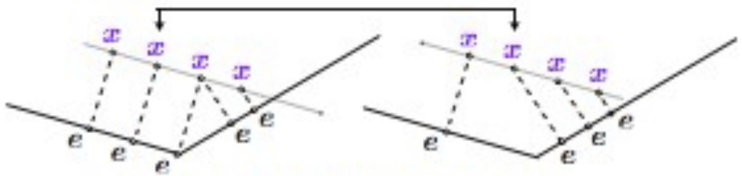
$$0 \in \sum_{i=1}^{4} \partial|x-h_i| \text{ where } \partial|x-h_i| = \begin{cases} B & \text{if } x = h_i \\ \frac{x-h_i}{|x-h_i|} & \text{if } x \neq h_i \end{cases}$$

Let us add one string going over the edge: Then

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

We have $x''(t) = g(x(t))$ with $g = \nabla V$ discontinuous.

And also local minima:



hysteresis (non-reversible dynamics)

Calculus of variations

The Basic Problem:

$$\min_{x(\cdot)} \int_a^b L(t, x(t), x'(t))\, dt,\ x(a) = A, x(b) = B$$

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.

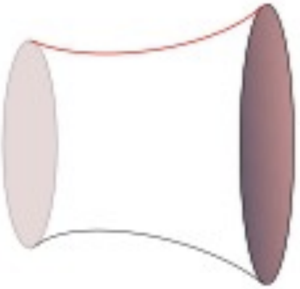
Leonhard Euler

1707-1783

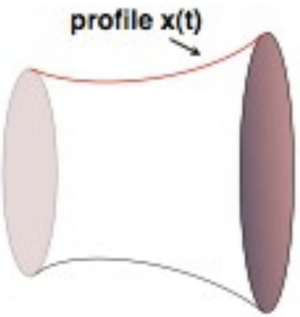
Example: soap bubble

Example: soap bubble

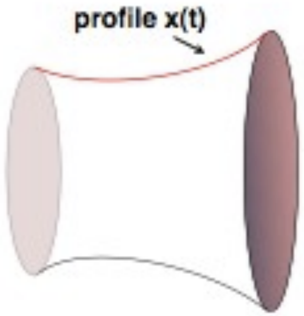
Example: soap bubble



Example: soap bubble



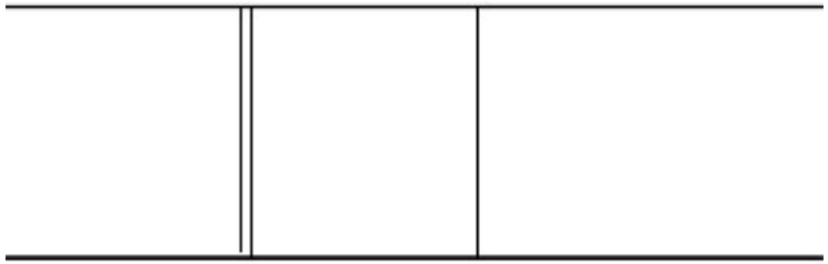
Example: soap bubble



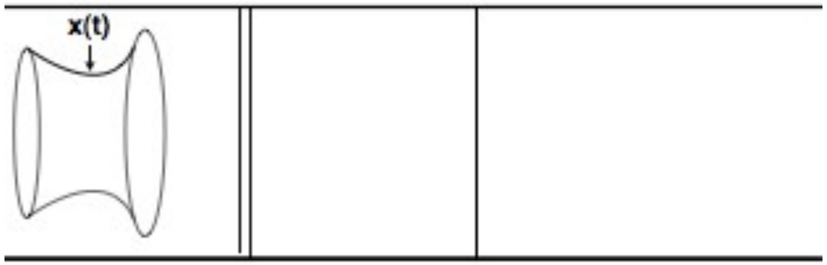
surface area | **Euler(-Lagrange) equation**

$$\min_{x(\cdot)} \int_a^b x(t)\sqrt{1+x'(t)^2}\,dt \quad \Rightarrow \quad \frac{d}{dt}\left\{\frac{x'(t)x(t)}{\sqrt{1+x'(t)^2}}\right\} = \sqrt{1+x'(t)^2}$$

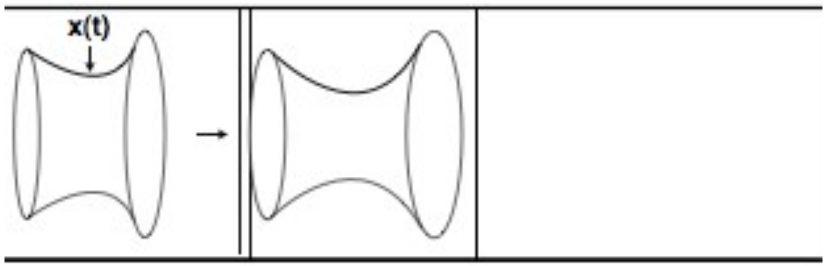
Solutions with corners



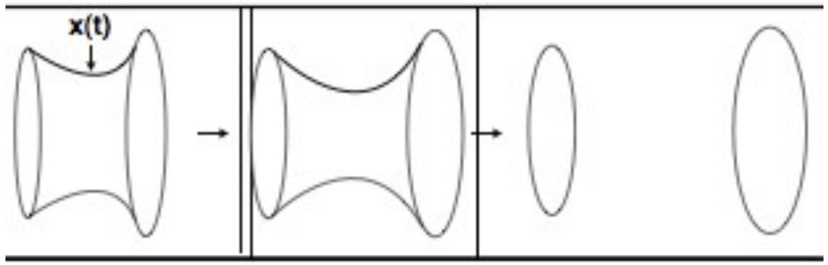
Solutions with corners



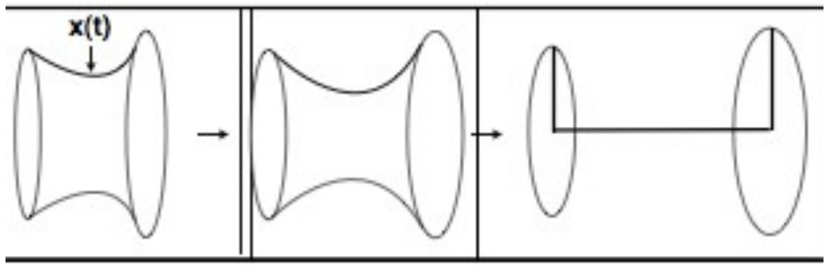
Solutions with corners



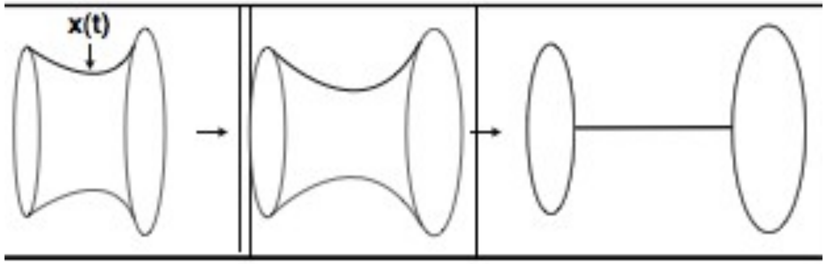
Solutions with corners



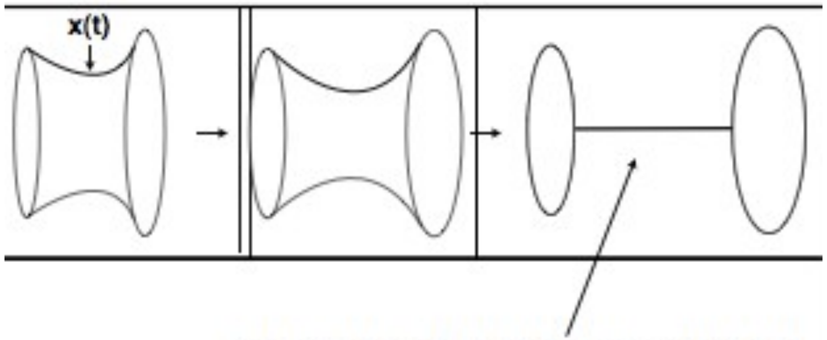
Solutions with corners



Solutions with corners



Solutions with corners



The Goldschmidt solution (1831)

A design problem

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.

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.

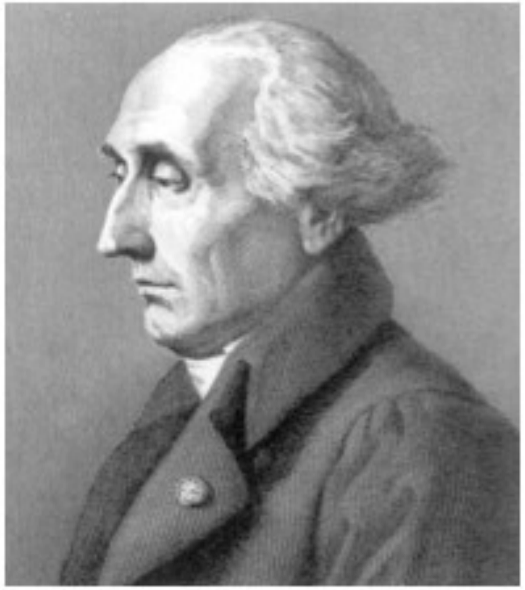
Lagrange (1770) *Sur la figure des colonnes*

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**

Lagrange

Lagrange

- After 20 years in Berlin, he joins the Paris Academy in 1786

Lagrange

- After 20 years in Berlin, he joins the Paris Academy in 1786
- During the revolution : metric system, Ecole Normale and Polytechnique

Lagrange

- 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

Lagrange

- 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

Lagrange

- 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

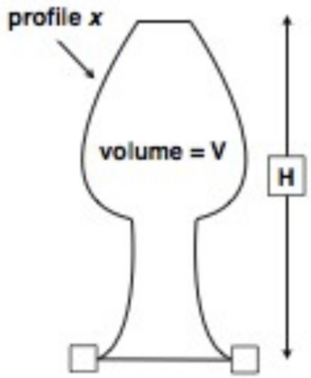
Designing an optimal column

1. Choose a profile *x*

Designing an optimal column

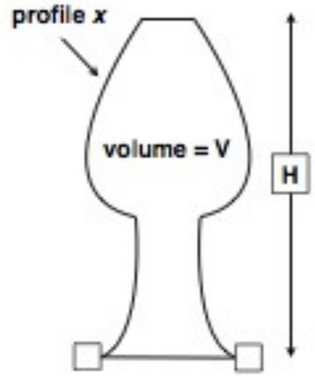
1. Choose a profile x
2. Rotate to generate a column $C(x)$

Designing an optimal column



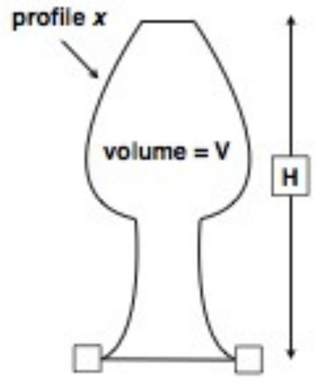
1. Choose a profile x
2. Rotate to generate a column $C(x)$
3. Respect the constraints on the volume, the height

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 $C(x)$

Designing an optimal column



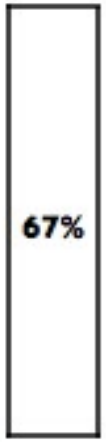
1. Choose a profile x
2. Rotate to generate a column $C(x)$
3. Respect the constraints on the volume, the height
4. Calculate (via Euler) the buckling strength $f(x)$ of the column $C(x)$
5. Maximize $f(x)$ over x

Three solutions

Three solutions

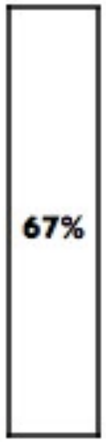
Lagrange
1770

Three solutions



Lagrange
1770

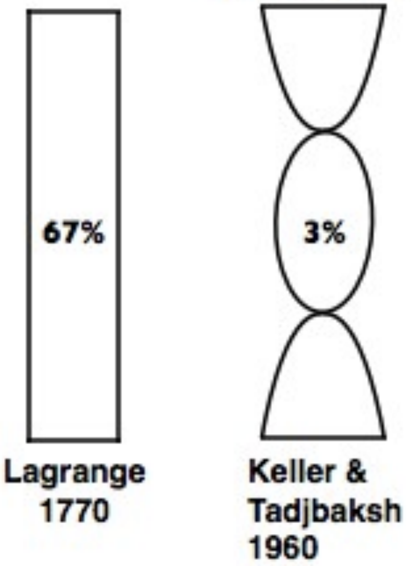
Three solutions



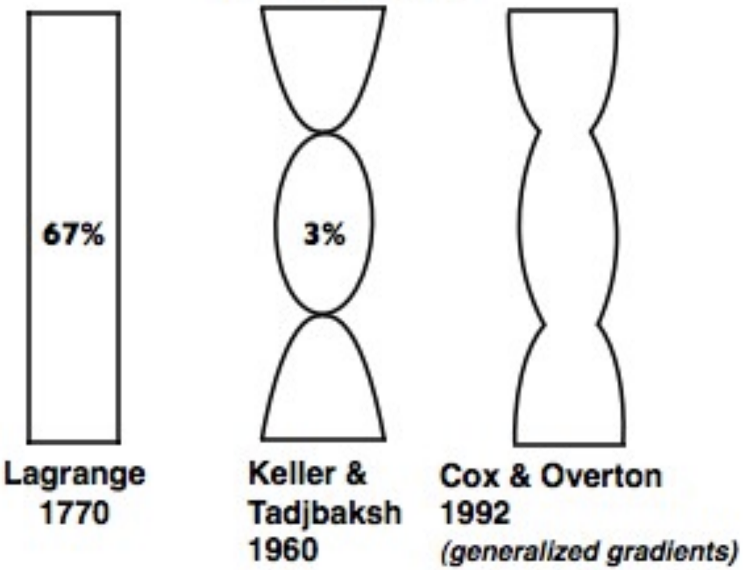
Lagrange
1770

Keller &
Tadjbaksh
1960

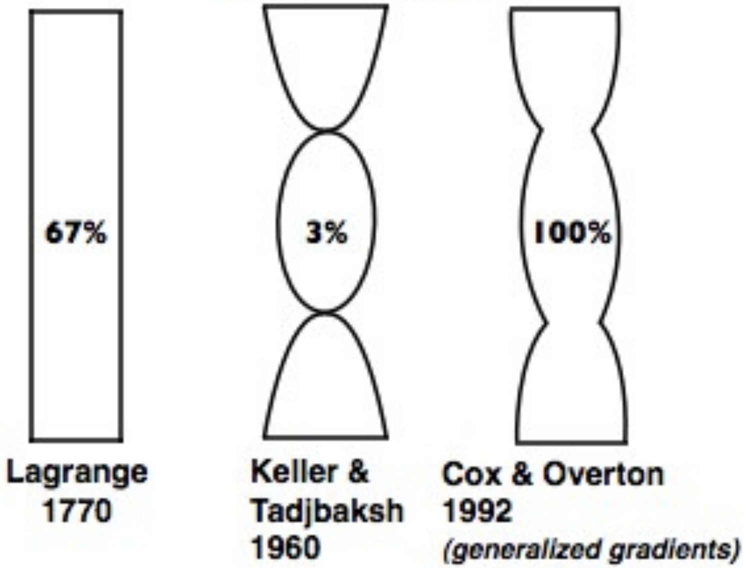
Three solutions



Three solutions



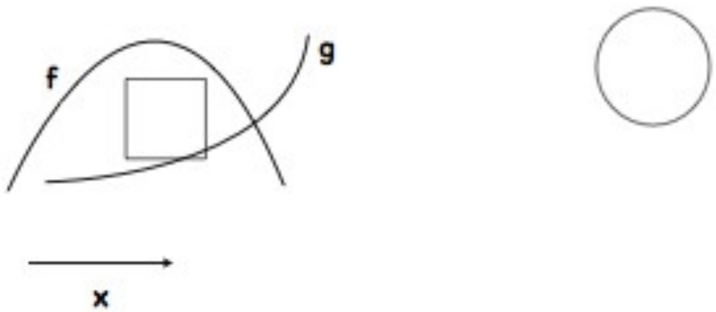
Three solutions



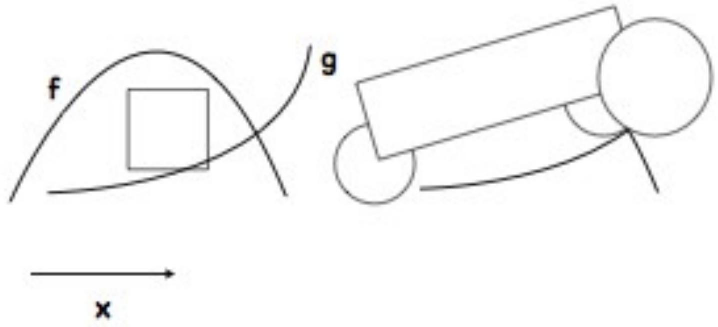
Envelopes of smooth functions are nonsmooth

