Use of piecewise linearization for (un)constrained optimization and ODE integration

Andreas Griewank

with thanks to

HUB: Torsten Bosse, Tom Streubel, Nikolai Strogies, ...
Uni Paderborn: Andrea Walther, Sabrina Fiege
MIT: Paul Barton, Kamil Khan

Recent Advances in Optimization, July 2013

Smooth stuff I won't talk about (very much)
- Cubic Overestimation with Hessian Updating
- Gauss-Transposed-Broyden for Least Squares

Generalized Derivatives and Semismoothness
- Background and Motivation
- Generalized differentiation rules
- Semismooth Newton Result

Piecewise linearization Approach
- Piecewise linearization rules
- Approximation and Continuity
- Computing Generalized Jacobians

Applications to fundamental tasks
- Nonsmooth equation solving
- (Un)constrained Optimization
- Integration of Lipschitzian Dynamics

Observations on Generalized Hessians

Alternative stabilizations of quadratic model

$$f(x+s)-f(x) \approx m(s) \equiv g^\top s+\tfrac{1}{2}s^\top Bs \text{ with } g \equiv \nabla f(x),\ B \approx \nabla^2 f(x)$$

Line-search with $B \succ 0$

$$s = \mathbf{argmin}\{m(s)\}$$

$$\alpha \approx \mathbf{argmin}\{f(x+\alpha s)\} \quad \text{and} \quad x_+ = x+\alpha s$$

Trust Region with radius $\Delta > 0$

$$s = \mathbf{argmin}\{m(s) : \|s\| \leq \Delta\}$$

$$\text{if} \quad f(x+s) \ll f(x) \quad \text{then} \quad x_+ = x+s \quad \text{else } \Delta_+ \ll \Delta$$

Cubic overestimation with Lipschitz constant $2q > 0$

$$s = \mathbf{argmin}\{m(s) + q\|s\|^3/3\}$$

$$\text{if} \quad f(x+s) \ll f(x) \quad \text{then} \quad x_+ = x+s \quad \text{else } q_+ \gg q$$

Alternative stabilizations of quadratic model

$$f(x+s)-f(x) \approx m(s) \equiv g^\top s+\tfrac{1}{2}s^\top Bs \text{ with } g \equiv \nabla f(x),\ B \approx \nabla^2 f(x)$$

Line-search with $B \succ 0$

$$s = \textbf{argmin}\{m(s)\}$$

$$\alpha \approx \textbf{argmin}\{f(x+\alpha s)\} \quad \text{and} \quad x_+ = x+\alpha s$$

Trust Region with radius $\Delta > 0$

$$s = \textbf{argmin}\{m(s) : \|s\| \le \Delta\}$$

$$\text{if} \quad f(x+s) \ll f(x) \quad \text{then} \quad x_+ = x+s \quad \text{else } \Delta_+ \ll \Delta$$

Cubic overestimation with Lipschitz constant $2q > 0$

$$s = \textbf{argmin}\{m(s) + q\|s\|^3/3\}$$

$$\text{if} \quad f(x+s) \ll f(x) \quad \text{then} \quad x_+ = x+s \quad \text{else } q_+ \gg q$$

Alternative stabilizations of quadratic model

$$f(x+s)-f(x) \approx m(s) \equiv g^\top s+\tfrac{1}{2}s^\top Bs \text{ with } g \equiv \nabla f(x),\ B \approx \nabla^2 f(x)$$

Line-search with $B \succ 0$

$$s = \mathbf{argmin}\{m(s)\}$$

$$\alpha \approx \mathbf{argmin}\{f(x+\alpha s)\} \quad \text{and} \quad x_+ = x + \alpha s$$

Trust Region with radius $\Delta > 0$

$$s = \mathbf{argmin}\{m(s) : \|s\| \leq \Delta\}$$

$$\text{if} \quad f(x+s) \ll f(x) \quad \text{then} \quad x_+ = x + s \quad \text{else } \Delta_+ \ll \Delta$$

Cubic overestimation with Lipschitz constant $2q > 0$

$$s = \mathbf{argmin}\{m(s) + q\|s\|^3/3\}$$

$$\text{if} \quad f(x+s) \ll f(x) \quad \text{then} \quad x_+ = x + s \quad \text{else } q_+ \gg q$$

Alternative stabilizations of quadratic model

$$f(x+s)-f(x) \;\approx\; m(s) \equiv g^\top s+\tfrac{1}{2}s^\top Bs \text{ with } g \equiv \nabla f(x),\; B \approx \nabla^2 f(x)$$

Line-search with $B \succ 0$

$$s = \textbf{argmin}\{m(s)\}$$

$$\alpha \approx \textbf{argmin}\{f(x+\alpha s)\} \quad \text{and} \quad x_+ = x+\alpha s$$

Trust Region with radius $\Delta > 0$

$$s = \textbf{argmin}\{m(s) : \|s\| \leq \Delta\}$$

$$\text{if} \quad f(x+s) \ll f(x) \quad \text{then} \quad x_+ = x+s \quad \text{else } \Delta_+ \ll \Delta$$

Cubic overestimation with Lipschitz constant $2q > 0$

$$s = \textbf{argmin}\{m(s) + q\|s\|^3/3\}$$

$$\text{if} \quad f(x+s) \ll f(x) \quad \text{then} \quad x_+ = x+s \quad \text{else } q_+ \gg q$$

Step Computation and Properties

As shown by G. and others global minimum of the cubic model

$$g(x)^\top s + \tfrac{1}{2} s^\top B s + \tfrac{1}{3} q \|s\|^3$$

is attained at all solutions of the appended linear system

$$(B + \lambda I) s = -g \quad \text{with} \quad \lambda \; = \; q\|s\|$$

for which $B(\lambda) \equiv B + \lambda I$ is positive semi-definite.

By eliminating $s(\lambda)$ we obtain a secular equation

$$\varphi(\lambda) \; \equiv \; \|s(\lambda)\|^2 \; \equiv \; g^\top (B + \lambda I)^{-2} g \; = \; \lambda^2 / \, q^2 .$$

where LHS/RHS are convex and monotonic falling/growing.

Need to solve systems:

$$(B + \lambda I)s = -g \quad \text{for various } \lambda \in \mathbb{R}$$

with B subject to shifted BFGS, SR1 or compromise update.
Hessenberg form etc does not work, i.e. generates cost of $\mathcal{O}(n^3)$.

First way out: Updating the EVD

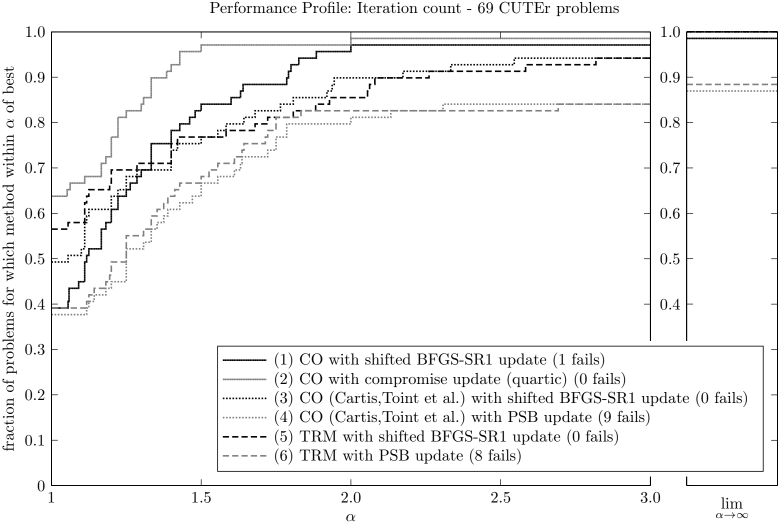
$$B \;=\; V \Lambda V^\top \quad \text{with diagonal } \Lambda \text{ and orthogonal } V$$

$$B_\pm \;=\; B \pm cc^\top \;=\; V_\pm \Lambda_\pm V_\pm^\top$$

requires $\mathcal{O}(n^2)$ operations for $\Lambda_\pm$ via the secular equation and $\mathcal{O}((n \log n)^2)$ operations for $V_\pm$ using fast polynomial arithmetic.
But method is barely numerically stable even when B is symmetric.

True way out: Nocedal Recurrence ?

Keep B in limited memory product format and apply modified NR!


Performance Profile: Iteration count - 69 CUTEr problems
fraction of problems for which method within α of best
α
lim α→∞
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1
1.5
2.0
2.5
3.0
(1) CO with shifted BFGS-SR1 update (1 fails)
(2) CO with compromise update (quartic) (0 fails)
(3) CO (Cartis,Toint et al.) with shifted BFGS-SR1 update (0 fails)
(4) CO (Cartis,Toint et al.) with PSB update (9 fails)
(5) TRM with shifted BFGS-SR1 update (0 fails)
(6) TRM with PSB update (8 fails)

Basic Approach (see also Eldon Haber)

Data assimilation and other least squares require solutions of

$$\min_x \varphi(x) \equiv \tfrac{1}{2}\|F(x)\|^2 \quad \text{for} \quad F : \mathbb{R}^n \mapsto \mathbb{R}^m$$

quasi-Gauss Newton approach for computing step s at point x

$$B^\top B\, s = -F'(x)^\top F(x) = -\nabla\varphi(x) \in \mathbb{R}^n$$

using $F'(x) \approx B \in \mathbb{R}^{m\times n}$ and only derivative vectors

$$z \equiv F'(x)^\top t \in \mathbb{R}^n \quad \text{and} \quad y \equiv F'(x)\, s \in \mathbb{R}^m$$

Two-sided rank one formula

$$B_+(z) \;=\; B + r\left[r^\top \left(F'(x_+) - B\right)\right] \Big/ \|r\|^2 \quad \text{with} \quad r = y - B\,s$$

Satisfies primal secant condition:

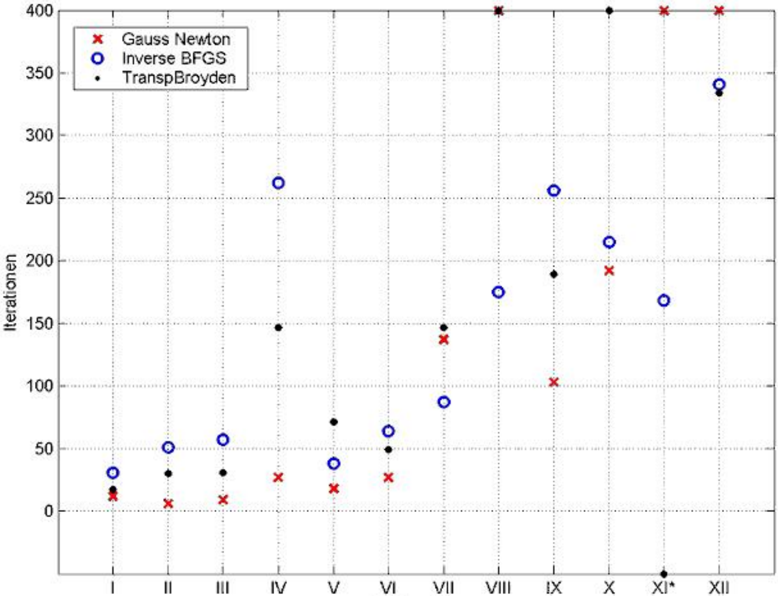
$$B_+\,s = r + B\,s = F'(x_+)s \;=\; F(x+s) - F(x) + O(\|s\|^2)$$

Satisfies dual secant condition:

$$B_+^\top z \;=\; F'(x_+)^\top z \quad \text{for} \quad r = z \in \mathbb{R}^m$$

General Properties:

Fixed scale least change $\Longrightarrow$ bounded deterioration and
Heredity in affine case yielding optimal rate in square case.


