Between Piecewise Smoothness and Linear Complementarity

Andreas Griewank,

with thanks to

A. Walther, T. Bosse, N. Strogies, S. Fiege, F. Kerkoff, J.U. Bernt, M. Radons,T. Streubel, R. Hasenfelder, P. Boeck, B. Lenser, ...

Institute for Applied Mathematics, Humboldt Universität zu Berlin, Germany

griewank@math.hu-berlin.de

6th International Conference on Complementarity Problems

August 4-8, 2014, Berlin.

August 8, 2014

1. Observations/Opinions of a Johnny Come Lately
2. Piecewise Linearization/Differentiation
3. Representation of PL functions in abs-normal form
4. Computation of conical Jacobians and gradients
5. Solving PL systems of Equation and LCPs
6. (Un)constrained optimization by successive PL
7. Integration of Lipschitzian dynamics

Why not smoothen/regularize/parametrize?

- If scale too small it makes no algorithmic difference.

Why not smoothen/regularize/parametrize?

- If scale too small it makes no algorithmic difference.
- If scale too large the problem is significantly changed.

Why not smoothen/regularize/parametrize?

- If scale too small it makes no algorithmic difference.
- If scale too large the problem is significantly changed.
- If scale variable we have one more loose parameter.

Why not smoothen/regularize/parametrize?

- If scale too small it makes no algorithmic difference.
- If scale too large the problem is significantly changed.
- If scale variable we have one more loose parameter.
- Resulting solution paths may have sharp turns (LOP).

Why not smoothen/regularize/parametrize?

- If scale too small it makes no algorithmic difference.
- If scale too large the problem is significantly changed.
- If scale variable we have one more loose parameter.
- Resulting solution paths may have sharp turns (LOP).
- There may be spurious solutions (Bernardo et al).

Why not smoothen/regularize/parametrize?

- If scale too small it makes no algorithmic difference.
- If scale too large the problem is significantly changed.
- If scale variable we have one more loose parameter.
- Resulting solution paths may have sharp turns (LOP).
- There may be spurious solutions (Bernardo et al).

Lesson/Moral:

Let's face the combinatorial music!

(Reflected in the piecewise linearization)

My issues with generalized differentiation

- For composite nonsmoothness differentiation rules are only inclusions so analysis cannot be turned into algebra in an AD fashion.

My issues with generalized differentiation

- For composite nonsmoothness differentiation rules are only inclusions so analysis cannot be turned into algebra in an AD fashion.

- The inclusions point the right way for propagating semi-smoothness but the wrong way for calculating generalized derivative elements.

My issues with generalized differentiation

- For composite nonsmoothness differentiation rules are only inclusions so analysis cannot be turned into algebra in an AD fashion.

- The inclusions point the right way for propagating semi-smoothness but the wrong way for calculating generalized derivative elements.

- Semi-smooth Newton extremely local, bundle methods too heuristic.

My issues with generalized differentiation

- For composite nonsmoothness differentiation rules are only inclusions so analysis cannot be turned into algebra in an AD fashion.

- The inclusions point the right way for propagating semi-smoothness but the wrong way for calculating generalized derivative elements.

- Semi-smooth Newton extremely local, bundle methods too heuristic.

- Generalized derivatives are fickle since outer semi-continuity of multi-functions does not mean stability in any numerical sense.

My issues with generalized differentiation

- For composite nonsmoothness differentiation rules are only inclusions so analysis cannot be turned into algebra in an AD fashion.

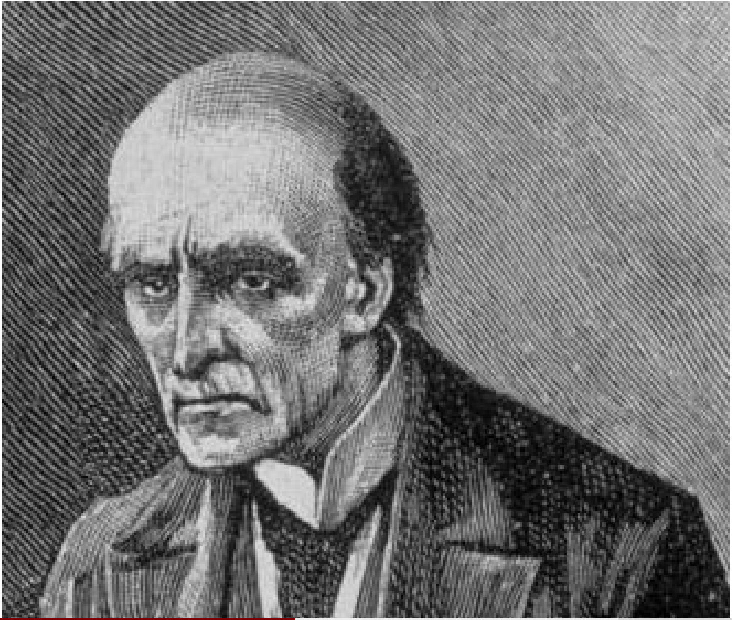
- The inclusions point the right way for propagating semi-smoothness but the wrong way for calculating generalized derivative elements.

- Semi-smooth Newton extremely local, bundle methods too heuristic.