Envelopes of smooth functions are nonsmooth

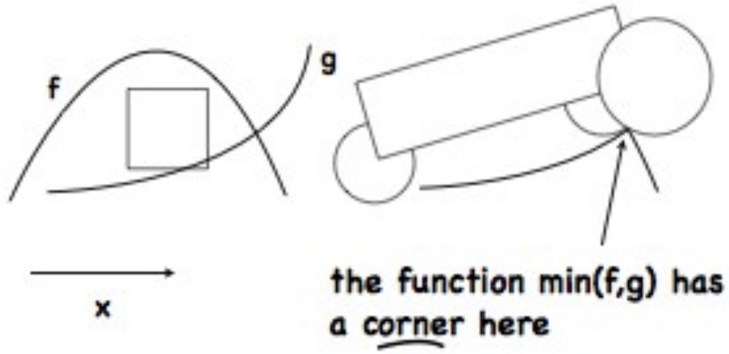
Envelopes of smooth functions are nonsmooth



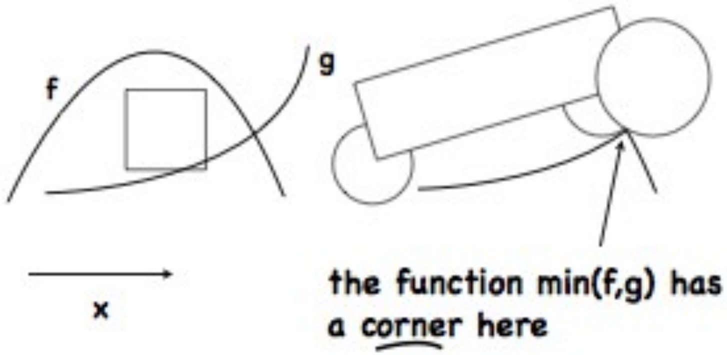
Envelopes of smooth functions are nonsmooth



Envelopes of smooth functions are nonsmooth



Envelopes of smooth functions are nonsmooth



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

Generalized gradients and proximal normals

Four definitions
Clarke 1973

Generalized gradients and proximal normals

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

Four definitions
Clarke 1973

Generalized gradients and proximal normals

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

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

Four definitions
Clarke 1973

Generalized gradients and proximal normals

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

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

Four definitions
Clarke 1973

$$\zeta \in N_S^P(x) \iff \begin{array}{l} \exists\, \sigma \geq 0 \text{ s.t.} \\ \langle \zeta, x' - x \rangle \leq \sigma |x' - x|^2 \; \forall x' \in S \end{array}$$

Generalized gradients and proximal normals

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

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

Four definitions
Clarke 1973

$$\zeta \in N_S^P(x) \iff \begin{array}{l} \exists\, \sigma \geq 0 \text{ s.t.} \\ \langle \zeta, x' - x \rangle \leq \sigma |x' - x|^2 \; \forall x' \in S \end{array}$$

S

Generalized gradients and proximal normals

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

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

Four definitions
Clarke 1973

$$\zeta \in N_S^P(x) \iff \begin{array}{l} \exists\, \sigma \geq 0 \text{ s.t.} \\ \langle \zeta, x' - x \rangle \leq \sigma |x' - x|^2 \; \forall x' \in S \end{array}$$

$\circ\, x$

S

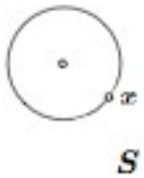
Generalized gradients and proximal normals

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

Four definitions
Clarke 1973

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

$$\zeta \in N_S^P(x) \iff \begin{array}{l} \exists\, \sigma \geq 0 \text{ s.t.} \\ \langle \zeta, x'-x\rangle \leq \sigma |x'-x|^2 \; \forall x' \in S \end{array}$$



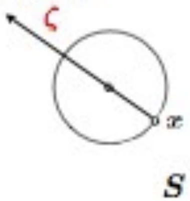
Generalized gradients and proximal normals

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

Four definitions
Clarke 1973

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

$$\zeta \in N_S^P(x) \iff \begin{array}{l} \exists\, \sigma \geq 0 \text{ s.t.} \\ \langle \zeta, x'-x\rangle \leq \sigma |x'-x|^2 \; \forall x' \in S \end{array}$$



Generalized gradients and proximal normals

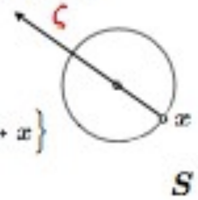
$$f^\circ(x;v) = \limsup_{t\downarrow 0, y\to x} \frac{f(y+tv)-f(y)}{t}$$

Four definitions
Clarke 1973

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

$$\zeta \in N_S^P(x) \iff \begin{array}{l} \exists\, \sigma \geq 0 \text{ s.t.} \\ \langle \zeta, x'-x\rangle \leq \sigma |x'-x|^2 \; \forall x' \in S \end{array}$$

$$N_S^C(x) = \operatorname{cl} \bigcup_{\lambda \geq 0} \lambda \partial_C d_S(x)$$
$$= \operatorname{cl}\operatorname{co}\left\{ \lim_{i\to\infty} \zeta_i : \zeta_i \in N_S^P(x_i), x_i \to x \right\}$$



Then

-
-
-
-
-
-

Then

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

Then

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

Then

- $\partial_C f(x)$ is compact convex nonempty
- $\partial_C(-f)(x) = -\partial_C f(x)$
- $\partial_C(f+g)(x) \subset \partial_C f(x) + \partial_C g(x)$

Then

- $\partial_C f(x)$ is compact convex nonempty
- $\partial_C(-f)(x) = -\partial_C f(x)$
- $\partial_C(f+g)(x) \subset \partial_C f(x) + \partial_C g(x)$
- $\partial_C \max\limits_{1 \le i \le n} f_i(x) \subset \dots$

Then

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

Then

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

Then

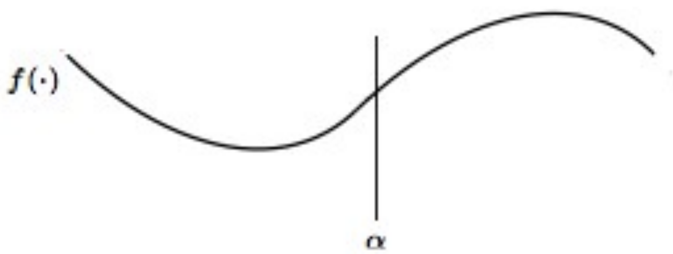
- $\partial_C f(x)$ is compact convex nonempty
- $\partial_C(-f)(x) = -\partial_C f(x)$
- $\partial_C(f+g)(x) \subset \partial_C f(x) + \partial_C g(x)$
- $\partial_C \max_{1\le i\le n} f_i(x) \subset \ldots$
- Mean value theorem
- Tangent vectors and normals to closed sets

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

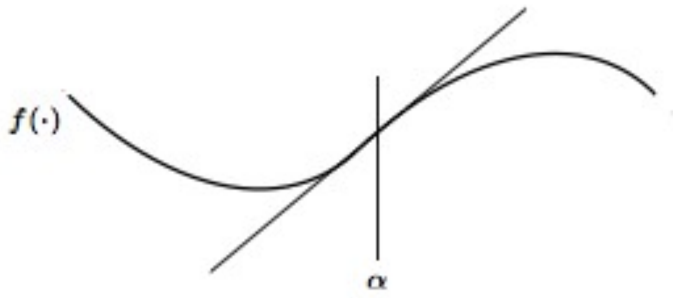
The classical derivative corresponds to a two-sided local approximation by an affine function.

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

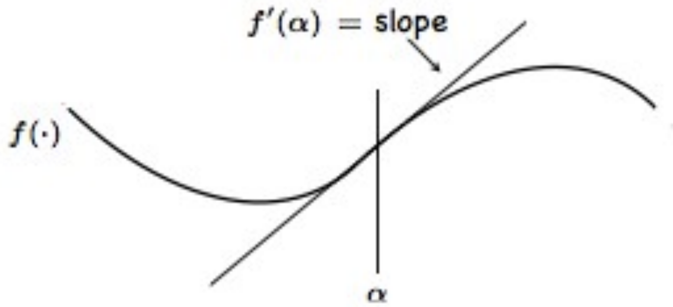
The classical derivative corresponds to a two-sided local approximation by an affine function.



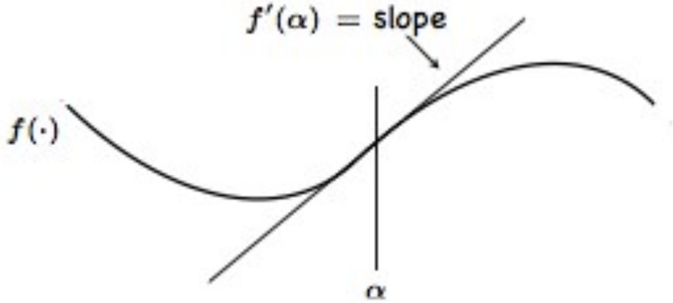
The classical derivative corresponds to a two-sided local approximation by an affine function.



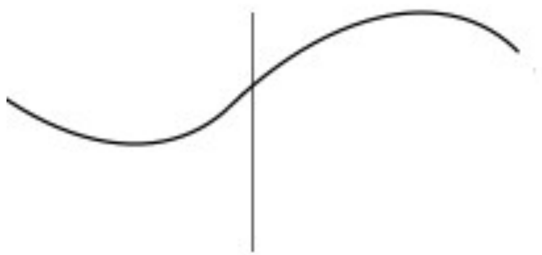
The classical derivative corresponds to a two-sided local approximation by an affine function.

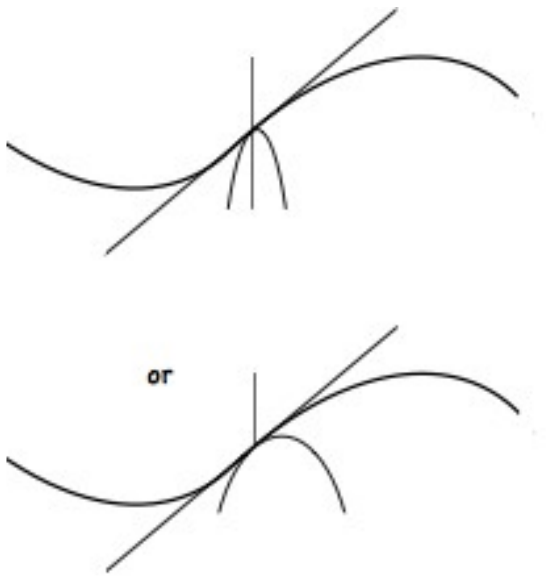


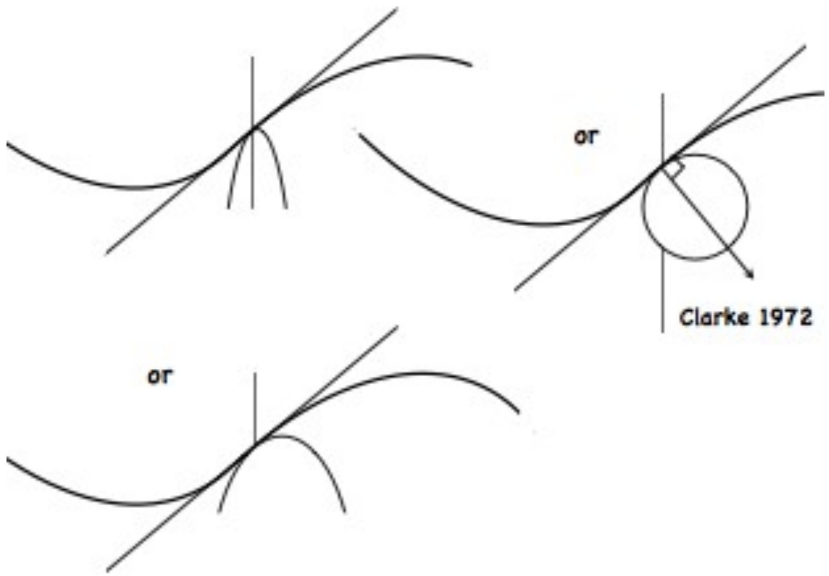
The classical derivative corresponds to a two-sided local approximation by an affine function.