Gauss Newton
Inverse BFGS
TranspBroyden
Iterationen
400
350
300
250
200
150
100
50
0
I
II
III
IV
V
VI
VII
VIII
IX
X
XI*
XII

Theoretical convergence result:

Assumption: Smoothness and injectivity on level set

$$F \in C^{1,1} \quad \text{and} \quad \inf_{\|F(x)\| \leq \|F(x_0)\| \geq \|F(z)\|} \left\{ \frac{\|F(x) - F(z)\|}{\|z - x\|} \right\} > 0$$

Wedin line-search $\Longrightarrow$ Square summability

$$\left| \frac{F_+^\top (F_+ - F)}{\|F_+ - F\|^2} - 1 \right| < \varepsilon \implies \sum \|F_+ - F\| < \infty \implies$$

$$\textbf{mean} \left\{ \tfrac{1}{\|s\|} \|(F'_+ - B)\, s\| + \tfrac{1}{\|r\|} \|(F'_+ - B)^\top r\| + \tfrac{1}{\|y\|} \|(F'_+ - B)^\top y\| \right\} = 0$$

Average Dennis and Moré $\Longrightarrow$ Gauss-Newton Rate

$$\textbf{mean} \left\{ [B^\top B - F'(x)^\top F'(x)]\, s \right\} = 0 \implies \|s - s^{GN}\| \big/ \|s\| \to 0$$

Can we turn this into Algebra!?

Directional derivative á la Dini, Hadamard, Clarke .....

$$F^{??}(\mathring{x}; \Delta x) \equiv \limsup_{\substack{x \to \mathring{x} \\ v \to \Delta x \\ t \searrow 0}} \left[\frac{F(x + tv) - F(x)}{t} \right]$$

Normal cone a la Mordukhovich in $\mathbb{R}^n$

$$\mathcal{N}(x; M) \equiv \limsup_{z \to x} \left\{ u^\top \in \mathbb{R}^m : \lim_{M \ni y \to z} \frac{u^\top (y - z)}{\|y - z\|} = 0 \right\}$$

Computational complexity?

Perturbations on x and Δx require exploration of F in full domain!!!

Can we turn this into Algebra!?

Directional derivative á la Dini, Hadamard, Clarke .....

$$F^{??}(\mathring{x}; \Delta x) \;\equiv\; \limsup_{\substack{x \to \mathring{x} \\ v \to \Delta x \\ t \searrow 0}} \left[\frac{F(x + tv) - F(x)}{t} \right]$$

Normal cone a la Mordukhovich in $\mathbb{R}^n$

$$\mathcal{N}(x; M) \;\equiv\; \limsup_{z \to x} \left\{ u^\top \in \mathbb{R}^m : \lim_{M \ni y \to z} \frac{u^\top (y - z)}{\|y - z\|} = 0 \right\}$$

Computational complexity?

Perturbations on x and Δx require exploration of F in full domain!!!

Can we turn this into Algebra!?

Directional derivative á la Dini, Hadamard, Clarke .....

$$F^{??}(\mathring{x};\Delta x) \;\equiv\; \limsup_{\substack{x \to \mathring{x} \\ v \to \Delta x \\ t \searrow 0}} \left[\frac{F(x+tv)-F(x)}{t} \right]$$

Normal cone a la Mordukhovich in $\mathbb{R}^n$

$$\mathcal{N}(x;M) \;\equiv\; \limsup_{z \to x} \left\{ u^\top \in \mathbb{R}^m : \lim_{M \ni y \to z} \frac{u^\top (y-z)}{\|y-z\|} = 0 \right\}$$

Computational complexity?

Perturbations on x and Δx require exploration of F in full domain!!!

Notational Zoo (Subspecies in Lipschitzian Habitat):

Fréchet Derivative: $\partial F(x) \equiv \partial F/\partial x : \mathcal{D} \mapsto \mathbb{R}^{m\times n} \cup \emptyset$

Limiting Jacobians: $\partial^L F(\mathring{x}) \equiv \overline{\lim}_{x\to\mathring{x}} \partial F(x) : \mathcal{D} \rightrightarrows \mathbb{R}^{m\times n}$

Clarke Jacobians: $\partial^C F(x) \equiv \mathbf{conv}(\partial^L F(x)) : \mathcal{D} \rightrightarrows \mathbb{R}^{m\times n}$

Bouligand:
$$\begin{aligned} F'(x;\Delta x) &\equiv \lim_{t\searrow 0}[F(x+t\Delta x)-F(x)]/t \\ &: \mathcal{D}\times\mathbb{R}^n \mapsto \mathbb{R}^m \\ &: \mathcal{D} \mapsto \mathbf{PL}_h(\mathbb{R}^n,\mathbb{R}^m) \end{aligned}$$

Piecewise linearization:

$$\begin{aligned} \Delta F(x;\Delta x) &: \mathcal{D}\times\mathbb{R}^n \mapsto \mathbb{R}^m \\ &: \mathcal{D} \mapsto \mathbf{PL}(\mathbb{R}^n,\mathbb{R}^m) \end{aligned}$$

Moriarty Effect by Rademacher $(\mathcal{C}^{0,1} = W^{1,\infty})$

Almost everywhere all concepts reduce to Fréchet, except PL!!

Notational Zoo (Subspecies in Lipschitzian Habitat):

Fréchet Derivative: $\partial F(x) \equiv \partial F/\partial x : \mathcal{D} \mapsto \mathbb{R}^{m\times n} \cup \emptyset$

Limiting Jacobians: $\partial^L F(\mathring{x}) \equiv \overline{\lim}_{x\to\mathring{x}} \partial F(x) : \mathcal{D} \rightrightarrows \mathbb{R}^{m\times n}$

Clarke Jacobians: $\partial^C F(x) \equiv \mathbf{conv}(\partial^L F(x)) : \mathcal{D} \rightrightarrows \mathbb{R}^{m\times n}$

Bouligand:
$$\begin{aligned} F'(x;\Delta x) &\equiv \lim_{t\searrow 0}[F(x+t\Delta x)-F(x)]/t \\ &: \mathcal{D}\times\mathbb{R}^n \mapsto \mathbb{R}^m \\ &: \mathcal{D} \mapsto \mathbf{PL}_h(\mathbb{R}^n,\mathbb{R}^m) \end{aligned}$$

Piecewise linearization:
$$\begin{aligned} \Delta F(x;\Delta x) &: \mathcal{D}\times\mathbb{R}^n \mapsto \mathbb{R}^m \\ &: \mathcal{D} \mapsto \mathbf{PL}(\mathbb{R}^n,\mathbb{R}^m) \end{aligned}$$

Moriarty Effect by Rademacher $(C^{0,1} = W^{1,\infty})$

Almost everywhere all concepts reduce to Fréchet, except PL!!

Notational Zoo (Subspecies in Lipschitzian Habitat):

Fréchet Derivative: $\partial F(x) \equiv \partial F/\partial x : \mathcal{D} \mapsto \mathbb{R}^{m\times n} \cup \emptyset$

Limiting Jacobians: $\partial^L F(\mathring{x}) \equiv \overline{\lim}_{x\to\mathring{x}} \partial F(x) : \mathcal{D} \rightrightarrows \mathbb{R}^{m\times n}$

Clarke Jacobians: $\partial^C F(x) \equiv \mathbf{conv}(\partial^L F(x)) : \mathcal{D} \rightrightarrows \mathbb{R}^{m\times n}$

Bouligand: $F'(x;\Delta x) \equiv \lim_{t\searrow 0}[F(x+t\Delta x) - F(x)]/t$

$$: \mathcal{D}\times\mathbb{R}^n \mapsto \mathbb{R}^m$$

$$: \mathcal{D} \mapsto \mathbf{PL}_h(\mathbb{R}^n,\mathbb{R}^m)$$

Piecewise linearization:

$$\Delta F(x;\Delta x) \; : \; \mathcal{D}\times\mathbb{R}^n \mapsto \mathbb{R}^m$$

$$: \mathcal{D} \mapsto \mathbf{PL}(\mathbb{R}^n,\mathbb{R}^m)$$

Moriarty Effect by Rademacher $(C^{0,1} = W^{1,\infty})$

Almost everywhere all concepts reduce to Fréchet, except PL!!

Notational Zoo (Subspecies in Lipschitzian Habitat):

Fréchet Derivative: $\partial F(x) \equiv \partial F/\partial x : \mathcal{D} \mapsto \mathbb{R}^{m\times n} \cup \emptyset$

Limiting Jacobians: $\partial^L F(\mathring{x}) \equiv \overline{\lim}_{x\to\mathring{x}} \partial F(x) : \mathcal{D} \rightrightarrows \mathbb{R}^{m\times n}$

Clarke Jacobians: $\partial^C F(x) \equiv \mathbf{conv}(\partial^L F(x)) : \mathcal{D} \rightrightarrows \mathbb{R}^{m\times n}$

Bouligand:
$$\begin{aligned} F'(x;\Delta x) &\equiv \lim_{t\searrow 0}[F(x+t\Delta x)-F(x)]/t \\ &: \mathcal{D}\times\mathbb{R}^n \mapsto \mathbb{R}^m \\ &: \mathcal{D} \mapsto \mathbf{PL}_h(\mathbb{R}^n,\mathbb{R}^m) \end{aligned}$$

Piecewise linearization:
$$\begin{aligned} \Delta F(x;\Delta x) &: \mathcal{D}\times\mathbb{R}^n \mapsto \mathbb{R}^m \\ &: \mathcal{D} \mapsto \mathbf{PL}(\mathbb{R}^n,\mathbb{R}^m) \end{aligned}$$

Moriarty Effect by Rademacher ($\mathcal{C}^{0,1} = W^{1,\infty}$)

Almost everywhere all concepts reduce to Fréchet, except PL!!

Notational Zoo (Subspecies in Lipschitzian Habitat):

Fréchet Derivative: $\partial F(x) \equiv \partial F/\partial x : \mathcal{D} \mapsto \mathbb{R}^{m\times n} \cup \emptyset$

Limiting Jacobians: $\partial^L F(\mathring{x}) \equiv \overline{\lim}_{x\to\mathring{x}} \partial F(x) : \mathcal{D} \rightrightarrows \mathbb{R}^{m\times n}$

Clarke Jacobians: $\partial^C F(x) \equiv \mathbf{conv}(\partial^L F(x)) : \mathcal{D} \rightrightarrows \mathbb{R}^{m\times n}$

Bouligand:
$$\begin{aligned} F'(x;\Delta x) \;&\equiv\; \lim_{t\searrow 0}[F(x+t\Delta x) - F(x)]/t \\ &:\; \mathcal{D}\times\mathbb{R}^n \mapsto \mathbb{R}^m \\ &:\; \mathcal{D} \mapsto \mathbf{PL}_h(\mathbb{R}^n,\mathbb{R}^m) \end{aligned}$$

Piecewise linearization:
$$\begin{aligned} \Delta F(x;\Delta x) \;&:\; \mathcal{D}\times\mathbb{R}^n \mapsto \mathbb{R}^m \\ &:\; \mathcal{D} \mapsto \mathbf{PL}(\mathbb{R}^n,\mathbb{R}^m) \end{aligned}$$

Moriarty Effect by Rademacher ($C^{0,1} = W^{1,\infty}$)

Almost everywhere all concepts reduce to Fréchet, except PL!!