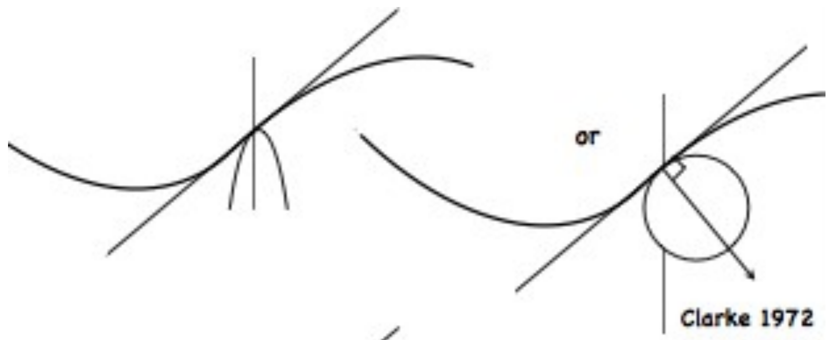


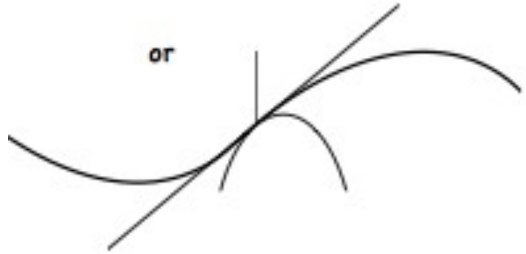
We may also approximate just from below, using nonlinear functions: proximal analysis











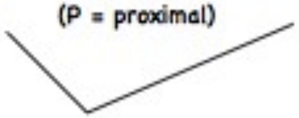
We can apply the `local lower-approximation by parabolas' idea to nonsmooth (lsc) functions

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

(P = proximal)

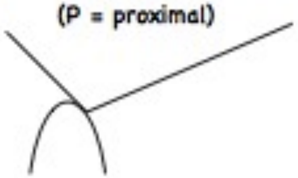
The set of all `contact slopes' of lower locally supporting parabolas is the proximal subdifferential $\partial_p f(\alpha)$

(P = proximal)



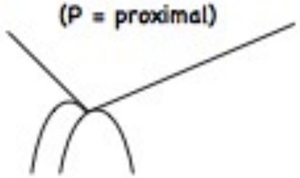
The set of all `contact slopes' of lower locally supporting parabolas is the proximal subdifferential $\partial_p f(\alpha)$

(P = proximal)



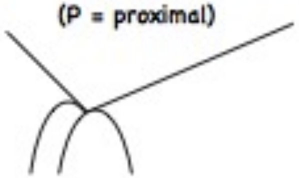
The set of all `contact slopes' of lower locally supporting parabolas is the proximal subdifferential $\partial_p f(\alpha)$

(P = proximal)



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

(P = proximal)



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

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

(P = proximal)

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

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

(P = proximal)

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

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

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

(P = proximal)

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

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

$\zeta \in \partial_P f(\alpha) \iff$

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

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

(P = proximal)

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

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

$\zeta \in \partial_P f(\alpha) \iff$

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

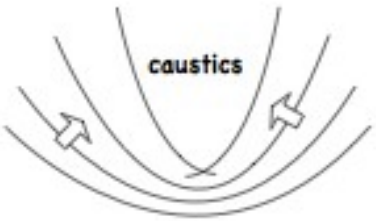
$\partial_P f$ has a very complete (but fuzzy!) theory and calculus...
Borwein, Ioffe, Ledyaev, Loewen, Rockafellar, Vinter, Zeidan...

The Hamilton–Jacobi equation: Various solution concepts

$$\phi_t(t,x) + H(t,x,\phi_x(t,x)) = 0 \qquad \text{(and bdry cdns)}$$

The Hamilton–Jacobi equation: Various solution concepts

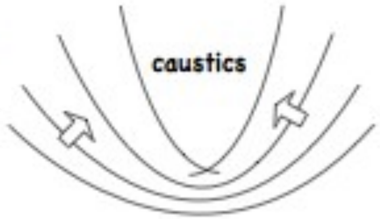
$$\phi_t(t,x) + H(t,x,\phi_x(t,x)) = 0 \qquad \text{(and bdry cdns)}$$



The Hamilton–Jacobi equation: Various solution concepts

$$\phi_t(t,x) + H(t,x,\phi_x(t,x)) = 0 \qquad \text{(and bdry cdns)}$$

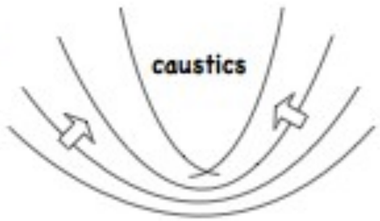
- **Classical**
  (φ smooth, pointwise equality)



The Hamilton–Jacobi equation: Various solution concepts

$$\phi_t(t,x) + H(t,x,\phi_x(t,x)) = 0 \qquad \text{(and bdry cdns)}$$

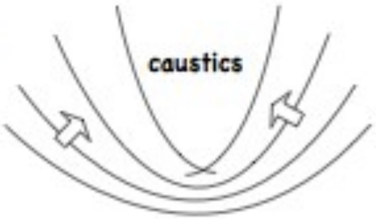
- **Classical**
  (φ smooth, pointwise equality)
- ***Almost everywhere* solutions**
  (φ Lipschitz)



The Hamilton–Jacobi equation: Various solution concepts

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

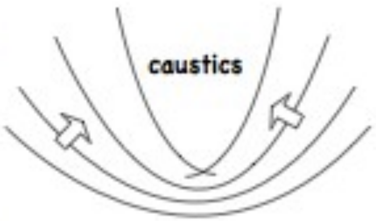
- **Classical**
  (φ smooth, pointwise equality)
- ***Almost everywhere*** **solutions**
  (φ Lipschitz)
- **Using generalized gradients**
  (Clarke 1977)



The Hamilton–Jacobi equation: Various solution concepts

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

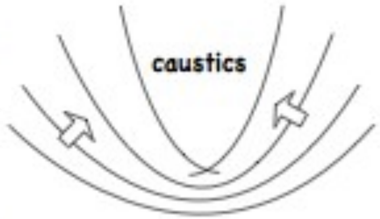
- **Classical**
  (φ smooth, pointwise equality)
- ***Almost everywhere*** **solutions**
  (φ Lipschitz)
- **Using generalized gradients**
  (Clarke 1977)
- **Using Dini derivates bilaterally**
  (Subbotin 1980)



The Hamilton–Jacobi equation: Various solution concepts

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

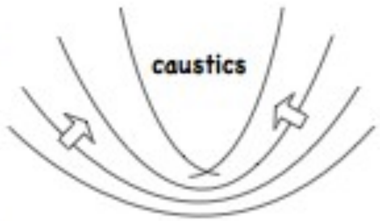
- **Classical**
  (φ smooth, pointwise equality)
- ***Almost everywhere*** **solutions**
  (φ Lipschitz)
- **Using generalized gradients**
  (Clarke 1977)
- **Using Dini derivates bilaterally**
  (Subbotin 1980)
- **Using sub- and superdifferentials**
  [*viscosity*] (Fleming, Crandall &
  Lions 1982,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



The Hamilton–Jacobi equation: Various solution concepts

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

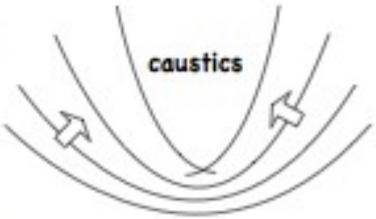
- **Classical**
  (φ smooth, pointwise equality)
- ***Almost everywhere*** **solutions**
  (φ Lipschitz)
- **Using generalized gradients**
  (Clarke 1977)
- **Using Dini derivates bilaterally**
  (Subbotin 1980)
- **Using sub- and superdifferentials**
  [*viscosity*] (Fleming, Crandall &
  Lions 1982,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



The Hamilton-Jacobi equation: Various solution concepts

$$\phi_t(t,x) + H(t,x,\phi_x(t,x)) = 0 \qquad \text{(and bdry cdns)}$$

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



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)

$$[\varphi'(x)]^2 - 1 = 0, \;\; \varphi(0) = \varphi(1) = 0$$

Example (n = 1)

$$[\varphi'(x)]^2 - 1 = 0, \;\; \varphi(0) = \varphi(1) = 0$$

- **No smooth solutions**

Example (n = 1)

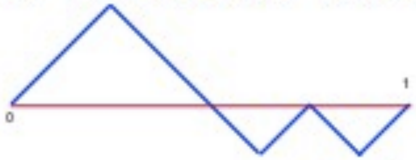
$$[\varphi'(x)]^2 - 1 = 0, \;\; \varphi(0) = \varphi(1) = 0$$

- **No smooth solutions**
- **Many "almost everywhere" solutions:**

Example (n = 1)

$$[\varphi'(x)]^2-1=0,\ \varphi(0)=\varphi(1)=0$$

- No smooth solutions
- Many "almost everywhere" solutions:



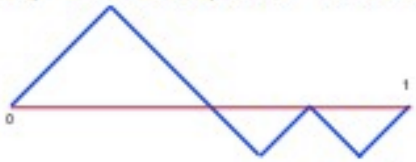
- A unique continuous φ satisfies $|\partial_P\varphi(x)|^2-1=0$

$(i.e. [\varsigma]^2-1=0\ \forall \varsigma\in\partial_P\varphi(x),\ x\in(0,1))$

Example (n = 1)

$$[\varphi'(x)]^2-1=0,\ \varphi(0)=\varphi(1)=0$$

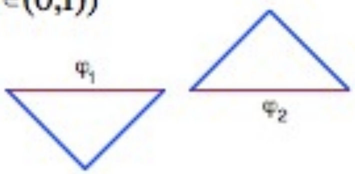
- No smooth solutions
- Many "almost everywhere" solutions:



- A unique continuous φ satisfies $|\partial_P\varphi(x)|^2-1=0$

$(i.e. [\varsigma]^2-1=0\ \forall \varsigma\in\partial_P\varphi(x),\ x\in(0,1))$

Hint: it is one of these two functions:



Optimal control : an example in bioeconomics

(Clark, Clarke, Munro / Econometrica)

$$x'(t)=G(x(t))-u(t)\,x(t)$$
$$\max\int_0^\infty e^{-\delta t}\{\pi\,x(t)-k\}\,u(t)\,dt$$
$$0\le u(t)\le E$$

x = biomass
u = fishing effort
G = natural growth
E = maximum fishing effort
δ = discount rate
π = resource price
k = effort cost

Optimal control : an example in bioeconomics

(Clark, Clarke, Munro / Econometrica)

$$x'(t)=G(x(t))-u(t)\,x(t)$$
$$\max\int_0^\infty e^{-\delta t}\{\pi\,x(t)-k\}\,u(t)\,dt$$
$$0\le u(t)\le E$$

x = biomass
u = fishing effort
G = natural growth
E = maximum fishing effort
δ = discount rate
π = resource price
k = effort cost

Optimal control : an example in bioeconomics

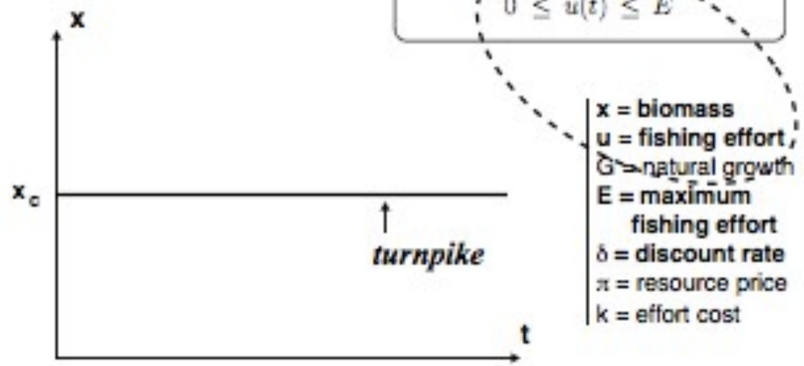
(Clark, Clarke, Munro / Econometrica)

$$x'(t) = G(x(t)) - u(t)\,x(t)$$
$$\max \int_0^\infty e^{-\delta t}\{\pi x(t) - k\}\,u(t)\,dt$$
$$0 \le u(t) \le E$$

Optimal control : an example in bioeconomics

(Clark, Clarke, Munro / Econometrica)

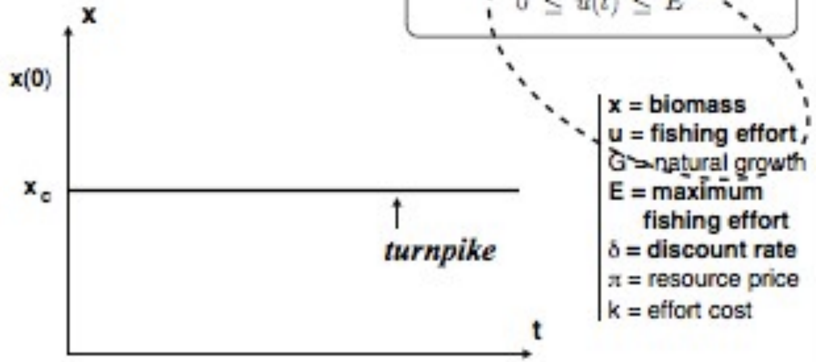
$$x'(t) = G(x(t)) - u(t)\,x(t)$$
$$\max \int_0^\infty e^{-\delta t}\{\pi x(t) - k\}\,u(t)\,dt$$
$$0 \le u(t) \le E$$



Optimal control : an example in bioeconomics

(Clark, Clarke, Munro / Econometrica)

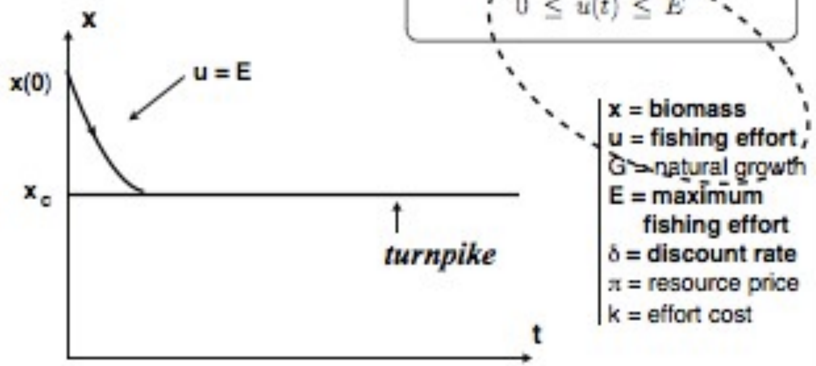
$$x'(t) = G(x(t)) - u(t)\,x(t)$$
$$\max \int_0^\infty e^{-\delta t}\{\pi x(t) - k\}\,u(t)\,dt$$
$$0 \le u(t) \le E$$



Optimal control : an example in bioeconomics

(Clark, Clarke, Munro / Econometrica)

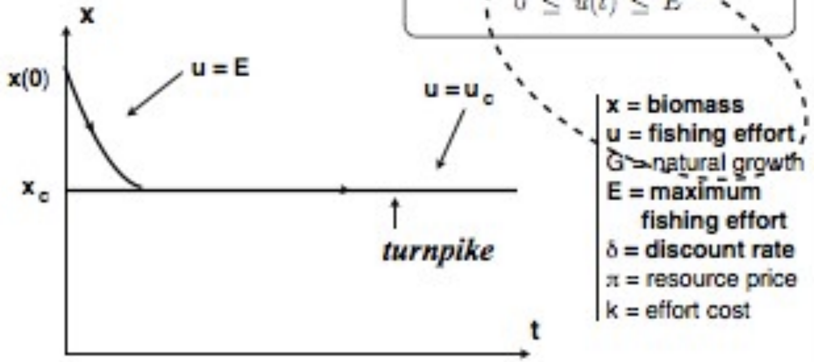
$$x'(t) = G(x(t)) - u(t)\,x(t)$$
$$\max \int_0^\infty e^{-\delta t}\{\pi x(t) - k\}\,u(t)\,dt$$
$$0 \le u(t) \le E$$