Notational Zoo (Subspecies in Lipschitzian Habitat):

Fréchet Derivative: $\partial F(x) \equiv \partial F/\partial x : \mathcal{D} \mapsto \mathbb{R}^{m\times n} \cup \emptyset$

Limiting Jacobians: $\partial^L F(\mathring{x}) \equiv \overline{\lim}_{x\to\mathring{x}} \partial F(x) : \mathcal{D} \rightrightarrows \mathbb{R}^{m\times n}$

Clarke Jacobians: $\partial^C F(x) \equiv \mathbf{conv}(\partial^L F(x)) : \mathcal{D} \rightrightarrows \mathbb{R}^{m\times n}$

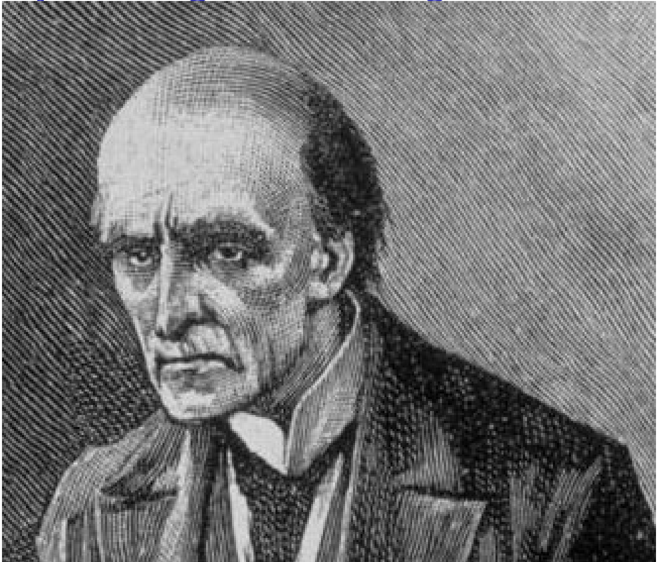
Bouligand:
$$
\begin{aligned}
F'(x;\Delta x) &\equiv \lim_{t\searrow 0}[F(x+t\Delta x)-F(x)]/t \\
&: \mathcal{D}\times\mathbb{R}^n \mapsto \mathbb{R}^m \\
&: \mathcal{D} \mapsto \mathbf{PL}_h(\mathbb{R}^n,\mathbb{R}^m)
\end{aligned}
$$

Piecewise linearization:
$$
\begin{aligned}
\Delta F(x;\Delta x) &: \mathcal{D}\times\mathbb{R}^n \mapsto \mathbb{R}^m \\
&: \mathcal{D} \mapsto \mathbf{PL}(\mathbb{R}^n,\mathbb{R}^m)
\end{aligned}
$$

Moriarty Effect by Rademacher $\left(\mathcal{C}^{0,1} = W^{1,\infty}\right)$

Almost everywhere all concepts reduce to Fréchet, except PL!!

Always lurking in the background: Prof. Moriarty



Relations holding for ∂^L with implications for $\partial^C \equiv conv(\partial^L)$

- $\partial^L(\alpha F) = \alpha \partial^L(F)$ for $\alpha \in \mathbb{R}$
- $\partial^L \binom{G}{F} \subseteq \partial^L F \times \partial^L G \equiv \left\{ \binom{A}{B} : A \in \partial^L(F), B \in \partial^L(G) \right\}$
- $\partial^L(G \circ F) = \partial^L G(F) \cdot \partial^L F$ if $G \in \mathcal{C}^1(R^m)$
- $\partial^L(F \pm G) \subseteq \partial^L F \pm \partial^L G = \{A \pm B : A \in \partial^L F, B \in \partial^L G\}$
- $\partial^L(f \cdot g) \subseteq g \cdot \partial^L f + f \cdot \partial^L g$
- $\partial^L|f| \quad \begin{cases} = \partial^L f & \text{when f>0} \\ \subseteq -\partial^L \cup \{0\} \cup \partial^L f & \text{when f=0} \\ = -\partial^L f & \text{when f<0} \end{cases}$

Relations holding for ∂^L with implications for $\partial^C \equiv conv(\partial^L)$

- $\partial^L(\alpha F) = \alpha \partial^L(F)$ for $\alpha \in \mathbb{R}$
- $\partial^L\binom{G}{F} \subseteq \partial^L F \times \partial^L G \equiv \left\{\binom{A}{B} : A \in \partial^L(F), B \in \partial^L(G)\right\}$
- $\partial^L(G \circ F) = \partial^L G(F) \cdot \partial^L F \quad$ if $\quad G \in \mathcal{C}^1(R^m)$
- $\partial^L(F \pm G) \subseteq \partial^L F \pm \partial^L G = \{A \pm B : A \in \partial^L F, B \in \partial^L G\}$
- $\partial^L(f \cdot g) \subseteq g \cdot \partial^L f + f \cdot \partial^L g$
- $\partial^L|f| \quad \begin{cases} = \partial^L f & \text{when f>0} \\ \subseteq -\partial^L \cup \{0\} \cup \partial^L f & \text{when f=0} \\ = -\partial^L f & \text{when f<0} \end{cases}$

Relations holding for ∂^L with implications for $\partial^C \equiv conv(\partial^L)$

- $\partial^L(\alpha F) = \alpha \partial^L(F)$ for $\alpha \in \mathbb{R}$
- $\partial^L \binom{G}{F} \subseteq \partial^L F \times \partial^L G \equiv \left\{ \binom{A}{B} : A \in \partial^L(F), B \in \partial^L(G) \right\}$
- $\partial^L(G \circ F) = \partial^L G(F) \cdot \partial^L F$ if $G \in \mathcal{C}^1(R^m)$
- $\partial^L(F \pm G) \subseteq \partial^L F \pm \partial^L G = \{A \pm B : A \in \partial^L F, B \in \partial^L G\}$
- $\partial^L(f \cdot g) \subseteq g \cdot \partial^L f + f \cdot \partial^L g$
- $\partial^L |f| \quad \begin{cases} = \partial^L f & \text{when f>0} \\ \subseteq -\partial^L \cup \{0\} \cup \partial^L f & \text{when f=0} \\ = -\partial^L f & \text{when f<0} \end{cases}$

Relations holding for ∂^L with implications for $\partial^C \equiv conv(\partial^L)$

- $\partial^L(\alpha F) = \alpha \partial^L(F)$ for $\alpha \in \mathbb{R}$
- $\partial^L \binom{G}{F} \subseteq \partial^L F \times \partial^L G \equiv \left\{ \binom{A}{B} : A \in \partial^L(F), B \in \partial^L(G) \right\}$
- $\partial^L(G \circ F) = \partial^L G(F) \cdot \partial^L F$ if $G \in \mathcal{C}^1(R^m)$
- $\partial^L(F \pm G) \subseteq \partial^L F \pm \partial^L G = \{A \pm B : A \in \partial^L F, B \in \partial^L G\}$
- $\partial^L(f \cdot g) \subseteq g \cdot \partial^L f + f \cdot \partial^L g$
- $\partial^L|f| \quad \begin{cases} = \partial^L f & \text{when f>0} \\ \subseteq -\partial^L \cup \{0\} \cup \partial^L f & \text{when f=0} \\ = -\partial^L f & \text{when f<0} \end{cases}$

Relations holding for ∂^L with implications for $\partial^C \equiv conv(\partial^L)$

- $\partial^L(\alpha F) = \alpha \partial^L(F)$ for $\alpha \in \mathbb{R}$
- $\partial^L \binom{G}{F} \subseteq \partial^L F \times \partial^L G \equiv \left\{ \binom{A}{B} : A \in \partial^L(F), B \in \partial^L(G) \right\}$
- $\partial^L(G \circ F) = \partial^L G(F) \cdot \partial^L F$ if $G \in \mathcal{C}^1(R^m)$
- $\partial^L(F \pm G) \subseteq \partial^L F \pm \partial^L G = \{A \pm B : A \in \partial^L F, B \in \partial^L G\}$
- $\partial^L(f \cdot g) \subseteq g \cdot \partial^L f + f \cdot \partial^L g$
- $\partial^L|f| \begin{cases} = \partial^L f & \text{when f>0} \\ \subseteq -\partial^L \cup \{0\} \cup \partial^L f & \text{when f=0} \\ = -\partial^L f & \text{when f<0} \end{cases}$

Relations holding for ∂^L with implications for $\partial^C \equiv conv(\partial^L)$

- $\partial^L(\alpha F) = \alpha \partial^L(F)$ for $\alpha \in \mathbb{R}$
- $\partial^L \binom{G}{F} \subseteq \partial^L F \times \partial^L G \equiv \left\{ \binom{A}{B} : A \in \partial^L(F), B \in \partial^L(G) \right\}$
- $\partial^L(G \circ F) = \partial^L G(F) \cdot \partial^L F$ if $G \in \mathcal{C}^1(R^m)$
- $\partial^L(F \pm G) \subseteq \partial^L F \pm \partial^L G = \{A \pm B : A \in \partial^L F, B \in \partial^L G\}$
- $\partial^L(f \cdot g) \subseteq g \cdot \partial^L f + f \cdot \partial^L g$
- $\partial^L|f| \quad \begin{cases} = \partial^L f & \text{when f>0} \\ \subseteq -\partial^L \cup \{0\} \cup \partial^L f & \text{when f=0} \\ = -\partial^L f & \text{when f<0} \end{cases}$

Direction of inclusions is:

Bad for evaluating (generalized) Jacobians:

since application may result in gross overestimation. Example:

$$\partial^L [|x| - |x|]_{x=0} = \{0\} \neq \{-2, 0, 2\} = \{-1, +1\} + \partial^L [|x|]_{x=0}$$

Good for propagating semi-smoothness:

$$\limsup_{J \in \partial^L F(x+s)} \|F(x+s) - F(x) - J\,s\| = o(\|s\|)$$

Consequence:

All compositions of smoothies and **abs**() are semismooth !!!

Direction of inclusions is:

Bad for evaluating (generalized) Jacobians:

since application may result in gross overestimation. Example:

$$\partial^L \left[|x| - |x|\right]_{x=0} = \{0\} \neq \{-2, 0, 2\} = \{-1, +1\} + \partial^L \left[|x|\right]_{x=0}$$

Good for propagating semi-smoothness:

$$\limsup_{J \in \partial^L F(x+s)} \|F(x+s) - F(x) - J\,s\| = o(\|s\|)$$

Consequence:

All compositions of smoothies and **abs**() are semismooth !!!