Optimal control : an example in bioeconomics

(Clark, Clarke, Munro / Econometrica)

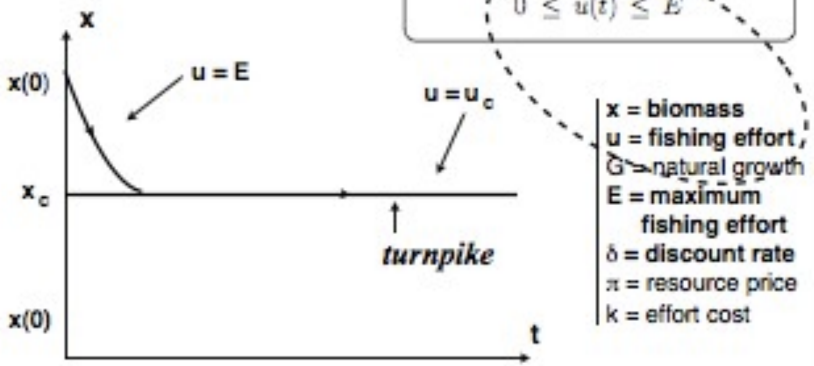
$$x'(t) = G(x(t)) - u(t)\,x(t)$$
$$\max \int_0^\infty e^{-\delta t}\{\pi x(t) - k\}\,u(t)\,dt$$
$$0 \le u(t) \le E$$



Optimal control : an example in bioeconomics

(Clark, Clarke, Munro / Econometrica)

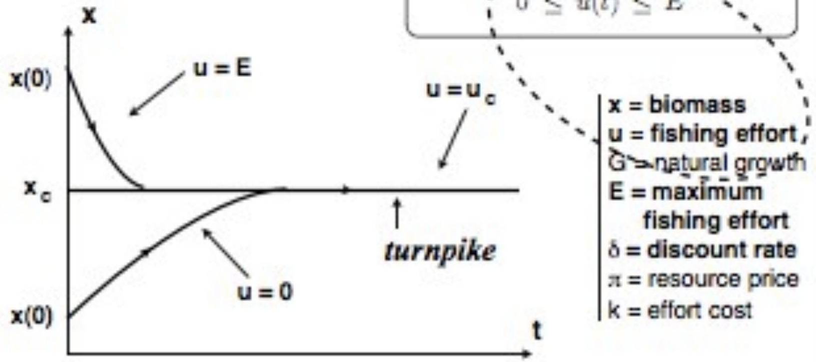
$$x'(t) = G(x(t)) - u(t)\,x(t)$$
$$\max \int_0^\infty e^{-\delta t}\{\pi x(t) - k\}\,u(t)\,dt$$
$$0 \le u(t) \le E$$



Optimal control : an example in bioeconomics

(Clark, Clarke, Munro / Econometrica)

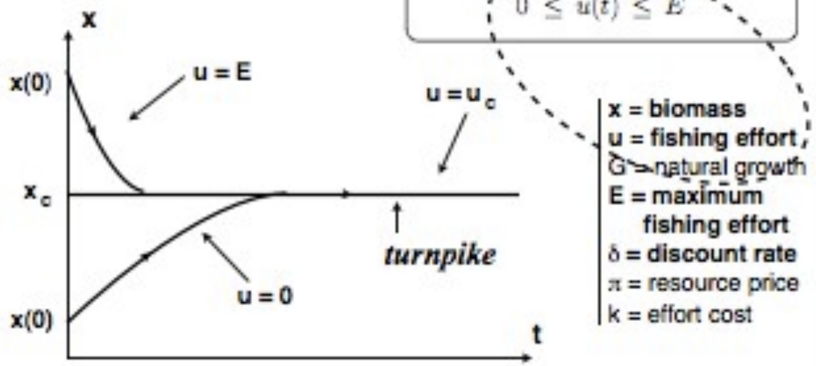
$$x'(t) = G(x(t)) - u(t)\,x(t)$$
$$\max \int_0^\infty e^{-\delta t}\{\pi x(t) - k\}\,u(t)\,dt$$
$$0 \le u(t) \le E$$



Optimal control : an example in bioeconomics

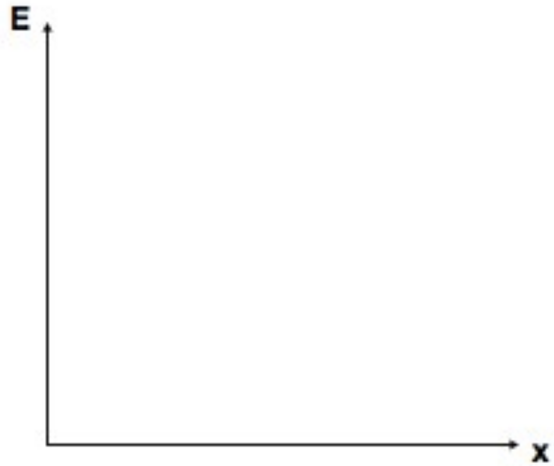
(Clark, Clarke, Munro / Econometrica)

$$x'(t) = G(x(t)) - u(t)\,x(t)$$
$$\max \int_0^\infty e^{-\delta t}\{\pi x(t) - k\}\,u(t)\,dt$$
$$0 \le u(t) \le E$$



If δ is sufficiently large, we have $x_c = 0$ (extinction)

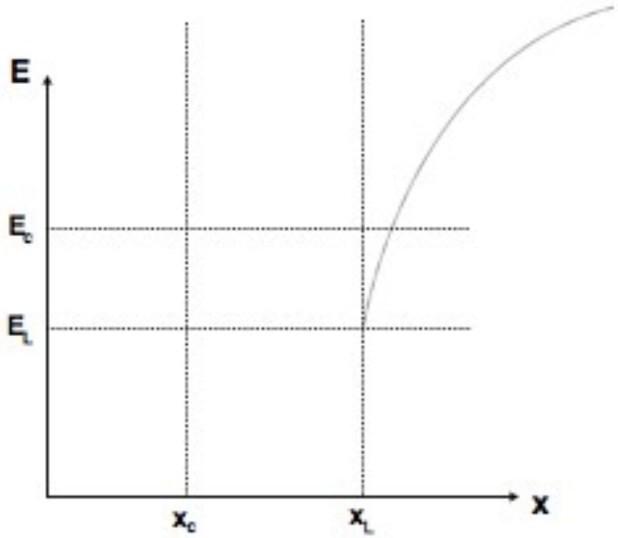
Example: Optimal fishing strategy in the presence of both investment and depreciation in boats
(Clark, Clarke, Munro / Econometrica)

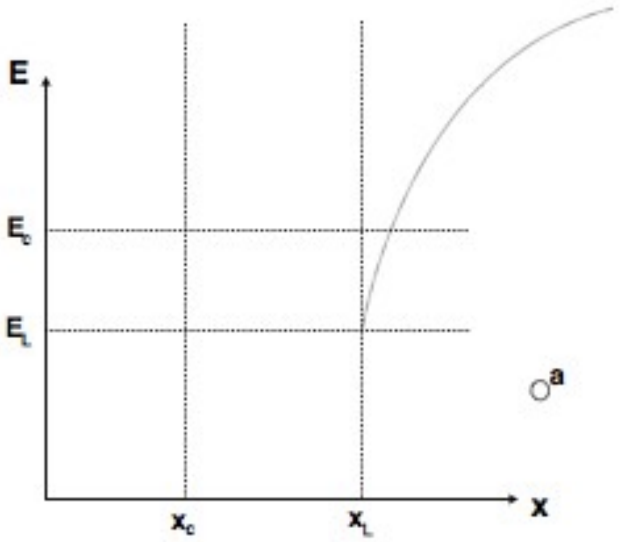


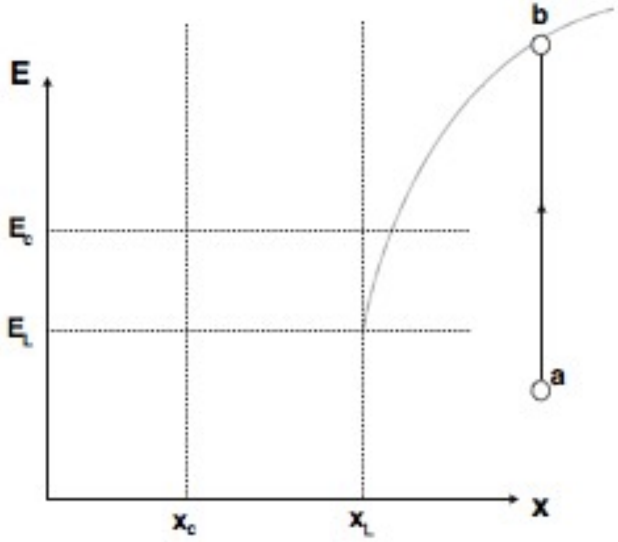
Example: Optimal fishing strategy in the presence of both investment and depreciation in boats
(Clark, Clarke, Munro / Econometrica)

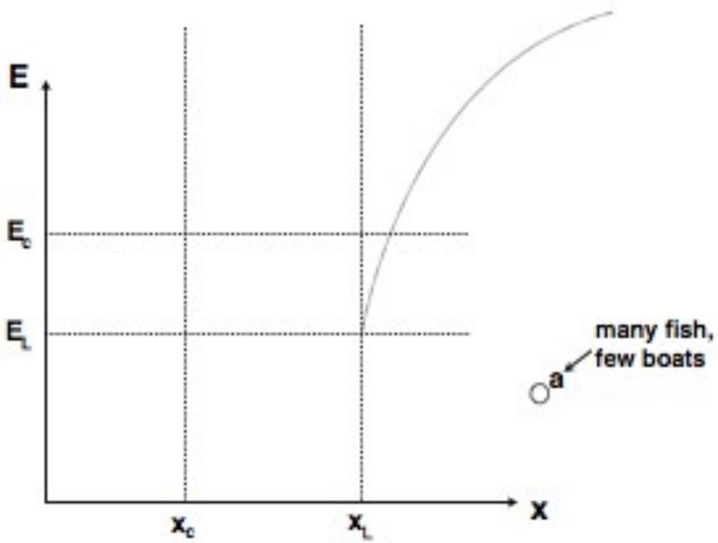
$$\max \int_0^\infty e^{-\delta t}\left\{(\pi x(t) - k)u(t) - cI(t)\right\} dt + \sum_i e^{-\delta t_i} \Delta E(t_i)$$

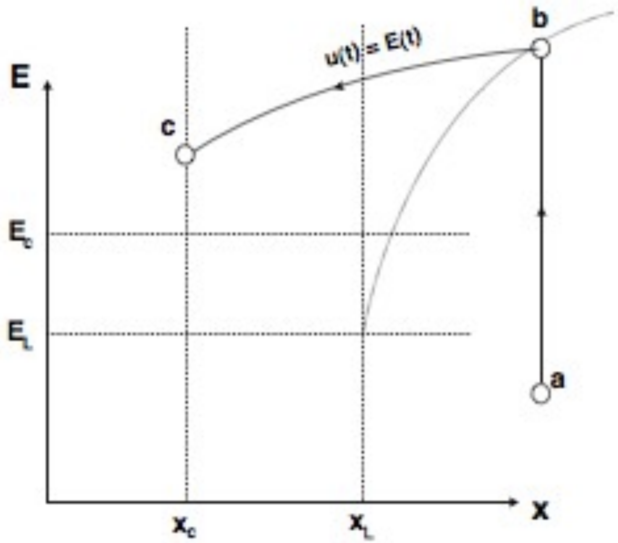
$$x'(t) = g(x(t)) - u(t)x(t),\ 0 \le u(t) \le E(t)$$
$$E'(t) = -\gamma E(t) + I(t),\ 0 \le I(t) \le +\infty$$