Proposition by Kummer, Qi, Sun,Kunisch et al

Semismoothness ensures that generalized Newton:

$$x_{k+1} = x_k - J^{-1}F(x_k) \quad \text{with} \quad J \in \partial^L F(x_k)$$

converges superlinearly to root $x_* \in F^{-1}(0)$ provided

$$\|x_0 - x_*\| \leq \rho \quad \text{and} \quad \|J^{-1}\| \leq M < \infty \quad \text{for } J \in \partial^L F(x_*)$$

Doubts concerning Applicability:

- How small is contraction radius $\rho > 0$?
- How can we calculate some $J \in \partial^L F(x)$?

Proposition by Kummer, Qi, Sun,Kunisch et al

Semismoothness ensures that generalized Newton:

$$x_{k+1} \;=\; x_k - J^{-1}F(x_k) \quad \text{with} \quad J \in \partial^L F(x_k)$$

converges superlinearly to root $x_* \in F^{-1}(0)$ provided

$$\|x_0 - x_*\| \leq \rho \quad \text{and} \quad \|J^{-1}\| \leq M < \infty \quad \text{for } J \in \partial^L F(x_*)$$

Doubts concerning Applicability:

- How small is contraction radius $\rho > 0$?
- How can we calculate some $J \in \partial^L F(x)$?

Proposition by Kummer, Qi, Sun,Kunisch et al

Semismoothness ensures that generalized Newton:

$$x_{k+1} \;=\; x_k - J^{-1}F(x_k) \quad \text{with} \quad J \in \partial^L F(x_k)$$

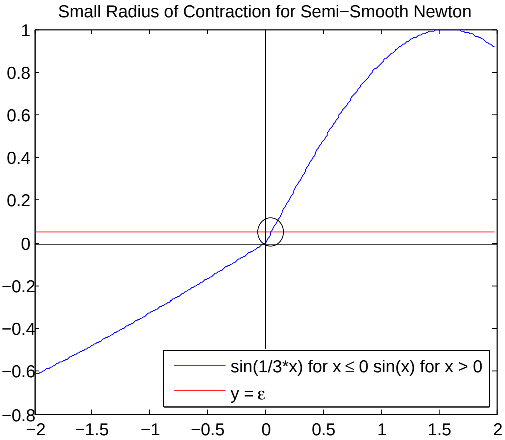
converges superlinearly to root $x_* \in F^{-1}(0)$ provided

$$\|x_0 - x_*\| \le \rho \quad \text{and} \quad \|J^{-1}\| \le M < \infty \quad \text{for } J \in \partial^L F(x_*)$$

Doubts concerning Applicability:

- How small is contraction radius $\rho > 0$?
- How can we calculate some $J \in \partial^L F(x)$?

Contraction radius $\leq$ distance to next kink



Tacit but realistic assumption:

$$y = F(x) : \mathcal{D} \subset \mathbb{R}^n \to \mathbb{R}^m$$

defined by *long* evaluation loop

$$\begin{array}{rrll} \text{input :} & v_{i-n} & = x_i & \text{for } i = 1 \ldots n \\ \text{evaluation :} & v_i & = \varphi_i \big(v_j\big)_{j \prec i} & \text{for } i = 1 \ldots \ell \\ \text{output :} & y_{m-i} & = v_{\ell-i} & \text{for } i = 0 \ldots m-1 \end{array}$$

where $v_i \in \mathbb{R}$ for $i = 1-n \ldots \ell$ and

$$\varphi_i \in \{+, -, *, /, \exp, \log, \sin, \cos, \ldots, \mathbf{abs}, \cdots\}$$

Partial pre-ordering

$$j \prec i \iff c_{ij} \equiv \frac{\partial}{\partial v_j} \varphi_i \not\equiv 0 \ .$$

abs covers min, max, $\|\cdot\|_1$, $\|\cdot\|_\infty$, table look-ups

Provided u and w are both finite one has

$$\begin{aligned} \max(u,w) &= \tfrac{1}{2}\left[u+w+\mathbf{abs}(u-w)\right] \\ \min(u,w) &= \tfrac{1}{2}\left[u+w-\mathbf{abs}(u-w)\right] \end{aligned}$$

and data (x_i, y_i) for $i = 0 \dots n$ with slopes s_0 and s_{n+1} on left and right are piecewise linearly interpolated by the formula

$$y = \frac{1}{2}\left[s_0(x-x_0)+y_0+\sum_{i=0}^{n}(s_{i+1}-s_i)\mathbf{abs}(x-x_i)+y_n+s_{n+1}(x-x_n)\right]$$

where $s_i = (y_{i+1}-y_i)/(x_{i+1}-x_i)$ represent the inner slopes.

Piecewise Linearization

We wish to determine for *base point* x and *increment* Δx

$$\boxed{\Delta y \;\equiv\; \Delta F(x;\Delta x) \;=\; F(x+\Delta x) - F(x) + \mathcal{O}(\|\Delta x\|^2)}$$

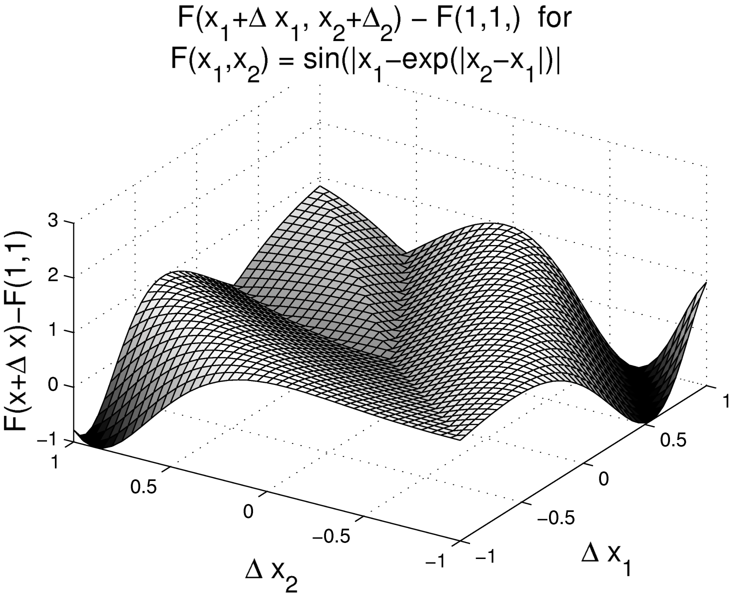
This can be done by propagating increments according to

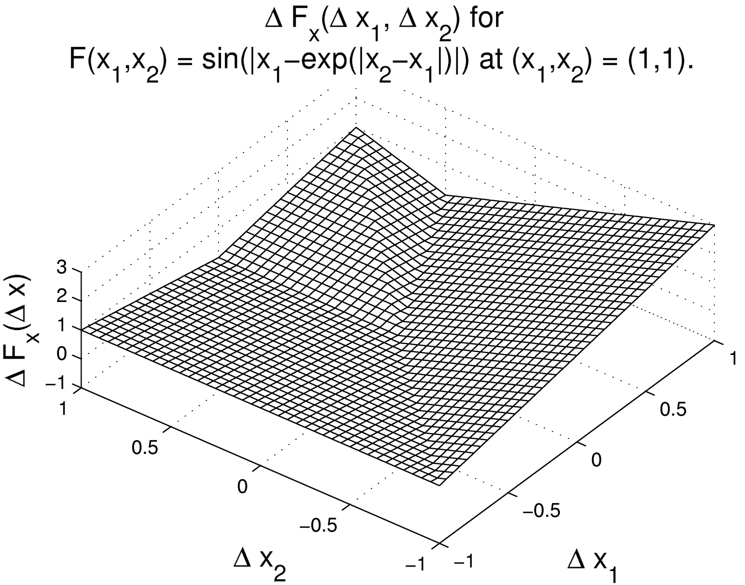
Smooth elementals

$$\begin{aligned}
\Delta v_i &= \Delta v_j \pm \Delta v_k \quad \text{for} \quad v_i = v_j \pm v_k \\
\Delta v_i &= v_j * \Delta v_k + \Delta v_j * v_k \quad \text{for} \quad v_i = v_j * v_k \\
\Delta v_i &= c_{ij}\,\Delta v_j \text{ with } c_{ij} \equiv \varphi_i'(v_j) \quad \text{for} \quad v_i = \varphi_i(v_j) \not\equiv \mathbf{abs}()
\end{aligned}$$

Lipschitz Elementals

$$\Delta v_i = \mathbf{abs}(v_j + \Delta v_j) - \mathbf{abs}(v_j) \quad \text{when} \quad v_i = \mathbf{abs}(v_j)\ .$$


F(x1+Δ x1, x2+Δ2) − F(1,1,) for
F(x1,x2) = sin(|x1−exp(|x2−x1|)|
F(x+Δ x)−F(1,1)
3
2
1
0
−1
1
0.5
0
−0.5
−1
Δ x2
−1
−0.5
0
0.5
1
Δ x1


Δ F_x(Δ x_1, Δ x_2) for
F(x_1,x_2) = sin(|x_1−exp(|x_2−x_1|)|) at (x_1,x_2) = (1,1).
Δ F_x(Δ x)
3
2
1
0
−1
1
0.5
0
−0.5
−1
Δ x_2
−1
−0.5
0
0.5
1
Δ x_1

Linearity and Product Rule

$$F, G : \mathcal{D} \subset \mathbb{R}^n \mapsto \mathbb{R}^m, \;\; \alpha, \beta \in \mathbb{R}$$

$$\Longrightarrow$$

$$\begin{aligned}
\Delta[\alpha F + \beta G](x; \Delta x) &= \alpha \Delta F(x, \Delta x) + \beta \Delta G(x, \Delta x) \\
\Delta[F^\top G](x; \Delta x) &= G(x)^\top \Delta F(x, \Delta x) + F(x)^\top \Delta G(x, \Delta x)
\end{aligned}$$

Chain Rule

$$F : \mathcal{D} \subset \mathbb{R}^n \mapsto \mathbb{R}^m \quad \text{and} \quad G : E \subset \mathbb{R}^m \mapsto \mathbb{R}^p \quad \text{with} \quad F(\mathcal{D}) \subset E$$

$$\Longrightarrow$$

$$\Delta[G \circ F](x; \Delta x) = \Delta G(F(x); \Delta F(x, \Delta x))$$

Linearity and Product Rule

$$F, G : \mathcal{D} \subset \mathbb{R}^n \mapsto \mathbb{R}^m,\ \alpha, \beta \in \mathbb{R}$$

$$\Longrightarrow$$

$$\begin{aligned} \Delta[\alpha F + \beta G](x; \Delta x) &= \alpha \Delta F(x, \Delta x) + \beta \Delta G(x, \Delta x) \\ \Delta[F^\top G](x; \Delta x) &= G(x)^\top \Delta F(x, \Delta x) + F(x)^\top \Delta G(x, \Delta x) \end{aligned}$$

Chain Rule

$$F : \mathcal{D} \subset \mathbb{R}^n \mapsto \mathbb{R}^m \quad \text{and} \quad G : E \subset \mathbb{R}^m \mapsto \mathbb{R}^p \quad \text{with} \quad F(\mathcal{D}) \subset E$$

$$\Longrightarrow$$

$$\Delta[G \circ F](x; \Delta x) = \Delta G(F(x); \Delta F(x, \Delta x))$$

Proposition (Approximation and Lipschitz Continuity)

Suppose F is composite Lipschitz on some open neighborhood $\mathcal{D}$ of a closed convex domain $\mathcal{K} \subset \mathbb{R}^n$. Then there exists a constant γ such that for all pairs $\mathring{x}, x \in \mathcal{K}$

$$\|F(x) - F(\mathring{x}) - \Delta F(\mathring{x}; x - \mathring{x})\| \leq \gamma \|x - \mathring{x}\|^2$$

Moreover, for any pair $\tilde{x}, \mathring{x} \in \mathcal{K}$, $\Delta x \in \mathbb{R}^n$, and a constant $\tilde{\gamma}$

$$\|\Delta F(\tilde{x}; \Delta x) - \Delta F(\mathring{x}; \Delta x)\| / (1 + \|\Delta x\|) \leq \tilde{\gamma} \|\tilde{x} - \mathring{x}\|$$

Finally there is a continuous radius $\rho(\mathring{x})$ such that

$$\Delta F(\mathring{x}; \Delta x) = F'(\mathring{x}; \Delta x) \quad \text{if} \quad \|\Delta x\| < \rho(\mathring{x})$$

Locally we reduce to the homogeneous piecewise linear $F'(x; \Delta x)$.

Proposition (Approximation and Lipschitz Continuity)

Suppose F is composite Lipschitz on some open neighborhood $\mathcal{D}$ of a closed convex domain $\mathcal{K} \subset \mathbb{R}^n$. Then there exists a constant γ such that for all pairs $\mathring{x}, x \in \mathcal{K}$

$$\|F(x) - F(\mathring{x}) - \Delta F(\mathring{x}; x - \mathring{x})\| \leq \gamma \|x - \mathring{x}\|^2$$

Moreover, for any pair $\tilde{x}, \mathring{x} \in \mathcal{K}$, $\Delta x \in \mathbb{R}^n$, and a constant $\tilde{\gamma}$

$$\|\Delta F(\tilde{x}; \Delta x) - \Delta F(\mathring{x}; \Delta x)\| / (1 + \|\Delta x\|) \leq \tilde{\gamma} \|\tilde{x} - \mathring{x}\|$$

Finally there is a continuous radius $\rho(\mathring{x})$ such that

$$\Delta F(\mathring{x}; \Delta x) = F'(\mathring{x}; \Delta x) \quad \textit{if} \quad \|\Delta x\| < \rho(\mathring{x})$$

Locally we reduce to the homogeneous piecewise linear $F'(x; \Delta x)$.

Proposition (Approximation and Lipschitz Continuity)

Suppose F is composite Lipschitz on some open neighborhood $\mathcal{D}$ of a closed convex domain $\mathcal{K} \subset \mathbb{R}^n$. Then there exists a constant γ such that for all pairs $\mathring{x}, x \in \mathcal{K}$

$$\|F(x) - F(\mathring{x}) - \Delta F(\mathring{x}; x - \mathring{x})\| \leq \gamma \|x - \mathring{x}\|^2$$

Moreover, for any pair $\tilde{x}, \mathring{x} \in \mathcal{K}$, $\Delta x \in \mathbb{R}^n$, and a constant $\tilde{\gamma}$

$$\|\Delta F(\tilde{x}; \Delta x) - \Delta F(\mathring{x}; \Delta x)\| / (1 + \|\Delta x\|) \leq \tilde{\gamma} \|\tilde{x} - \mathring{x}\|$$

Finally there is a continuous radius $\rho(\mathring{x})$ such that

$$\Delta F(\mathring{x}; \Delta x) = F'(\mathring{x}; \Delta x) \quad \textit{if} \quad \|\Delta x\| < \rho(\mathring{x})$$

Locally we reduce to the homogeneous piecewise linear $F'(x; \Delta x)$.

Reduced Representation in abs-normal form

After preaccumulation of smoothies

at fixed $\mathring{x}$ with strictly lower triangular $L \in \mathbb{R}^{s\times s}$

$$\begin{bmatrix} z \\ y \end{bmatrix} = \begin{bmatrix} \mathring{z} + Z(x-\mathring{x}) + L(|z| - |\mathring{z}|) \\ \mathring{y} + J(x-\mathring{x}) + Y(|z| - |\mathring{z}|) \end{bmatrix}$$

$$= \begin{bmatrix} c \\ b \end{bmatrix} + \begin{bmatrix} Z & L \\ J & Y \end{bmatrix} \begin{bmatrix} x \\ |z| \end{bmatrix} = \begin{bmatrix} c \\ b \end{bmatrix} + [\ C\ ] \begin{bmatrix} x \\ |z| \end{bmatrix}$$

The signature vector

$\sigma = \mathbf{sign}(z) \in \{-1, 0, 1\}^s$ characterizes control flow = selection.

Data c, b and sparse C computable

at cost $\leq (n+s)\mathbf{OPS}(F(x))$ by modification of e.g. ADOL-C

Beating the nonsmoothness superposition problem:

Proposition (Khan & Barton and A. G.)

$$\partial^K F(\mathring{x}) \;\equiv\; \partial^L_{\Delta x} \Delta F(\mathring{x}; \Delta x)\big|_{\Delta x = 0} \;\subset\; \partial^L F(x)\big|_{x=\mathring{x}}$$

contains those Jacobians $\partial F_\sigma(\mathring{x})$ for which the tangent cone

$$T_\sigma \;\equiv\; T_{\mathring{x}}\{x \in \mathcal{D} : F_\sigma(x) = F(x)\}$$

has a nonempty interior. (i.e. F_σ and ∂F_σ are conically active)

Remark

We can find several of them at cost $n\,OPS(F)$ in worst case.

All of them likely a stretch, there could be 2^s different ones.

Beating the nonsmoothness superposition problem:

Proposition (Khan & Barton and A. G.)

$$\partial^K F(\mathring{x}) \;\equiv\; \partial^L_{\Delta x} \Delta F(\mathring{x}; \Delta x)\big|_{\Delta x = 0} \;\subset\; \partial^L F(x)\big|_{x=\mathring{x}}$$

contains those Jacobians $\partial F_\sigma(\mathring{x})$ for which the tangent cone

$$T_\sigma \;\equiv\; T_{\mathring{x}}\{x \in \mathcal{D} : F_\sigma(x) = F(x)\}$$

has a nonempty interior. (i.e. F_σ and ∂F_σ are conically active)

Remark

We can find several of them at cost $n\,OPS(F)$ in worst case.
All of them likely a stretch, there could be 2^s different ones.

Nonsmooth equation solving

Hope:

$$F(\mathring{x}) \quad = \quad 0 \quad \text{and} \quad \partial^{??}F(\mathring{x}) \text{ 'invertible'}$$
$$\Longrightarrow$$
$$F(x) \quad = \quad y \approx 0 \quad \text{solvable by} \quad x \mathrel{+}= \partial^{??}F(\mathring{x})^{-1}(y - F(x))$$

Snag:

(Uniform) Invertibility of $F'(\mathring{x};\cdot)$ and coherent orientation of $\Delta F(\mathring{x};\cdot)$ is not stable w.r.t small perturbations in $\mathring{x}$.

Saving grace by Scholtes:

Invertibility of $F'(\mathring{x};\cdot) \implies$ openness of F at $\mathring{x}$,
i.e. nonunique solvability of $F(x) = y \approx 0$, but how to realize?

Nonsmooth equation solving

Hope:

$$F(\mathring{x}) = 0 \quad \text{and} \quad \partial^{??}F(\mathring{x}) \text{ 'invertible'}$$

$$\Longrightarrow$$

$$F(x) = y \approx 0 \quad \text{solvable by} \quad x \mathrel{+}= \partial^{??}F(\mathring{x})^{-1}(y - F(x))$$

Snag:

(Uniform) Invertibility of $F'(\mathring{x};\cdot)$ and coherent orientation of $\Delta F(\mathring{x};\cdot)$ is not stable w.r.t small perturbations in $\mathring{x}$.

Saving grace by Scholtes:

Invertibility of $F'(\mathring{x};\cdot) \Longrightarrow$ openness of F at $\mathring{x}$,
i.e. nonunique solvability of $F(x) = y \approx 0$, but how to realize?

Nonsmooth equation solving

Hope:

$$F(\mathring{x}) \quad = \quad 0 \quad \text{and} \quad \partial^{??}F(\mathring{x}) \text{ 'invertible'}$$

$$\Longrightarrow$$

$$F(x) \quad = \quad y \approx 0 \quad \text{solvable by} \quad x \mathrel{+}= \partial^{??}F(\mathring{x})^{-1}(y - F(x))$$

Snag:

(Uniform) Invertibility of $F'(\mathring{x};\cdot)$ and coherent orientation of $\Delta F(\mathring{x};\cdot)$ is not stable w.r.t small perturbations in $\mathring{x}$.

Saving grace by Scholtes:

Invertibility of $F'(\mathring{x};\cdot) \implies$ openness of F at $\mathring{x}$,
i.e. nonunique solvability of $F(x) = y \approx 0$, but how to realize?

Optimization with quadratic overestimation

Under our assumption there exists for given level set

$$\mathcal{N}_0 \;\equiv\; \{x \in \mathbb{R}^n : f(x) \leq f(x_0)\}$$

$$\hat{q}(x,s) \;\equiv\; |f(x+s) - f(x) - \Delta f(x;s)| \,/ \|s\|^2 \;\leq\; \bar{q}(\|s\|)$$

Consequence:

$$\begin{aligned} \Delta x \quad &\equiv \quad \underset{s}{\mathbf{argmin}}(\Delta f(x;s) + q\|s\|^2) \\ x \quad &\mathrel{+}= \quad \Delta x \quad \mathit{if} \quad f(x+\Delta x) < f(x) \\ q_+ \quad &= \quad \max(q, \hat{q}(x, \Delta x)) \end{aligned}$$

has stationary cluster point x_*, i.e.

$$\Delta f(x_*;s) \;\geq\; 0 \quad \text{for} \quad s \in \mathbb{R}^n$$

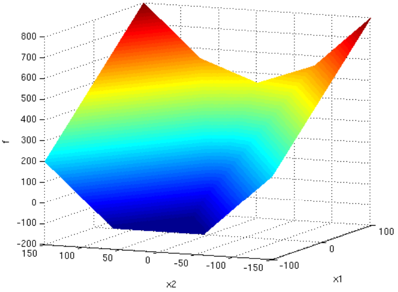
Local = Inner Problem

$$\min_{s \in \mathbb{R}^n} \Delta f(x; s) + \frac{q}{2}\|s\|^2$$

Here we only look at

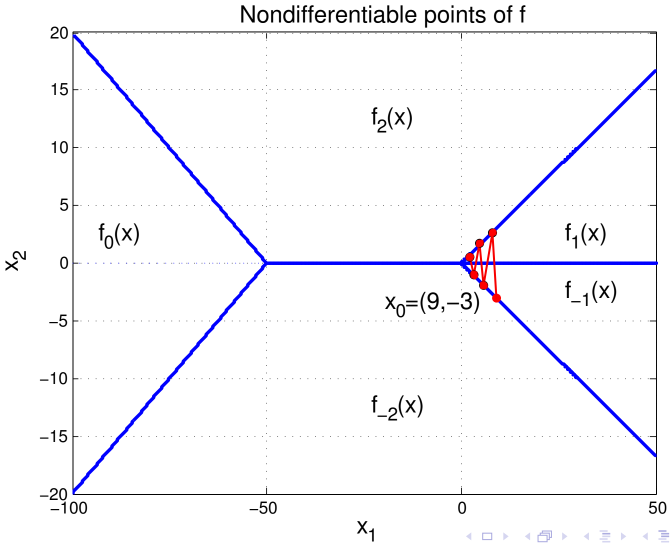
$$\min_{x \in \mathbb{R}^n} f(x)$$

with f continuous and PL.