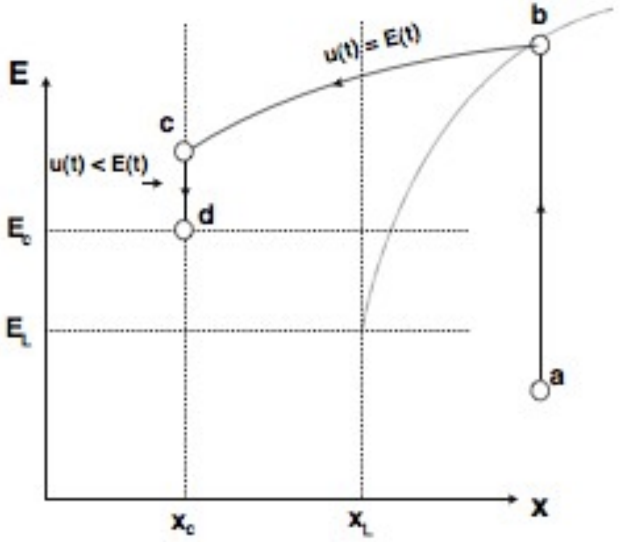


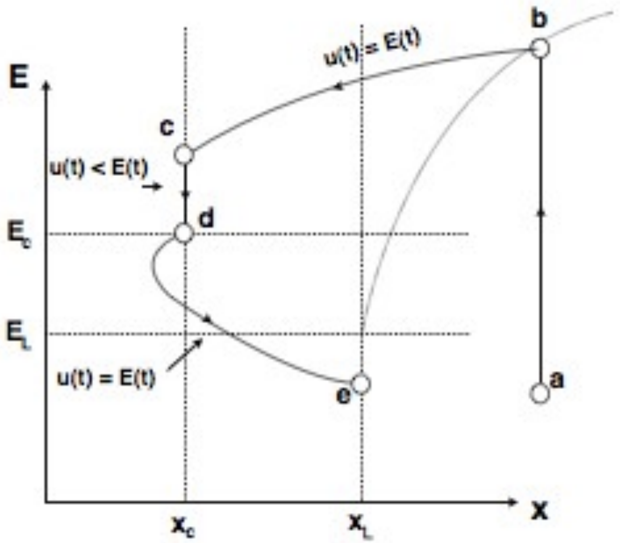

E
E_c
E_L
a
x_c
x_L
x

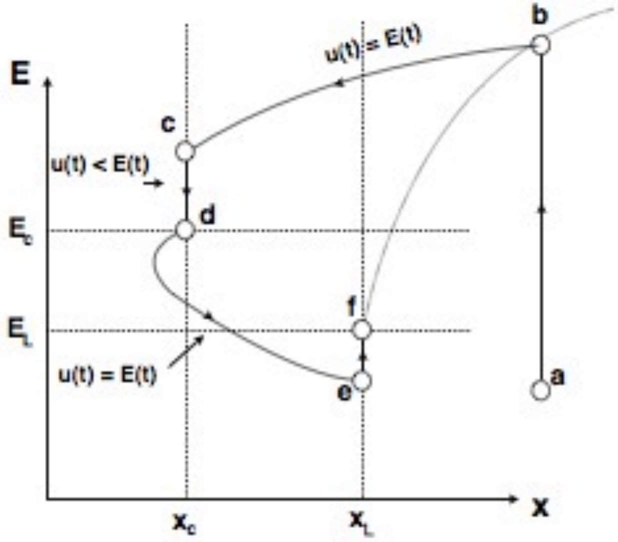

E
b
E_c
E_L
a
x_c
x_L
x

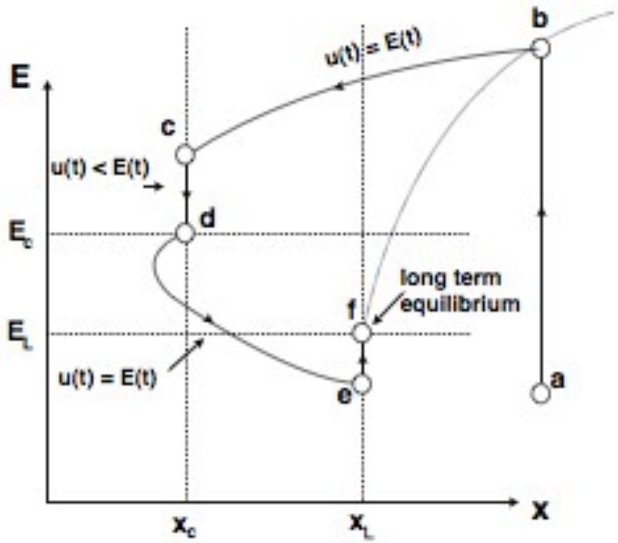

E
E_c
E_L
many fish,
few boats
a
x_c
x_L
x

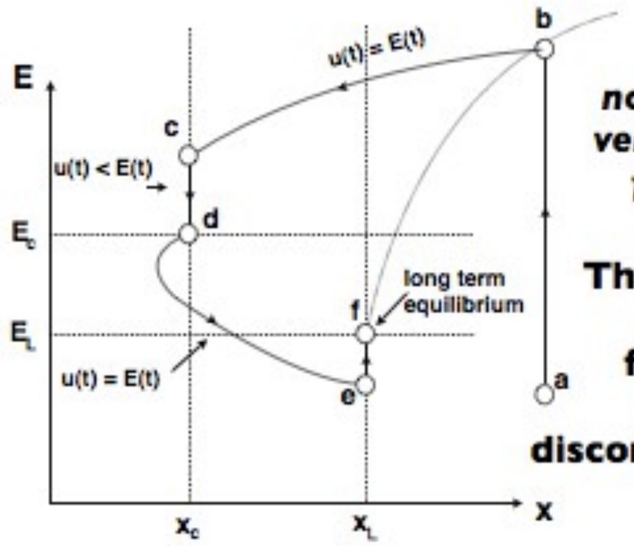

E
b
u(t) = E(t)
c
E_c
E_L
a
x_c
x_L
x


E
b
u(t) = E(t)
c
u(t) < E(t)
d
E_c
E_L
a
x_c
x_L
x


E
b
u(t) = E(t)
c
u(t) < E(t)
d
E_c
E_L
u(t) = E(t)
e
a
x_c
x_L
x


E
b
u(t) = E(t)
c
u(t) < E(t)
d
E_c
f
E_L
u(t) = E(t)
e
a
x_c
x_L
x


E
b
u(t) = E(t)
c
u(t) < E(t)
d
E_c
long term
equilibrium
f
E_L
u(t) = E(t)
e
a
x_c
x_L
x



Proof:
nonsmooth verification functions

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

Verification functions

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

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

Goal: verify that a candidate (x_*, u_*) is optimal

$$\min J(x,u) := \int_a^b L(t,x(t),u(t))\,dt \qquad x(a)=A$$

$$x'(t) = f(t,x(t),u(t)),\ u(t)\in U(t) \text{ a.e.} \qquad x(b)=B$$

Goal: verify that a candidate (x_*,u_*) is optimal

Method: exhibit a function ϕ satisfying

$$L(t,x,u) \ \ge\ \phi_t(t,x) + \langle \phi_x(t,x), f(t,x,u)\rangle \ \forall (t,x), u\in U(t)$$

$$(= \text{at } (t,x_*(t),u_*(t)))$$

$$\min J(x,u) := \int_a^b L(t,x(t),u(t))\,dt \qquad x(a)=A$$

$$x'(t) = f(t,x(t),u(t)),\ u(t)\in U(t) \text{ a.e.} \qquad x(b)=B$$

Goal: verify that a candidate (x_*,u_*) is optimal

Method: exhibit a function ϕ satisfying

$$L(t,x,u) \ \ge\ \phi_t(t,x) + \langle \phi_x(t,x), f(t,x,u)\rangle \ \forall (t,x), u\in U(t)$$

$$(= \text{at } (t,x_*(t),u_*(t)))$$

Proof: For any admissible (x,u) we have

$$L(t,x(t),u(t)) \ge \phi_t(t,x(t)) + \langle \phi_x(t,x(t)), f(t,x(t),u(t))\rangle$$

$$= d/dt\,\{\phi(t,x(t))\}$$

$$\min J(x,u) := \int_a^b L(t,x(t),u(t))\,dt \qquad x(a)=A$$

$$x'(t) = f(t,x(t),u(t)),\ u(t)\in U(t) \text{ a.e.} \qquad x(b)=B$$

Goal: verify that a candidate (x_*,u_*) is optimal

Method: exhibit a function ϕ satisfying

$$L(t,x,u) \ \ge\ \phi_t(t,x) + \langle \phi_x(t,x), f(t,x,u)\rangle \ \forall (t,x), u\in U(t)$$

$$(= \text{at } (t,x_*(t),u_*(t)))$$

Proof: For any admissible (x,u) we have

$$L(t,x(t),u(t)) \ge \phi_t(t,x(t)) + \langle \phi_x(t,x(t)), f(t,x(t),u(t))\rangle$$

$$= d/dt\,\{\phi(t,x(t))\}$$

$$\Longrightarrow J(x,u) \ge \phi(b,B) - \phi(a,A)$$

$$\min J(x,u) := \int_a^b L(t,x(t),u(t))\,dt \qquad x(a)=A$$

$$x'(t) = f(t,x(t),u(t)),\ u(t)\in U(t) \text{ a.e.} \qquad x(b)=B$$

Goal: verify that a candidate (x_*,u_*) is optimal

Method: exhibit a function ϕ satisfying

$$L(t,x,u) \ \ge\ \phi_t(t,x) + \langle \phi_x(t,x), f(t,x,u)\rangle \ \forall (t,x), u\in U(t)$$

$$(= \text{at } (t,x_*(t),u_*(t)))$$

Proof: For any admissible (x,u) we have

$$L(t,x(t),u(t)) \ge \phi_t(t,x(t)) + \langle \phi_x(t,x(t)), f(t,x(t),u(t))\rangle$$

$$= d/dt\,\{\phi(t,x(t))\}$$

$$\Longrightarrow J(x,u) \ge \phi(b,B) - \phi(a,A)$$

$$(= \text{for } (x,u) = (x_*,u_*))$$ QED

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

Goal: verify that a candidate (x_*, u_*) is optimal

Method: exhibit a function ϕ satisfying

$$L(t,x,u) \;\geq\; \phi_t(t,x) + \langle \phi_x(t,x), f(t,x,u)\rangle \;\forall (t,x), u \in U(t)$$
$$(= \text{at } (t, x_*(t), u_*(t)))$$

Proof: For any admissible (x,u) we have

$$L(t,x(t),u(t)) \geq \phi_t(t,x(t)) + \langle \phi_x(t,x(t)), f(t,x(t),u(t))\rangle$$
$$= d/dt\,\{\phi(t,x(t))\}$$
$$\Longrightarrow J(x,u) \geq \phi(b,B) - \phi(a,A)$$
$$(= \text{for } (x,u) = (x_*,u_*)) \quad \text{QED}$$

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

clarke@math.univ-lyon1.fr

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

Generalized Gradients and Proximal analysis

Francis Clarke

Institut universitaire de France et Université de Lyon

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 <u>functions</u> is the generalized directional derivative:

Generalized gradients and associated geometry

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

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

Generalized gradients and associated geometry

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

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

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

$$\begin{aligned} f^\circ(x; v+w) &\le f^\circ(x;v) + f^\circ(x;w)\ \forall v, w \\ f^\circ(x; tv) &= t f^\circ(x;v)\ \forall t \ge 0 \end{aligned}$$

Generalized gradients and associated geometry

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

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

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

$$f^\circ(x;v+w) \;\le\; f^\circ(x;v) + f^\circ(x;w)\ \forall v, w$$
$$f^\circ(x;tv) \;=\; t f^\circ(x;v)\ \forall t \ge 0$$

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

$$f^\circ(x;v) \;=\; H_Z(v)\ \forall v \in X$$

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

$$f^\circ(x;v) \;=\; H_Z(v)\ \forall v \in X$$

We denote this set by $\partial_C f(x)$, the generalized gradient. The following duality holds:

$$\partial_C f(x) = \left\{\zeta \in X^* : f^\circ(x;v) \ge \langle \zeta, v\rangle\ \forall v\right\}$$
$$f^\circ(x;v) \;=\; \max_{\zeta\in\partial_C f(x)} \langle \zeta, v\rangle$$

$\partial_C f(x)$ 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

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

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

$$N_S^C(x) := \mathrm{cl}\,\{\, t\,\partial_C d_S(x) : t \geq 0\}$$

$$\begin{aligned} T_S^C(x) &= \left[N_S^C(x)\right]^\circ \\ &:= \{v : \langle \zeta, v\rangle \leq 0 \;\; \forall \zeta \in N_S^C(x)\} \\ &= \{v : d_S^\circ(x; v) = 0\} \end{aligned}$$

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

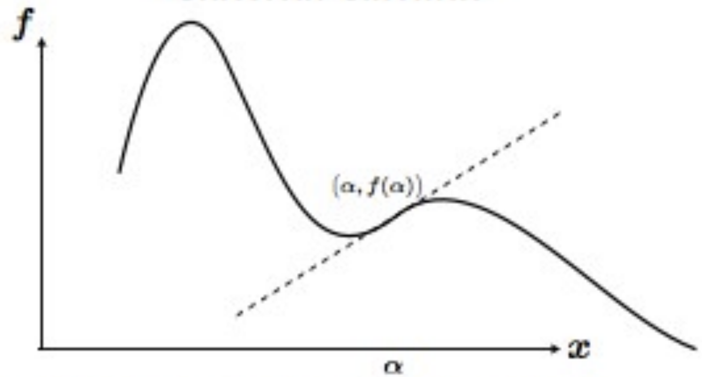
$$\begin{aligned} T_S^C(x) := \{v : \forall\, x_i \to_S x,\ \forall\, t_i \downarrow 0, \\ \exists\, v_i \to v \;/\; x_i + t_i v_i \in S\} \end{aligned}$$

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

$$T_S^C(x) := \{v : \forall\, x_i \to_S x,\ \forall\, t_i \downarrow 0,\ \exists\, v_i \to v \,/\, x_i + t_i v_i \in S\}$$

$$N_S^C(x) := \left[T_S^C(x)\right]^\circ = \{\zeta : \langle \zeta, v\rangle \le 0\ \forall v \in T_S^C(x)\}$$

Classical Calculus



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

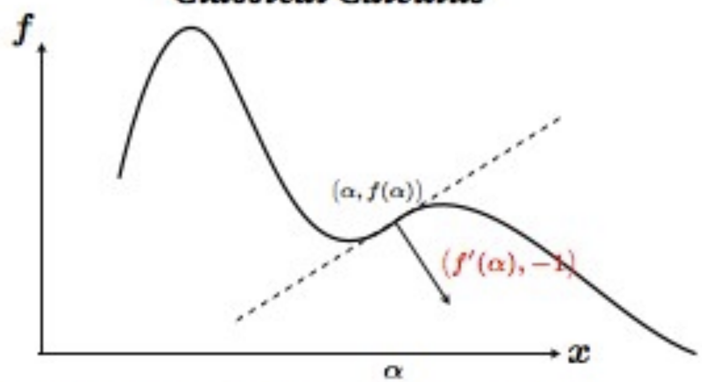
If we wish to start with geometry, the last shall be first:

$$T_S^C(x) := \{v : \forall\, x_i \to_S x,\ \forall\, t_i \downarrow 0,\ \exists\, v_i \to v \,/\, x_i + t_i v_i \in S\}$$

$$N_S^C(x) := \left[T_S^C(x)\right]^\circ = \{\zeta : \langle \zeta, v\rangle \le 0\ \forall v \in T_S^C(x)\}$$

How do we recover the functional constructs?

Classical Calculus



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

Dually, the value ζ such that $(\zeta, -1)$ is normal to the graph of f

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

$$T_S^C(x) := \{v : \forall\, x_i \to_S x,\ \forall\, t_i \downarrow 0,\ \exists\, v_i \to v \;/\; x_i + t_i v_i \in S\}$$

$$N_S^C(x) := \left[T_S^C(x)\right]^\circ = \{\zeta : \langle \zeta, v\rangle \le 0\ \ \forall v \in T_S^C(x)\}$$

How do we recover the functional constructs?

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

$$T_S^C(x) := \{v : \forall\, x_i \to_S x,\ \forall\, t_i \downarrow 0,\ \exists\, v_i \to v \;/\; x_i + t_i v_i \in S\}$$

$$N_S^C(x) := \left[T_S^C(x)\right]^\circ = \{\zeta : \langle \zeta, v\rangle \le 0\ \ \forall v \in T_S^C(x)\}$$

How do we recover the functional constructs?

$$\partial_C f(x) := \left\{\zeta : (\zeta, -1) \in N^C_{\mathrm{epi}\, f}(x, f(x))\right\}$$

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

$$T_S^C(x) := \{v : \forall\, x_i \to_S x,\ \forall\, t_i \downarrow 0,\ \exists\, v_i \to v \;/\; x_i + t_i v_i \in S\}$$

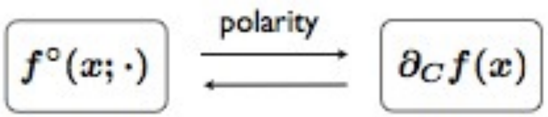
$$N_S^C(x) := \left[T_S^C(x)\right]^\circ = \{\zeta : \langle \zeta, v\rangle \le 0\ \ \forall v \in T_S^C(x)\}$$

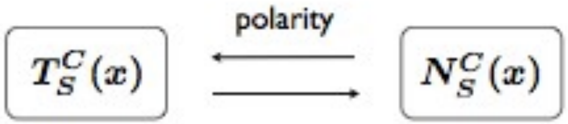
How do we recover the functional constructs?

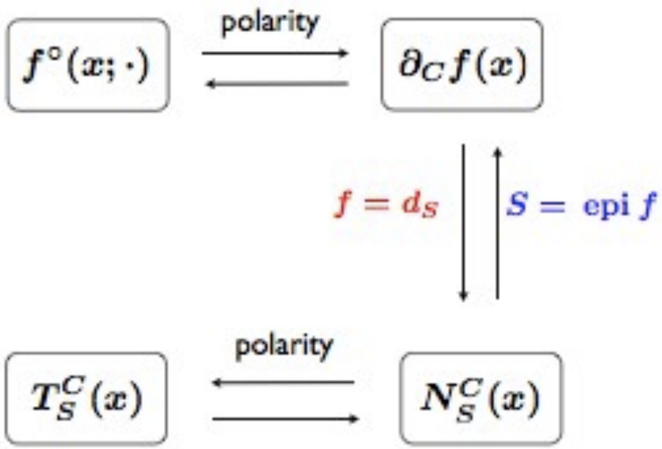
$$\partial_C f(x) := \left\{\zeta : (\zeta, -1) \in N^C_{\mathrm{epi}\, f}(x, f(x))\right\}$$

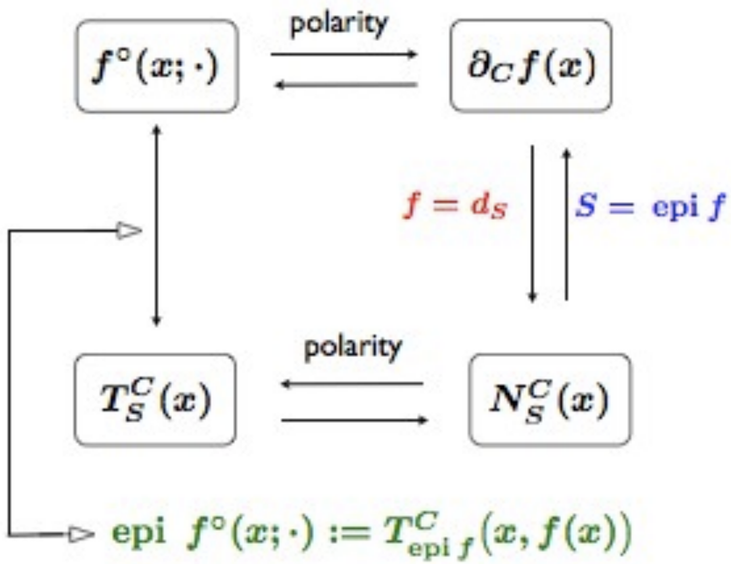
(and then $f^\circ(x;\cdot)$ is the support function of $\partial_C f(x)$)

$f^\circ(x;\cdot)$	$\partial_C f(x)$
$T_S^C(x)$	$N_S^C(x)$









The smooth case

The smooth case

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

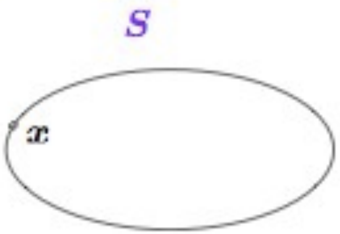
$$\langle f'(x), v\rangle = f'(x;v) = f^\circ(x;v) = \max_{\zeta\in\partial_C f(x)} \langle \zeta, v\rangle$$

The smooth case

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

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

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

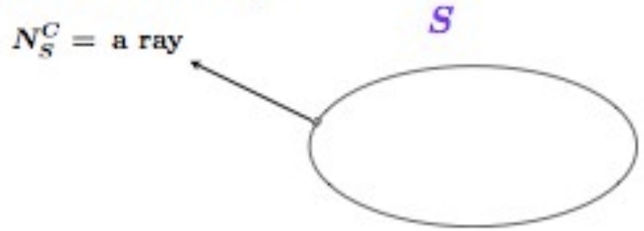


The smooth case

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

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

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

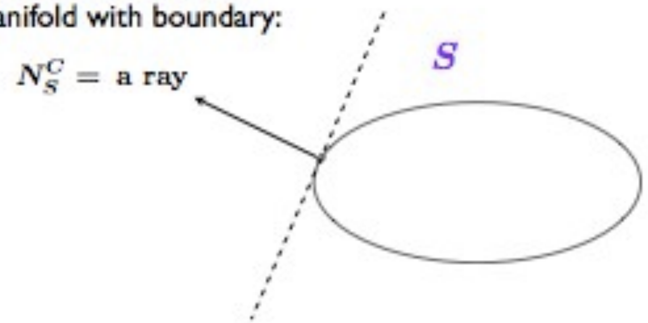


The smooth case

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

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

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

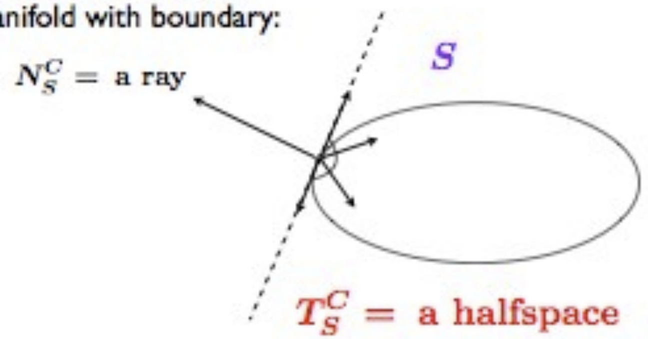


The smooth case

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

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

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



The convex case

The convex case

If f is convex, then

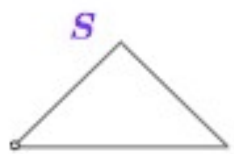
$$\partial_C f(x) = \partial f(x) \quad \text{the subdifferential}$$
$$= \{\zeta : f(y) - f(x) \geq \langle \zeta, y - x\rangle \;\forall\, y \in X\}$$

The convex case

If f is convex, then

$$\partial_C f(x) = \partial f(x) \quad \text{the subdifferential}$$
$$= \{\zeta : f(y) - f(x) \geq \langle \zeta, y - x\rangle \;\forall\, y \in X\}$$

If S is convex, then



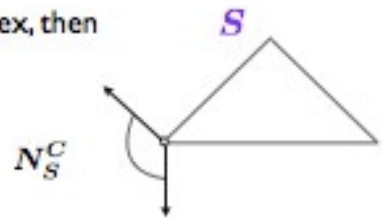
The convex case

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



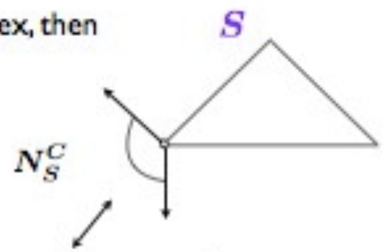
The convex case

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



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

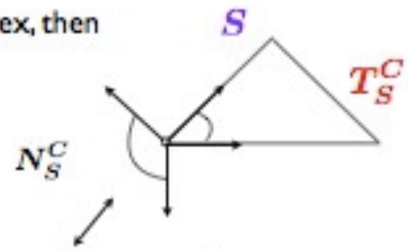
The convex case

If f is convex, then

$$\partial_C f(x) = \partial f(x)$$ *the subdifferential*

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

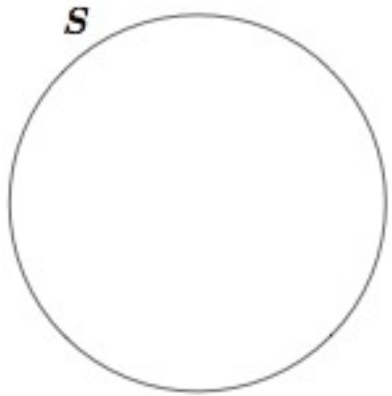
If S is convex, then



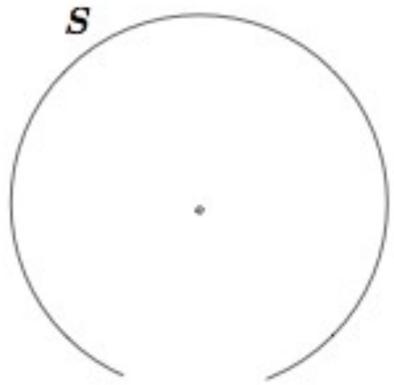
$$\zeta \in N_S(x) \iff \langle \zeta, y - x \rangle \leq 0 \; \forall \, y \in S$$

An example which is neither smooth nor convex

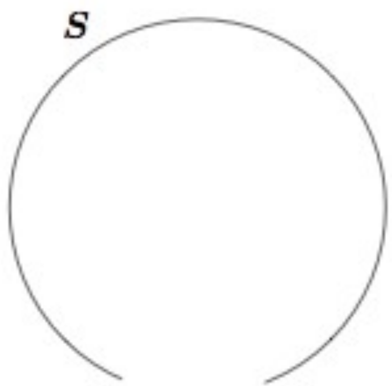
An example which is neither smooth nor convex



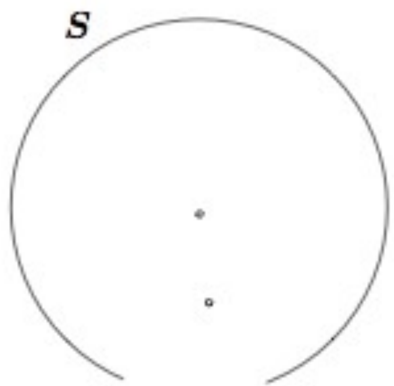
An example which is neither smooth nor convex



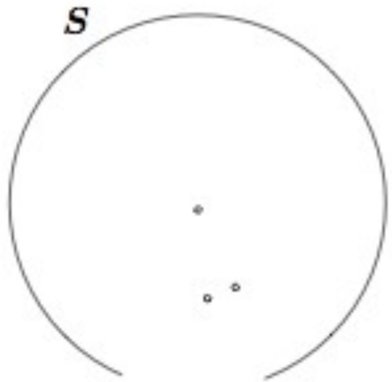
An example which is neither smooth nor convex



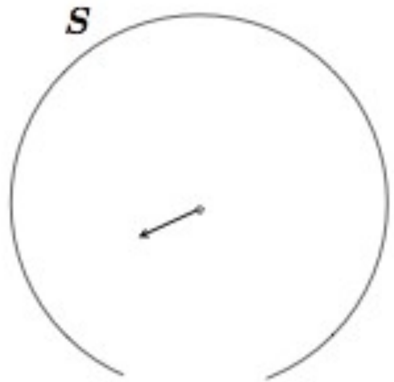
An example which is neither smooth nor convex



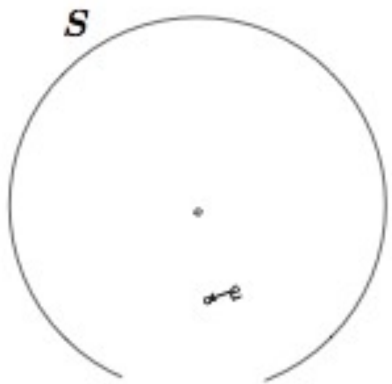
An example which is neither smooth nor convex



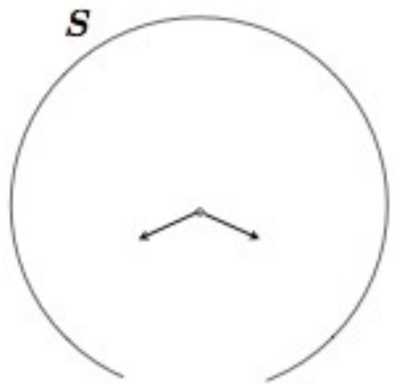
An example which is neither smooth nor convex



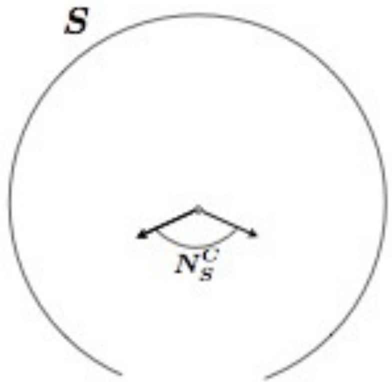
An example which is neither smooth nor convex



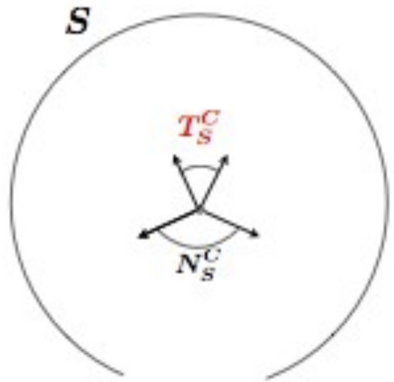
An example which is neither smooth nor convex



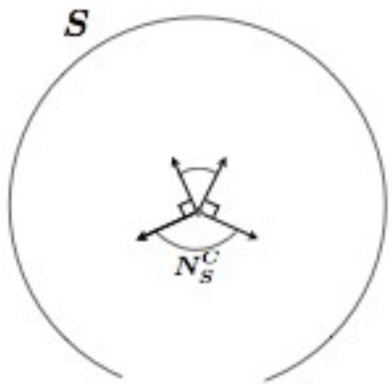
An example which is neither smooth nor convex



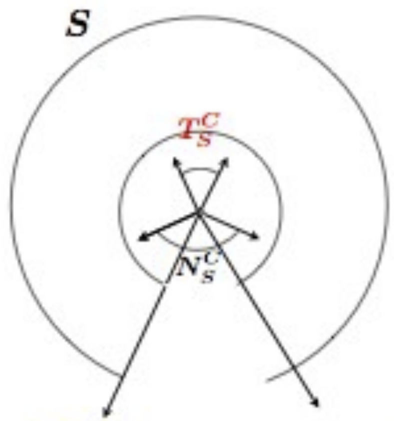
An example which is neither smooth nor convex



An example which is neither smooth nor convex



An example which is neither smooth nor convex



$$T_S(x) := \{ \lim_{i \to \infty} \frac{x_i - x}{t_i} : x_i \to_S x,\ t_i \downarrow 0 \}$$

Bouligand contingent cone

Some calculus

Some calculus

Sums:

$$\partial_C(f_1 + f_2)(x) \subset \partial_C f_1(x) + \partial_C f_2(x)$$

(equality when f_1, f_2 ***regular***)

$f^\circ = f'$

Some calculus

Sums:

$$\partial_C(f_1 + f_2)(x) \subset \partial_C f_1(x) + \partial_C f_2(x)$$

(equality when f_1, f_2 ***regular***)

Some calculus

Sums:

$$\partial_C(f_1 + f_2)(x) \subset \partial_C f_1(x) + \partial_C f_2(x)$$

(equality when f_1, f_2 ***regular***)

Some calculus

Sums:

$$\partial_C(f_1 + f_2)(x) \subset \partial_C f_1(x) + \partial_C f_2(x)$$

(equality when f_1, f_2 ***regular***)

Mean value theorem:

$$\exists\, z \in (x,y) \,/\, f(y) - f(x) \in \langle \zeta, y - x \rangle$$

Some calculus

Sums:

$$\partial_C(f_1 + f_2)(x) \subset \partial_C f_1(x) + \partial_C f_2(x)$$

(equality when f_1, f_2 ***regular***)