- At least, global minimization is NP-hard ($\leftarrow$ SAT3)
- Classical steepest descent with exact line search may fail even when f is convex as demonstrated by Bonnans et al.

Zig-zagging of Steepest Descent



True steepest descent

Let the search trajectory with starting point $x(0) = x_0$ be defined as

$$-\dot{x} = -d(x) \equiv \mathbf{short}(\partial f(x)) \equiv \mathbf{argmin}\{\|g\| : g \in \partial f(x)\}$$

which are particular solution of differential inclusion $\dot{x} \in -\partial f(x)$.

Solution $x(t)$ satisfies a.e. $g(x(t))^\top \dot{x}(t) \leq -\|d(x(t))\|^2 \implies$

$$0 \;=\; \operatorname*{\mathbf{essinf}}_{t>0} \left| \frac{d}{dt} f(x(t)) \right| \;\geq\; \operatorname*{\mathbf{essinf}}_{t>0} \|d(x(t))\|^2$$

because $f \geq f_*$ on the bounded level set $\mathcal{N}_0 = f^{-1}[f_*, f(x_0)]$.
Thus $x(t)$ has stationary cluster point or limit $x_* \in \mathcal{N}_0$.

Problem: Zeno behaviour possible, i.e. a trajectory that includes an infinite number of direction changes in a finite amount of time.

Implementation if PL Case

Abs-normal form yields for any pair $x, d \neq 0$

- *directionally active gradient* $g = \nabla f(x, d)$
- *a maximal multiplier* $t_c \in [0, \infty]$ s.t.

$$g \in \partial f(x) \quad \text{and} \quad f(x) + t\, g^\top d = f(x + td) \quad \text{for} \quad 0 \leq t \leq t_c$$

Use bundle subset $G \subset \partial^L f(x)$

define direction as $d = -\mathbf{short}(G) \equiv -\mathbf{argmin}\{\|g\| : g \in G\}$
make sure $\nabla f(x, d) \in G$ before taking serious step
and reduce subsequently $G = \{g \in G : g^T d = -\|d\|^2\}$

Proposition (Griewank, Walther)

Finite convergence to minimizer if $f \in C(\mathbb{R}^n)$ convex and PL.

Implementation if PL Case

Abs-normal form yields for any pair $x, d \neq 0$

- *directionally active gradient* $g = \nabla f(x, d)$
- *a maximal multiplier* $t_c \in [0, \infty]$ s.t.

$$g \in \partial f(x) \quad \text{and} \quad f(x) + t\, g^\top d = f(x + td) \quad \text{for} \quad 0 \leq t \leq t_c$$

Use bundle subset $G \subset \partial^L f(x)$

define direction as $d = -\mathbf{short}(G) \equiv -\mathbf{argmin}\{\|g\| : g \in G\}$
make sure $\nabla f(x, d) \in G$ before taking serious step
and reduce subsequently $G = \{g \in G : g^T d = -\|d\|^2\}$

Proposition (Griewank, Walther)

Finite convergence to minimizer if $f \in C(\mathbb{R}^n)$ convex and PL.

Implementation if PL Case

Abs-normal form yields for any pair $x, d \neq 0$

- *directionally active gradient* $g = \nabla f(x, d)$
- *a maximal multiplier* $t_c \in [0, \infty]$ s.t.

$$g \in \partial f(x) \quad \text{and} \quad f(x) + t\, g^\top d = f(x + td) \quad \text{for} \quad 0 \leq t \leq t_c$$

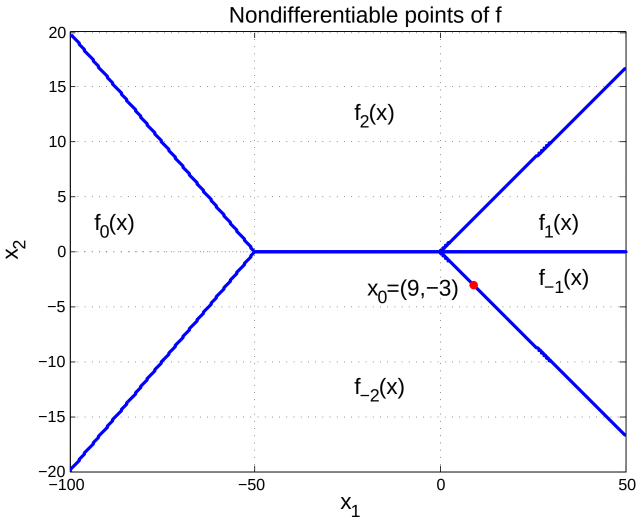
Use bundle subset $G \subset \partial^L f(x)$

define direction as $d = -\mathbf{short}(G) \equiv -\mathbf{argmin}\{\|g\| : g \in G\}$
make sure $\nabla f(x, d) \in G$ before taking serious step
and reduce subsequently $G = \{g \in G : g^T d = -\|d\|^2\}$

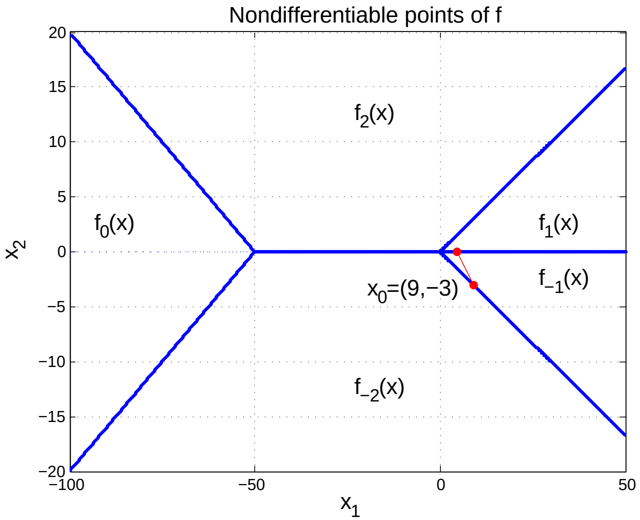
Proposition (Griewank, Walther)

Finite convergence to minimizer if $f \in C(\mathbb{R}^n)$ convex and PL.

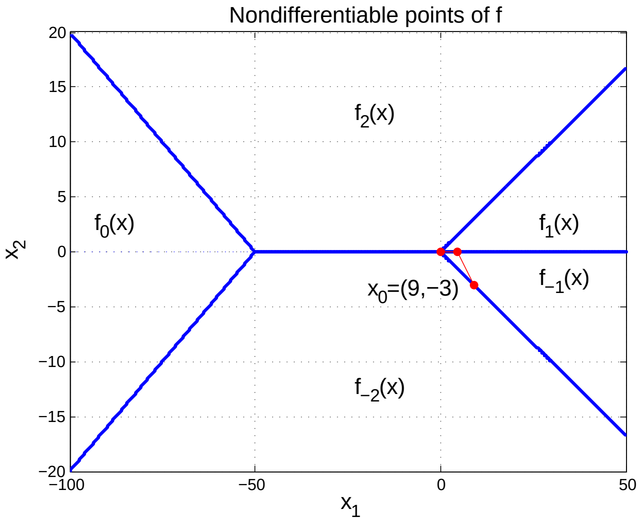
Iteration 1


Nondifferentiable points of f
20
15
10
5
0
−5
−10
−15
−20
−100
−50
0
50
x_2
x_1
$f_2(x)$
$f_0(x)$
$f_1(x)$
$f_{-1}(x)$
$x_0=(9,-3)$
$f_{-2}(x)$

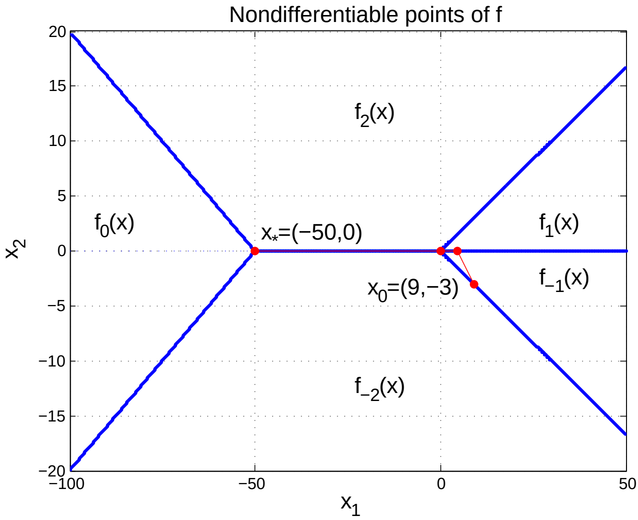
Iteration 2

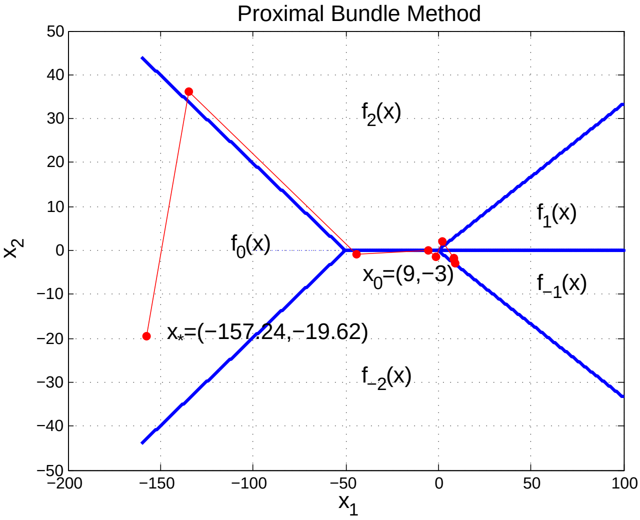

Nondifferentiable points of f
$f_2(x)$
$f_0(x)$
$f_1(x)$
$f_{-1}(x)$
$x_0=(9,-3)$
$f_{-2}(x)$
x_2
x_1
20
15
10
5
0
−5
−10
−15
−20
−100
−50
0
50

Iteration 3


Nondifferentiable points of f
20
15
10
5
0
−5
−10
−15
−20
−100
−50
0
50
x2
x1
f_2(x)
f_0(x)
f_1(x)
f_-1(x)
f_-2(x)
x0=(9,−3)

Iteration 4: Reached optimal point




Proximal Bundle Method
$f_2(x)$
$f_1(x)$
$f_0(x)$
$x_0=(9,-3)$
$f_{-1}(x)$
$x_*=(-157.24,-19.62)$
$f_{-2}(x)$
x_2
x_1
50
40
30
20
10
0
−10
−20
−30
−40
−50
−200
−150
−100
−50
0
50
100

ODE integration with Lipschitzian RHS

Possibly after space discretization of PDE:

$$\dot{x} \equiv \frac{d}{dt}x(t) = F(x(t)) \quad \text{with} \quad F \in \mathcal{C}^{0,1} = W^{1,\infty}$$

Generalized midpoint rule

With $\check{x}$ current point, $\hat{x}$ next point, $\mathring{x} = (\check{x} + \hat{x})/2$ and step h

$$\hat{x} - \check{x} = h \int_{-1/2}^{1/2} [F(\mathring{x}) + \Delta F\,(\mathring{x}; (\hat{x} - \check{x})\, t)]\, dt$$

maintains global second order with automatic event handling, realizable by Picard if $1 > h\, Lipschitz(RHS)$, i.e. nonstiffness.

Rolling Stone

$$\ddot{x} = -V'(x) \quad \text{with}$$

$$V(x) = \begin{cases} \frac{1}{2}(1-x)^2 & \text{if } x \geq 1 \\ \frac{1}{2}(1+x)^2 & \text{if } x \leq -1 \\ 0 & \text{else} \end{cases}$$

$$-V'(x) = \begin{cases} 1-x & \text{if } x \geq 1 \\ -x-1 & \text{if } x \leq -1 \\ 0 & \text{else} \end{cases}$$

$$= \min(\max(-x-1, 0), 1-x)$$

$$= -x - \tfrac{1}{2}|x-1| + \tfrac{1}{2}|x+1|$$

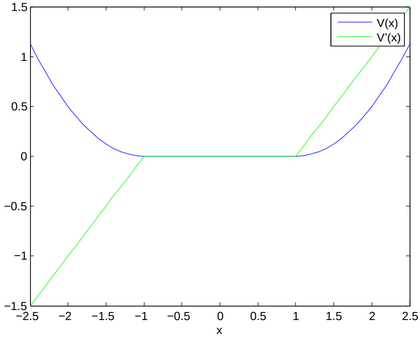


Figure: Rolling Stones Right hand side

Exact solution

$x(0) = 1, \quad \dot{x}(0) = 1$

$$x(t) = \begin{cases} 1 + \sin(t) & \text{if } t \in [0, \pi) \\ 1 - (t - \pi) & \text{if } t \in [\pi, \pi + 2) \\ -2 - \sin(2 - t) & \text{if } t \in [\pi + 2, 2\pi + 2) \\ t - 3 - 2\pi & \text{if } t \in [2\pi + 2, 2\pi + 4) \end{cases}$$

The total period is $2\pi + 4$.

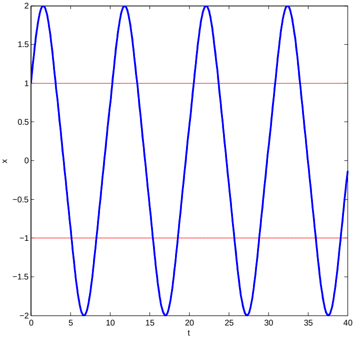
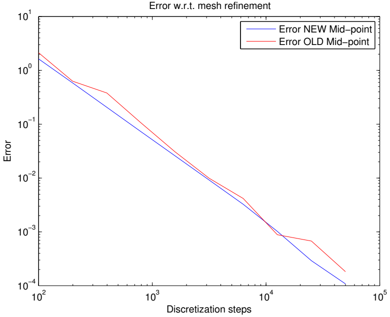
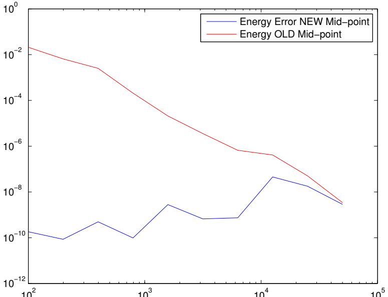


Figure: Exact solution


Error w.r.t. mesh refinement
Error NEW Mid−point
Error OLD Mid−point
Error
Discretization steps


Energy Error NEW Mid−point
Energy OLD Mid−point

Problem Definition

$$F(x) = \begin{pmatrix} \dot{x} \\ \dot{y} \\ \dot{z} \end{pmatrix} = \begin{pmatrix} \alpha(y - x - f(x)) \\ x - y + z \\ -\beta y \end{pmatrix}$$

$$f(x) = m_1 x + \frac{1}{2}(m_0 - m_1)(|x+1| - |x-1|)$$

- x, y are the voltages across C_1 and C_2
- z is the intensity of the electrical current at I
- $f(x)$ is the electrical response of the resistor
- constants are $\alpha = 15.6, \beta = 28, m_0 = -1.143, m_1 = -0.714$

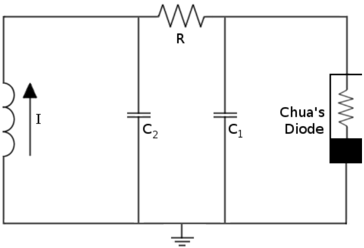
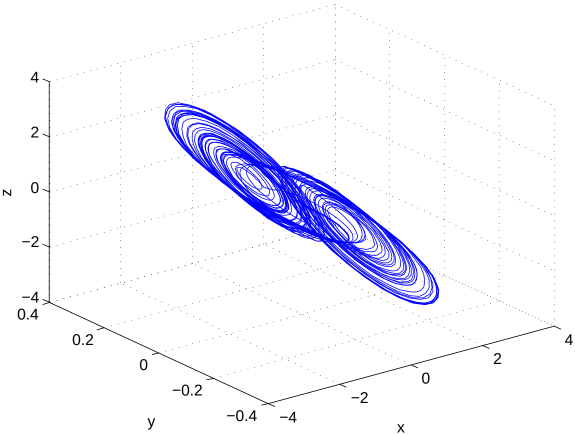


Figure: Chua circuit

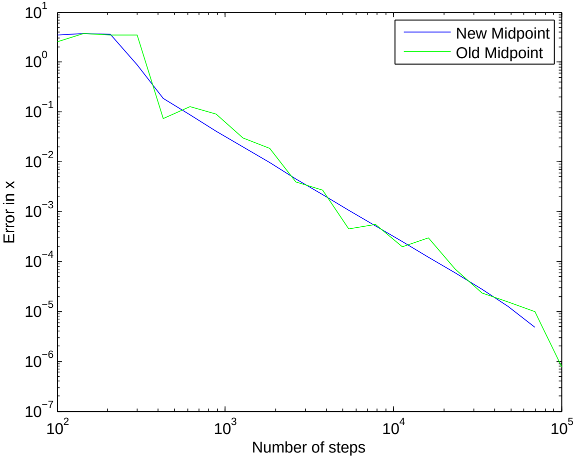
taken from

`http://www.chuacircuits.com/`

Chua circuit



Convergence



When are Hessians symmetric ??

- Euler, Clairault, Bernoulli, Cauchy, and others tried to prove that matrices of second derivatives are symmetric.
- Lindelöf demonstrated in 1857 that all their assertions and/or proofs were wrong. Beginner's analysis errors !!
- A. H. Schwarz, student of Weierstrass proved in 1863

$$g = \nabla f \in C^1(D) \implies (g')^\top = g' = \nabla^2 f$$

- Peano provided counter example where in 'some sense'

$$\nabla^2 f(0,0) = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \quad \text{for} \quad f(x,y) = x\,y\,\frac{(x^2-y^2)}{(x^2+y^2)}$$

When are Hessians symmetric ??

- Euler, Clairault, Bernoulli, Cauchy, and others tried to prove that matrices of second derivatives are symmetric.
- Lindelöf demonstrated in 1857 that all their assertions and/or proofs were wrong. Beginner's analysis errors !!
- A. H. Schwarz, student of Weierstrass proved in 1863

$$g = \nabla f \in C^1(D) \implies (g')^\top = g' = \nabla^2 f$$

- Peano provided counter example where in 'some sense'

$$\nabla^2 f(0,0) = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \quad \text{for} \quad f(x,y) = x\,y\,\frac{(x^2 - y^2)}{(x^2 + y^2)}$$

When are Hessians symmetric ??

- Euler, Clairault, Bernoulli, Cauchy, and others tried to prove that matrices of second derivatives are symmetric.
- Lindelöf demonstrated in 1857 that all their assertions and/or proofs were wrong. Beginner's analysis errors !!
- A. H. Schwarz, student of Weierstrass proved in 1863

$$g = \nabla f \in C^1(D) \implies (g')^\top = g' = \nabla^2 f$$

- Peano provided counter example where in 'some sense'

$$\nabla^2 f(0,0) = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \quad \text{for} \quad f(x,y) = x\,y\,\frac{(x^2 - y^2)}{(x^2 + y^2)}$$

When are Hessians symmetric ??

- Euler, Clairault, Bernoulli, Cauchy, and others tried to prove that matrices of second derivatives are symmetric.
- Lindelöf demonstrated in 1857 that all their assertions and/or proofs were wrong. Beginner's analysis errors !!
- A. H. Schwarz, student of Weierstrass proved in 1863

$$g = \nabla f \in \mathcal{C}^1(D) \implies (g')^\top = g' = \nabla^2 f$$

- Peano provided counter example where in 'some sense'

$$\nabla^2 f(0,0) = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \quad \text{for} \quad f(x,y) = x\,y\,\frac{(x^2-y^2)}{(x^2+y^2)}$$

When are Hessians symmetric ??

- Euler, Clairault, Bernoulli, Cauchy, and others tried to prove that matrices of second derivatives are symmetric.
- Lindelöf demonstrated in 1857 that all their assertions and/or proofs were wrong. Beginner's analysis errors !!
- A. H. Schwarz, student of Weierstrass proved in 1863

$$g = \nabla f \in \mathcal{C}^1(D) \implies (g')^\top = g' = \nabla^2 f$$

- Peano provided counter example where in 'some sense'

$$\nabla^2 f(0,0) = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \quad \text{for} \quad f(x,y) = x\,y\,\frac{(x^2 - y^2)}{(x^2 + y^2)}$$

When are Hessians symmetric ??

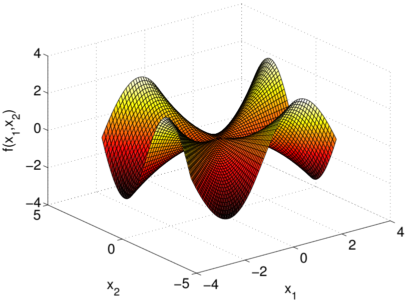
- Euler, Clairault, Bernoulli, Cauchy, and others tried to prove that matrices of second derivatives are symmetric.
- Lindelöf demonstrated in 1857 that all their assertions and/or proofs were wrong. Beginner's analysis errors !!
- A. H. Schwarz, student of Weierstrass proved in 1863

$$g = \nabla f \in \mathcal{C}^1(D) \quad \Longrightarrow \quad (g')^\top = g' = \nabla^2 f$$

- Peano provided counter example where in 'some sense'

$$\nabla^2 f(0,0) = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \quad \text{for} \quad f(x,y) \; = \; x\,y\,\frac{(x^2 - y^2)}{(x^2 + y^2)}$$

Look at the Peano Example



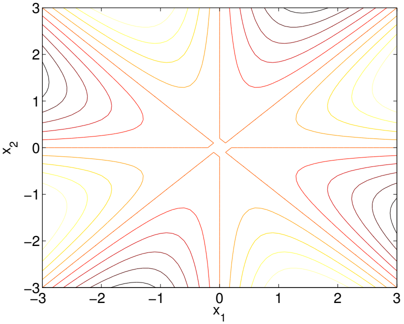


Figure: Peano function and its contour plot

Real Hessians are always symmetric !!