Mean value theorem:

$$\exists\, z \in (x,y) \,/\, f(y) - f(x) \in \langle \zeta, y - x \rangle$$

Maximum functions: $f(x) = \max_{1 \le i \le n} f_i(x)$ (each f_i smooth)

$$I(x) = \{ i \in \{1, 2, \ldots, n\} : f_i(x) = f(x) \}$$

Then $\partial_C f(x) = \mathrm{co}\, \{ f_i'(x) : i \in I(x) \}$

Some calculus

Sums:

$$\partial_C(f_1 + f_2)(x) \subset \partial_C f_1(x) + \partial_C f_2(x)$$

(equality when f_1, f_2 ***regular***)

Mean value theorem:

$$\exists\, z \in (x,y) \,/\, f(y) - f(x) \in \langle \zeta, y - x \rangle$$

Maximum functions: $f(x) = \max_{1 \le i \le n} f_i(x)$ (each f_i smooth)

$$I(x) = \{ i \in \{1, 2, \ldots, n\} : f_i(x) = f(x) \}$$

Then $\partial_C f(x) = \mathrm{co}\, \{ f_i'(x) : i \in I(x) \}$

Optimization:

$$\min_S f \text{ at } x \implies 0 \in \partial_C f(x) + N_S^C(x)$$

(more generally, ***Lagrange multipliers***)

Some calculus

Mean value theorem:

$$\exists\, z \in (x,y) \,/\, f(y) - f(x) \in \langle \zeta, y - x \rangle$$

Sums:

$$\partial_C(f_1 + f_2)(x) \subset \partial_C f_1(x) + \partial_C f_2(x)$$

(equality when f_1, f_2 ***regular***)

Maximum functions: $f(x) = \max_{1 \le i \le n} f_i(x)$ (each f_i smooth)

$$I(x) = \{ i \in \{1, 2, \ldots, n\} : f_i(x) = f(x) \}$$

Then $\partial_C f(x) = \mathrm{co}\, \{ f_i'(x) : i \in I(x) \}$

Optimization:

$$\min_S f \text{ at } x \implies 0 \in \partial_C f(x) + N_S^C(x)$$

(more generally, ***Lagrange multipliers***)

Some calculus

Mean value theorem:

$\exists\, z \in (x,y) \,/\, f(y) - f(x) \in \langle \zeta, y - x \rangle$

Sums:

$\partial_C (f_1 + f_2)(x) \subset \partial_C f_1(x) + \partial_C f_2(x)$

(equality when f_1, f_2 **regular**)

Maximum functions: $f(x) = \max_{1 \le i \le n} f_i(x)$ (each f_i smooth)

$I(x) = \{ i \in \{1, 2, \ldots, n\} : f_i(x) = f(x) \}$

Then $\partial_C f(x) = \mathrm{co}\{ f_i'(x) : i \in I(x) \}$

Optimization:

$\min_S f \text{ at } x \implies 0 \in \partial_C f(x) + N_S^C(x)$

(more generally, **Lagrange multipliers**)

Graph-closed:

$$\begin{array}{ccc} \zeta_i & \in & \partial_C f(x_i) \\ \downarrow & & \downarrow \\ \zeta & & x \end{array}$$

Some calculus

Mean value theorem:

$\exists\, z \in (x,y) \,/\, f(y) - f(x) \in \langle \zeta, y - x \rangle$

Sums:

$\partial_C (f_1 + f_2)(x) \subset \partial_C f_1(x) + \partial_C f_2(x)$

(equality when f_1, f_2 **regular**)

Maximum functions: $f(x) = \max_{1 \le i \le n} f_i(x)$ (each f_i smooth)

$I(x) = \{ i \in \{1, 2, \ldots, n\} : f_i(x) = f(x) \}$

Then $\partial_C f(x) = \mathrm{co}\{ f_i'(x) : i \in I(x) \}$

Optimization:

$\min_S f \text{ at } x \implies 0 \in \partial_C f(x) + N_S^C(x)$

(more generally, **Lagrange multipliers**)

Graph-closed:

$$\begin{array}{ccc} \zeta_i & \in & \partial_C f(x_i) \\ \downarrow & & \downarrow \\ \zeta & & x \end{array} \quad \Longrightarrow \quad \zeta \in \partial_C f(x)$$

Gradient formula

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

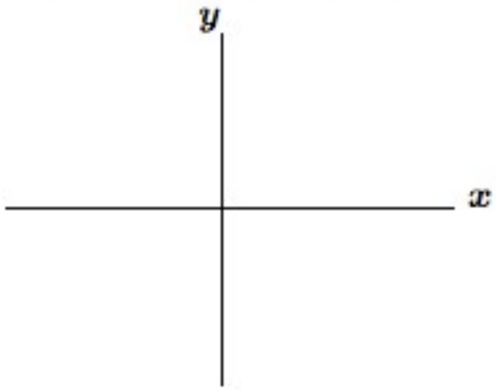
$$\partial_C f(x) = \mathrm{co}\,\{ \lim_{i \to \infty} \nabla f(x_i) : x_i \to x,\ x_i \notin \Omega \}$$

("blind to sets of measure 0").

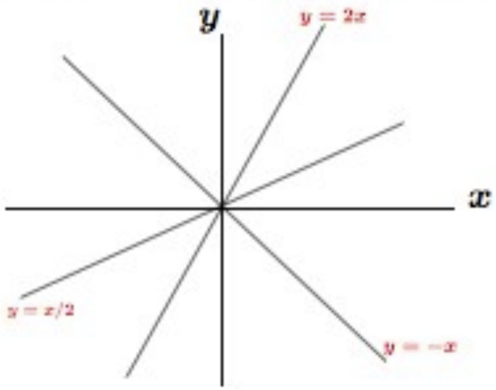
This is a useful tool for calculation.

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

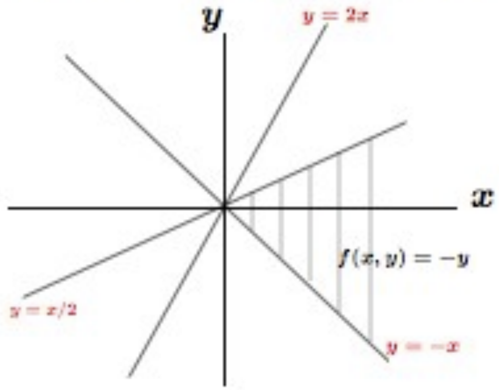
Example $f(x, y) = \max\{\min[x, -y], y - x\}$



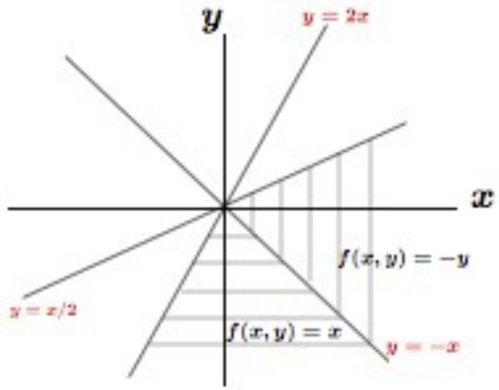
Example $f(x, y) = \max\{\min[x, -y], y - x\}$



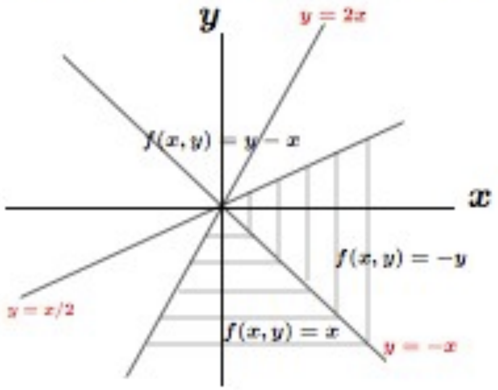
Example $f(x, y) = \max\{\min[x, -y], y - x\}$



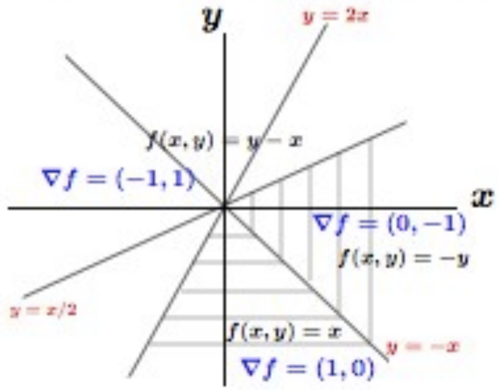
Example $f(x, y) = \max\{\min[x, -y], y - x\}$



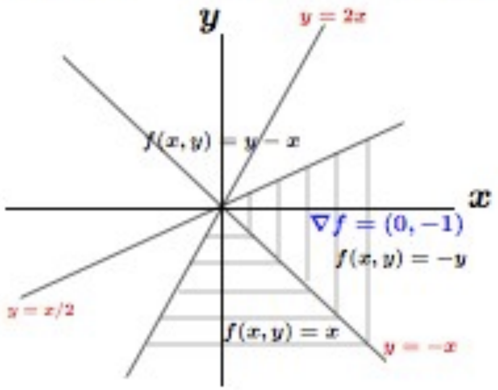
Example $f(x,y) = \max\{\min[x,-y], y-x\}$



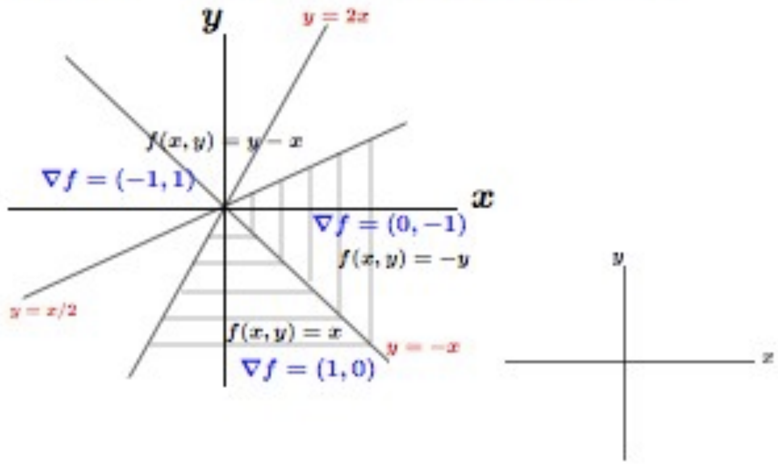
Example $f(x,y) = \max\{\min[x,-y], y-x\}$



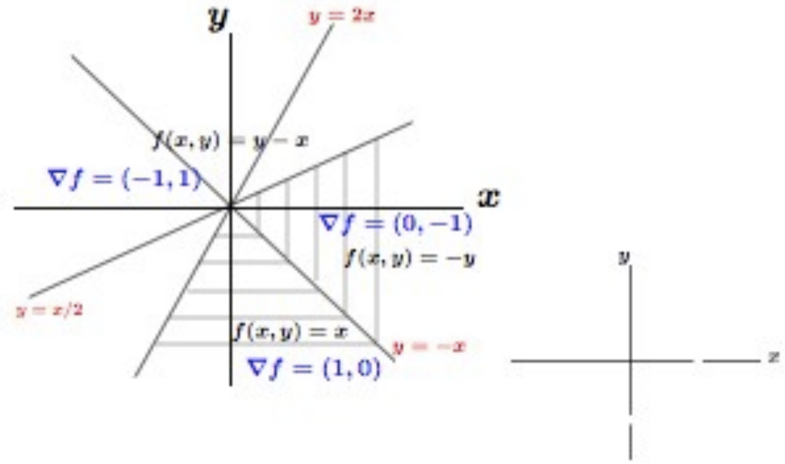
Example $f(x,y) = \max\{\min[x,-y], y-x\}$



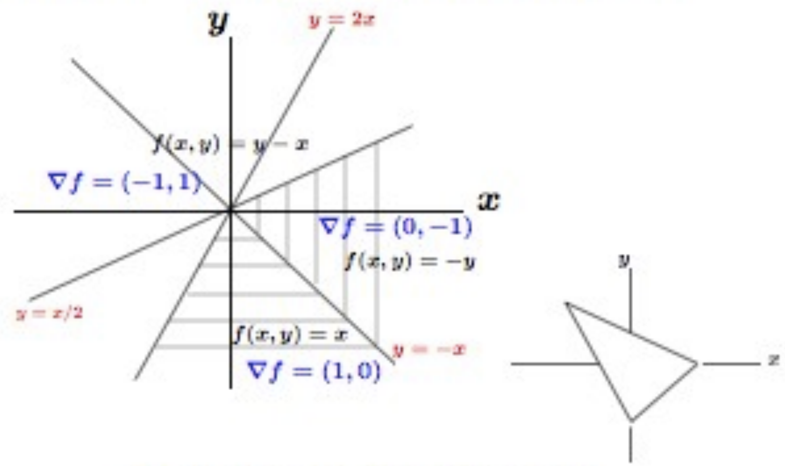
Example $f(x,y) = \max\{\min[x,-y], y-x\}$



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



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



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

Gradient formula

When f if locally Lipschitz on $\mathbb{R}^n$, then f is differentiable a.e. (Rademacher). Let Ω 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.

Gradient formula

When f if locally Lipschitz on $\mathbb{R}^n$, then f is differentiable a.e. (Rademacher). Let Ω 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/Zeidan]

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

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

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

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

Example

$$F(x,y) = \left[|x| + y, 2x + |y|\right]$$

$$\partial_C F(0,0) = \left\{ \begin{bmatrix} s & 1 \\ 2 & t \end{bmatrix} : -1 \le s \le 1,\ -1 \le t \le 1 \right\}$$

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

Calculus of sets

Let $x_0 \in S := \{x : f(x) \le 0\}$. If $0 \notin \partial_C f(x_0)$, then

$$T_S^C(x_0) \supset \{v \in X : f^\circ(x_0; v) \le 0\}.$$

If in addition f is regular at x_0, then

$$T_S^C(x_0) = \{v \in X : f'(x_0; v) \le 0\} \text{ and}$$

$$N_S^C(x_0) = \{\lambda\zeta : \lambda \ge 0, \zeta \in \partial_C f(x_0)\}.$$

Calculus of sets

Let $x_0 \in S := \{x : f(x) \le 0\}$. If $0 \notin \partial_C f(x_0)$, then

$$T_S^C(x_0) \supset \{v \in X : f^\circ(x_0; v) \le 0\}.$$

If in addition f is regular at x_0, then

$$T_S^C(x_0) = \{v \in X : f'(x_0; v) \le 0\} \text{ and}$$

$$N_S^C(x_0) = \{\lambda\zeta : \lambda \ge 0, \zeta \in \partial_C f(x_0)\}.$$

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

$$T_S^C(x_0) = \{v \in X : \langle F'(x_0), v\rangle = 0\} \text{ and}$$

$$N_S^C(x_0) = \{\theta F'(x_0) : \theta \in Y^*\}.$$

Calculus of sets

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

$$T_S^C(x_0) = \{v \in X : \langle F'(x_0), v\rangle = 0\} \text{ and}$$

$$N_S^C(x_0) = \{\theta F'(x_0) : \theta \in Y^*\}.$$

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

$$T_S^C(x_0) \supset \{v \in X : f^\circ(x_0; v) \leq 0\}.$$

If in addition f is regular at x_0, then

$$T_S^C(x_0) = \{v \in X : f'(x_0; v) \leq 0\} \text{ and}$$

$$N_S^C(x_0) = \{\lambda\zeta : \lambda \geq 0, \zeta \in \partial_C f(x_0)\}.$$

If $N_{S_1}^C(x) \bigcap -N_{S_2}^C(x) = \{0\}$, then

$$N_{S_1 \cap S_2}^C(x) \subset N_{S_1}^C(x) + N_{S_2}^C(x) \text{ and}$$
$$T_{S_1 \cap S_2}^C(x) \supset T_{S_1}^C(x) \cap T_{S_2}^C(x),$$

with equality when S_1 and S_2 are regular.