- Peano Hessian is algebraic fluke, not a Fréchet derivative:

$$g(x + \Delta x) - g(x) \neq g'(x)\,\Delta x + o(\|\Delta x\|)$$

- Dieudonné (1960) showed that derivatives of gradients are symmetric where they exist $\Longleftrightarrow$ No Perpetuum Mobile !!
- Limiting and Convexification maintain: $(\partial^C g)^\top = \partial^C g$
- Griewank et al (2013) are showing the converse, i.e.

$$g \in \mathcal{C}^{0,1}(D) \text{ with } (\partial^C g)^\top = \partial^C g \quad \Longrightarrow \quad g = \nabla f$$

Real Hessians are always symmetric !!

- Peano Hessian is algebraic fluke, not a Fréchet derivative:

$$g(x+\Delta x) - g(x) \neq g'(x)\,\Delta x + o(\|\Delta x\|)$$

- Dieudonné (1960) showed that derivatives of gradients are symmetric where they exist $\Longleftrightarrow$ No Perpetuum Mobile !!
- Limiting and Convexification maintain: $(\partial^C g)^\top = \partial^C g$
- Griewank et al (2013) are showing the converse, i.e.

$$g \in C^{0,1}(D) \text{ with } (\partial^C g)^\top = \partial^C g \quad \Longrightarrow \quad g = \nabla f$$

Real Hessians are always symmetric !!

- Peano Hessian is algebraic fluke, not a Fréchet derivative:
$$g(x + \Delta x) - g(x) \neq g'(x)\,\Delta x + o(\|\Delta x\|)$$
- Dieudonné (1960) showed that derivatives of gradients are symmetric where they exist $\iff$ No Perpetuum Mobile !!
- Limiting and Convexification maintain: $(\partial^C g)^\top = \partial^C g$
- Griewank et al (2013) are showing the converse, i.e.
$$g \in C^{0,1}(D) \text{ with } (\partial^C g)^\top = \partial^C g \quad \Longrightarrow \quad g = \nabla f$$

Real Hessians are always symmetric !!

- Peano Hessian is algebraic fluke, not a Fréchet derivative:

$$g(x+\Delta x) - g(x) \neq g'(x)\,\Delta x + o(\|\Delta x\|)$$

- Dieudonné (1960) showed that derivatives of gradients are symmetric where they exist $\Longleftrightarrow$ No Perpetuum Mobile !!
- Limiting and Convexification maintain: $(\partial^C g)^\top = \partial^C g$
- Griewank et al (2013) are showing the converse, i.e.

$$g \in C^{0,1}(D) \text{ with } (\partial^C g)^\top = \partial^C g \quad \Longrightarrow \quad g = \nabla f$$

Real Hessians are always symmetric !!

- Peano Hessian is algebraic fluke, not a Fréchet derivative:

$$g(x+\Delta x) - g(x) \neq g'(x)\,\Delta x + o(\|\Delta x\|)$$

- Dieudonné (1960) showed that derivatives of gradients are symmetric where they exist $\iff$ No Perpetuum Mobile !!
- Limiting and Convexification maintain: $(\partial^C g)^\top = \partial^C g$
- Griewank et al (2013) are showing the converse, i.e.

$$g \in \mathcal{C}^{0,1}(D) \text{ with } (\partial^C g)^\top = \partial^C g \quad \Longrightarrow \quad g = \nabla f$$

Real Hessians are always symmetric !!

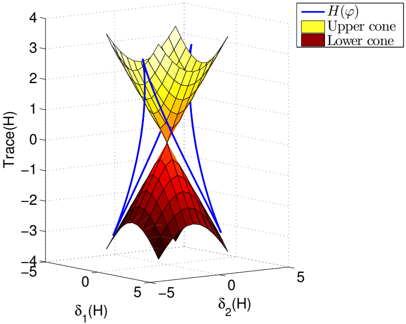
- Peano Hessian is algebraic fluke, not a Fréchet derivative:

$$g(x+\Delta x) - g(x) \neq g'(x)\,\Delta x + o(\|\Delta x\|)$$

- Dieudonné (1960) showed that derivatives of gradients are symmetric where they exist $\Longleftrightarrow$ No Perpetuum Mobile !!
- Limiting and Convexification maintain: $(\partial^C g)^\top = \partial^C g$
- Griewank et al (2013) are showing the converse, i.e.

$$g \in \mathcal{C}^{0,1}(D) \text{ with } (\partial^C g)^\top = \partial^C g \quad \Longrightarrow \quad g = \nabla f$$

Generalized Hessian of Peano



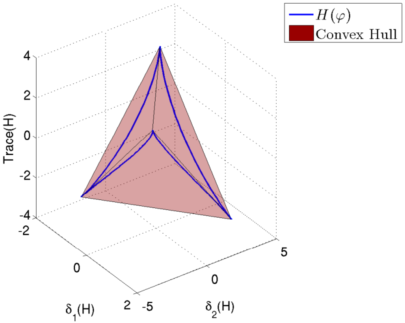


Figure: Hessians circling definite cone with four extremal cusps

Summary and tentative conclusions

- Practical functions are semi-smooth and their linearization goes further than we thought, but not quite far enough.
- Yes, we can compute generalized Jacobians! They are not only essential in the sense of Scholtes but conically active.
- But, semi-smooth Newton only yields convergence from points where combinatorial aspects have been resolved.
- Piecewise linearization facilitates nonsmooth equation solving, optimization, integration of Lipschitzian ODEs...
- Lipschitzian gradients have symmetric generalized Hessians that are computable by piecewise linearization
- Next on the agenda: solving algebraic and differential inclusions as well as bang-bang optimal control problems.

Summary and tentative conclusions

- Practical functions are semi-smooth and their linearization goes further than we thought, but not quite far enough.
- Yes, we can compute generalized Jacobians! They are not only essential in the sense of Scholtes but conically active.
- But, semi-smooth Newton only yields convergence from points where combinatorial aspects have been resolved.
- Piecewise linearization facilitates nonsmooth equation solving, optimization, integration of Lipschitzian ODEs...
- Lipschitzian gradients have symmetric generalized Hessians that are computable by piecewise linearization
- Next on the agenda: solving algebraic and differential inclusions as well as bang-bang optimal control problems.

Summary and tentative conclusions

- Practical functions are semi-smooth and their linearization goes further than we thought, but not quite far enough.
- Yes, we can compute generalized Jacobians! They are not only essential in the sense of Scholtes but conically active.
- But, semi-smooth Newton only yields convergence from points where combinatorial aspects have been resolved.
- Piecewise linearization facilitates nonsmooth equation solving, optimization, integration of Lipschitzian ODEs...
- Lipschitzian gradients have symmetric generalized Hessians that are computable by piecewise linearization
- Next on the agenda: solving algebraic and differential inclusions as well as bang-bang optimal control problems.

Summary and tentative conclusions

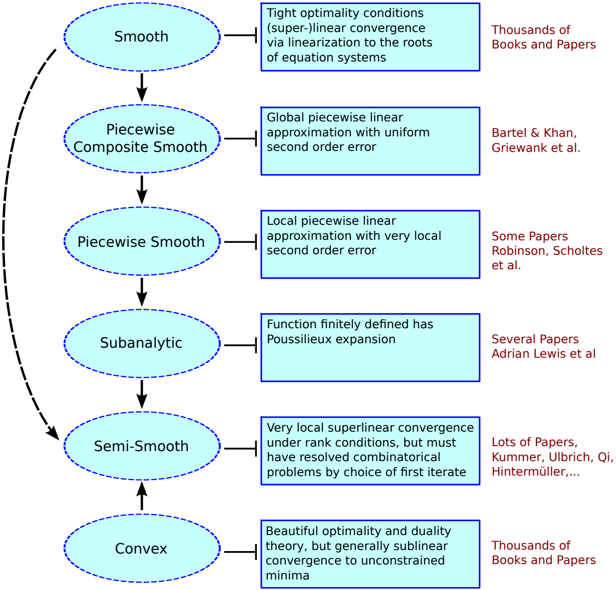
- Practical functions are semi-smooth and their linearization goes further than we thought, but not quite far enough.
- Yes, we can compute generalized Jacobians! They are not only essential in the sense of Scholtes but conically active.
- But, semi-smooth Newton only yields convergence from points where combinatorial aspects have been resolved.
- Piecewise linearization facilitates nonsmooth equation solving, optimization, integration of Lipschitzian ODEs...
- Lipschitzian gradients have symmetric generalized Hessians that are computable by piecewise linearization
- Next on the agenda: solving algebraic and differential inclusions as well as bang-bang optimal control problems.

Summary and tentative conclusions

- Practical functions are semi-smooth and their linearization goes further than we thought, but not quite far enough.
- Yes, we can compute generalized Jacobians! They are not only essential in the sense of Scholtes but conically active.
- But, semi-smooth Newton only yields convergence from points where combinatorial aspects have been resolved.
- Piecewise linearization facilitates nonsmooth equation solving, optimization, integration of Lipschitzian ODEs...
- Lipschitzian gradients have symmetric generalized Hessians that are computable by piecewise linearization
- Next on the agenda: solving algebraic and differential inclusions as well as bang-bang optimal control problems.

Summary and tentative conclusions

- Practical functions are semi-smooth and their linearization goes further than we thought, but not quite far enough.
- Yes, we can compute generalized Jacobians! They are not only essential in the sense of Scholtes but conically active.
- But, semi-smooth Newton only yields convergence from points where combinatorial aspects have been resolved.
- Piecewise linearization facilitates nonsmooth equation solving, optimization, integration of Lipschitzian ODEs...
- Lipschitzian gradients have symmetric generalized Hessians that are computable by piecewise linearization
- Next on the agenda: solving algebraic and differential inclusions as well as bang-bang optimal control problems.


Smooth
Tight optimality conditions (super-)linear convergence via linearization to the roots of equation systems
Thousands of Books and Papers
Piecewise Composite Smooth
Global piecewise linear approximation with uniform second order error
Bartel & Khan, Griewank et al.
Piecewise Smooth
Local piecewise linear approximation with very local second order error
Some Papers Robinson, Scholtes et al.
Subanalytic
Function finitely defined has Poussilieux expansion
Several Papers Adrian Lewis et al
Semi-Smooth
Very local superlinear convergence under rank conditions, but must have resolved combinatorical problems by choice of first iterate
Lots of Papers, Kummer, Ulbrich, Qi, Hintermüller,...
Convex
Beautiful optimality and duality theory, but generally sublinear convergence to unconstrained minima
Thousands of Books and Papers

Final greetings from Prof. Moriarty