Wedged (or epi-Lipschitz) sets

A set S is said to be ***wedged*** if:

$$\text{int } T_S^C(x) \neq \emptyset \;\forall\, x \in S$$

Wedged (or epi-Lipschitz) sets

A set S is said to be ***wedged*** if:

$$\text{int } T_S^C(x) \neq \emptyset \;\forall\, x \in S$$

not wedged

Wedged (or epi-Lipschitz) sets

A set S is said to be ***wedged*** if:

$$\text{int } T_S^C(x) \neq \emptyset \;\forall\, x \in S$$

not wedged

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

- $\text{int } S \neq \emptyset$
- $S = \text{cl}\,\{\text{int } S\}$
- $T_S^C(x) = \mathbb{R}^n$ 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

$$\phi(x) \in T_{B(0,1)}(x) \; \forall \, x \in B(0,1).$$

Then there exists $x_0 \in B(0,1)$ such that $\phi(x_0) = 0$.

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

Let $\phi : \mathbb{R}^n \to \mathbb{R}^n$ be a Lipschitz function satisfying

$$\phi(x) \in T_{B(0,1)}(x) \; \forall \, x \in B(0,1).$$

Then there exists $x_0 \in B(0,1)$ such that $\phi(x_0) = 0$.

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

Since Brouwer's Theorem holds when the ball is replaced by a set S homeomorphic to it, we can ask if the above still holds in that case.

Let $\phi : \mathbb{R}^n \to \mathbb{R}^n$ be a Lipschitz function satisfying

$$\phi(x) \in T_{B(0,1)}(x) \; \forall \, x \in B(0,1).$$

Then there exists $x_0 \in B(0,1)$ such that $\phi(x_0) = 0$.

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

Since Brouwer's Theorem holds when the ball is replaced by a set S homeomorphic to it, we can ask if the above still holds in that case.

Let $\phi : \mathbb{R}^n \to \mathbb{R}^n$ be a Lipschitz function satisfying

$$\phi(x) \in T_S^C(x) \; \forall \, x \in S,$$

where S is wedged and homeomorphic to $B(0,1)$.
Then there exists $x_0 \in S$ such that $\phi(x_0) = 0$.

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

Boundary analysis: Inner and outer sphere conditions, lower C^2 property, reach, semiconcavity, ϕ-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:

Boundary analysis: Inner and outer sphere conditions, lower C^2 property, reach, semiconcavity, ϕ-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?

Boundary analysis: Inner and outer sphere conditions, lower C^2 property, reach, semiconcavity, ϕ-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, ϕ-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, $0 < r < R$?

Boundary analysis: Inner and outer sphere conditions, lower C^2 property, reach, semiconcavity, ϕ-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, ϕ-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:

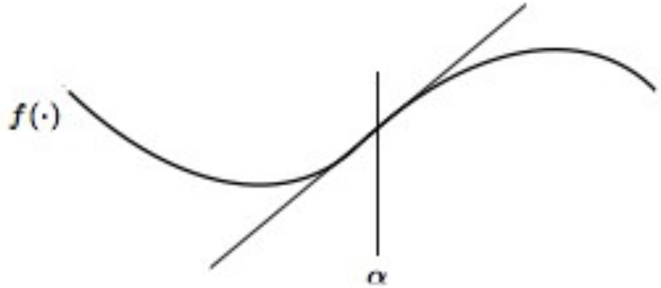
Is S then the union of balls of radius R? No in general

Is S the union of balls of radius r, 0 < r < R? Yes, if S is wedged and bounded

Does $r = \dfrac{nR}{2\sqrt{n^2-1}}$ suffice?

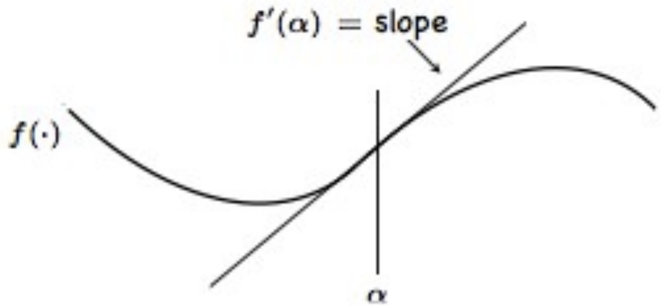
Proximal theory

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



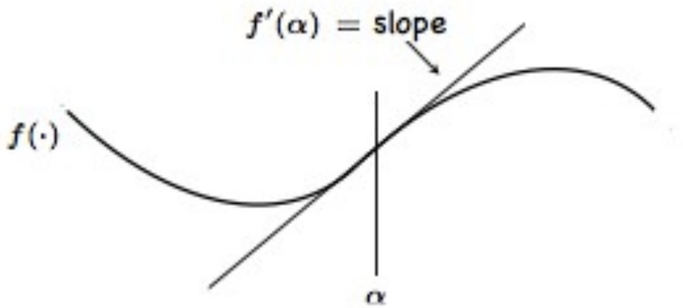
Proximal theory

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



Proximal theory

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



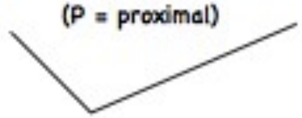
We may also approximate just from below, using nonlinear functions: proximal analysis

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

(P = proximal)

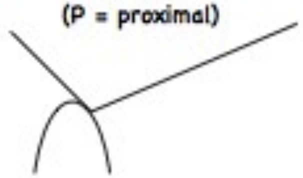
The set of all `contact slopes' of lower locally supporting parabolas is the proximal subdifferential $\partial_p f(\alpha)$

(P = proximal)



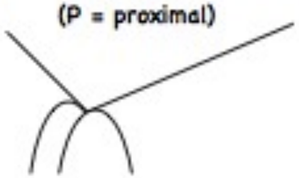
The set of all `contact slopes' of lower locally supporting parabolas is the proximal subdifferential $\partial_p f(\alpha)$

(P = proximal)



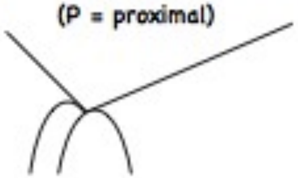
The set of all `contact slopes' of lower locally supporting parabolas is the proximal subdifferential $\partial_p f(\alpha)$

(P = proximal)



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

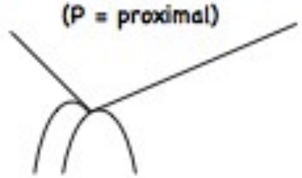
(P = proximal)



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

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

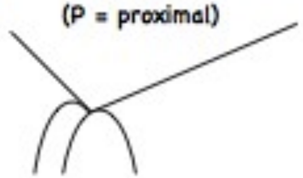
(P = proximal)



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

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

(P = proximal)



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

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

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

(P = proximal)

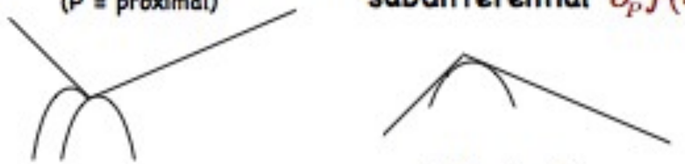
$\partial_P f(\alpha) = [-2, 1]$ $\qquad$ $\partial_P f(\alpha) = \emptyset$

$\zeta \in \partial_P f(\alpha) \iff$

$$f(x) \geq \langle \zeta, x - \alpha \rangle + f(\alpha) - \sigma |x - \alpha|^2 \text{ locally}$$

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

(P = proximal)



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

$\zeta \in \partial_P f(\alpha) \iff$

$$f(x) \geq \langle \zeta, x - \alpha \rangle + f(\alpha) - \sigma |x - \alpha|^2 \text{ locally}$$

$\partial_P f$ 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)!

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)!

$$\zeta \in \partial_P(f_1+f_2)(x) \implies$$
$$\forall\, \epsilon > 0\, \exists\, x_1, x_2, \eta :$$

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)!

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

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)!

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

and

$$\zeta \in \partial_P f_1(x_1) + \partial_P f_2(x_2) + \eta$$

We cannot expect to have, in general:

$$\partial_P(f_1+f_2)(x) \subset \partial_P f_1(x) + \partial_P f_2(x)$$

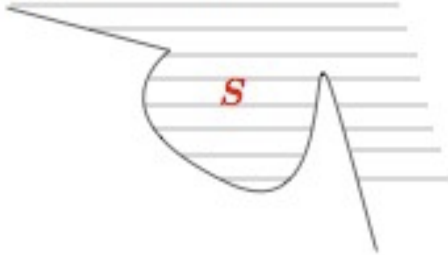
In fact, the ***opposite*** is true (but not useful)!

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

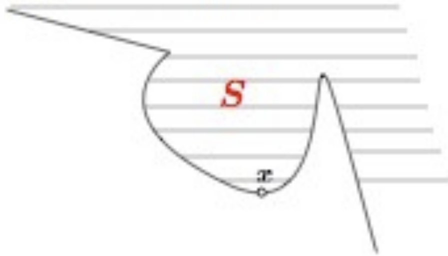
and

$$\zeta \in \partial_P f_1(\underset{\text{near } x}{x_1}) + \partial_P f_2(\underset{\text{near } x}{x_2}) + \underset{\text{small}}{\eta}$$

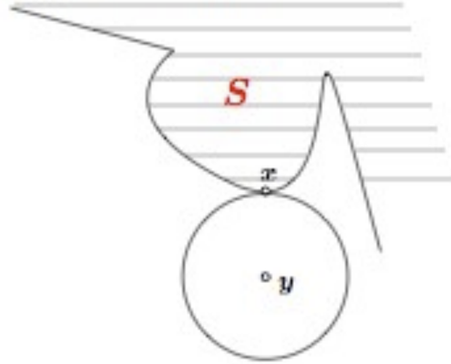
The geometry of proximal normals



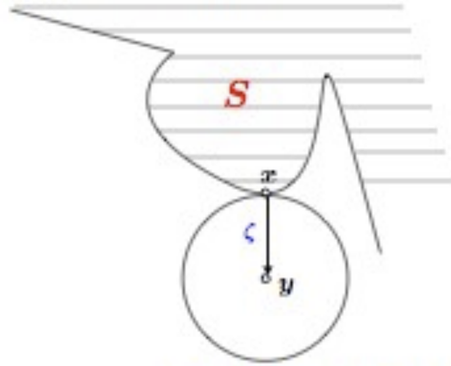
The geometry of proximal normals



The geometry of proximal normals

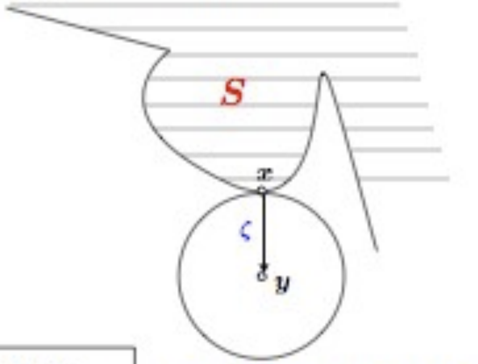


The geometry of proximal normals



ζ is a proximal normal to S at x

The geometry of proximal normals

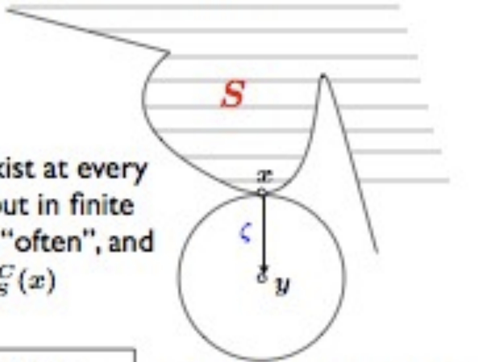


$\zeta \in N_S^P(x) \iff \exists \sigma \geq 0 :$

$\langle \zeta, y - x \rangle \leq \sigma |y - x|^2 \ \ \forall y \in S$

ζ is a proximal normal to S at x

The geometry of proximal normals



Such normals don't exist at every boundary point of S, but in finite dimensions they exist "often", and generate the cone $N_S^C(x)$

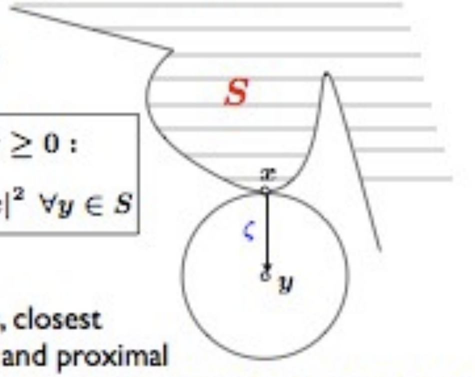
$\zeta \in N_S^P(x) \iff \exists \sigma \geq 0 :$

$\langle \zeta, y - x \rangle \leq \sigma |y - x|^2 \ \ \forall y \in S$

ζ is a proximal normal to S at x

The geometry of proximal normals

Such normals don't exist at every boundary point of S, but in finite dimensions they exist "often", and generate the cone $N_S^C(x)$



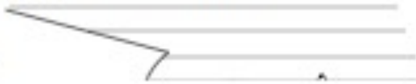
$\zeta \in N_S^P(x) \iff \exists \sigma \geq 0 :$

$\langle \zeta, y - x \rangle \leq \sigma |y - x|^2 \ \ \forall y \in S$

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*

ζ is a proximal normal to S at x

The geometry of proximal normals



Fact:
if f is lower semicontinuous, finite at x, then

$$\zeta \in \partial_P f(x) \iff (\zeta, -1) \in N^P_{\mathrm{epi}\, f}(x, f(x))$$

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*

ζ is a proximal normal to S at x

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 = Limiting

Limiting constructs

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

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

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

L = Limiting

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

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

$\partial_C f$ vis-à-vis $\partial_P f$ / $\partial_L f$

f

$\partial_C f$ 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

$\partial_C f$ 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°, 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°, 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'

$\partial_C f$ 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°, 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

$$\partial_C f = \mathrm{co}\, \partial_L f$$

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