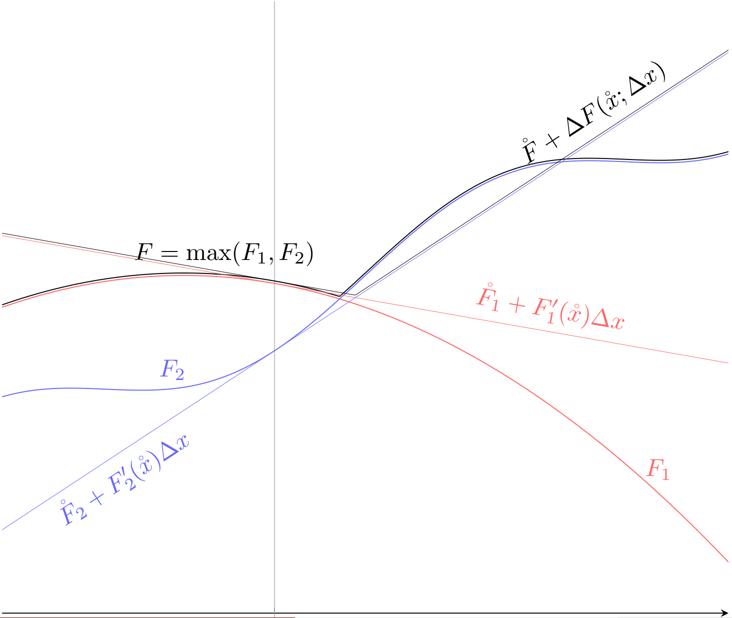
- Generalized derivatives are fickle since outer semi-continuity of multi-functions does not mean stability in any numerical sense.

- Rademacher says $F \in C^{0,1}(\mathbb{R}^n) \equiv W^{1,\infty}(\mathbb{R}^n)$, hence generalized derivatives are almost everywhere normal derivatives.

Lurking in the background: Prof. Moriarty



Basic idea of tangent linearization:



abs covers min, max and table look-ups

Provided u and w are both finite one has

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

data (x_i, y_i) for $0 \le i \le n$ are piecewise linearly interpolated by the formula

$$\begin{aligned} y &= \tfrac{1}{2}\left[y_0 + s_1\,\mathbf{abs}(x-x_0) + y_n + s_n\,\mathbf{abs}(x-x_n)\right. \\ &+ \left.\sum_{i=1}^{n-1}(s_{i+1}-s_i)\,\mathbf{abs}(x-x_i)\right] \quad \text{whose ???} \end{aligned}$$

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

- Every continuous PL function can be expressed as composition of affine functions and several **abs**(). That representation is not unique.

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

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

and correspondingly for max() und min().

Continuous Piecewise Differentiation Rules

Linearity and Product Rule

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

$$\Longrightarrow$$

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

Continuous Piecewise Differentiation Rules

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

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

$$\Longrightarrow$$

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

Second order error and Lipschitz continuity

Proposition

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 $x, x + \Delta x \in \mathcal{K}$

$$\|F(x + \Delta x) - F(x) - \Delta F(x; \Delta x)\| \ \leq \ \gamma \|\Delta x\|^2$$

Second order error and Lipschitz continuity

Proposition

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 $x, x + \Delta x \in \mathcal{K}$

$$\|F(x + \Delta x) - F(x) - \Delta F(x; \Delta x)\| \leq \gamma \|\Delta x\|^2$$

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

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

Second order error and Lipschitz continuity

Proposition

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 $x, x + \Delta x \in \mathcal{K}$

$$\|F(x + \Delta x) - F(x) - \Delta F(x; \Delta x)\| \leq \gamma \|\Delta x\|^2$$

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

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

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

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

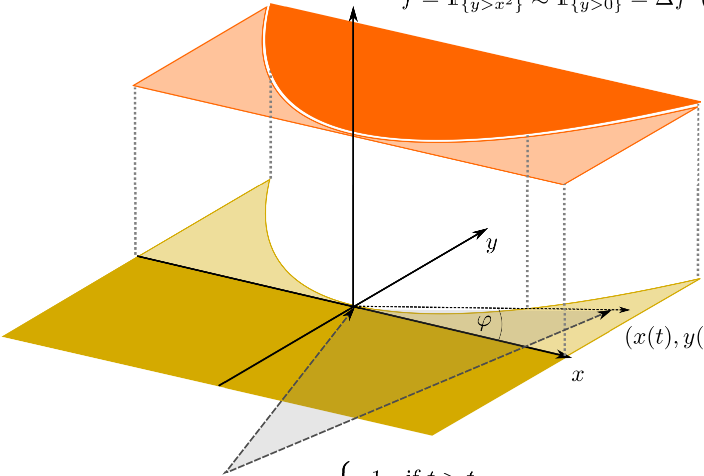
Locally we reduce to the homogeneous piecewise linear $F'(x; \Delta x)$.

Differentiation Concepts on Euclidean Spaces

Function Space:	Diff.Op.:	Model Space:	Discrepancy:
Smooth = S	$\overset{\partial\vert_{\mathring{x}}}{\underset{\text{Lip}}{\longmapsto}}$	L = linear	uniform
$\cap$		$\cap$	
CompPS = CPS	$\overset{\Delta\vert_{\mathring{x}}}{\underset{\text{Lip}}{\longmapsto}}$	PL = Piecewise L	uniform
$\cap$		$\downarrow\ \partial^B\vert_{\mathring{x}}$	
LipschitzPS = LPS	$\overset{\partial^B\vert_{\mathring{x}}}{\underset{???}{\longmapsto}}$	PL_h = homog. PL	nonuniform
$\cap$			
PiecewiseS = DCPS	$\overset{\Delta\vert_{\mathring{x}}}{\underset{???}{\longmapsto}}$	DPL = discont. PL	nonuniform

Piecewise Linearization of Discontinuous f

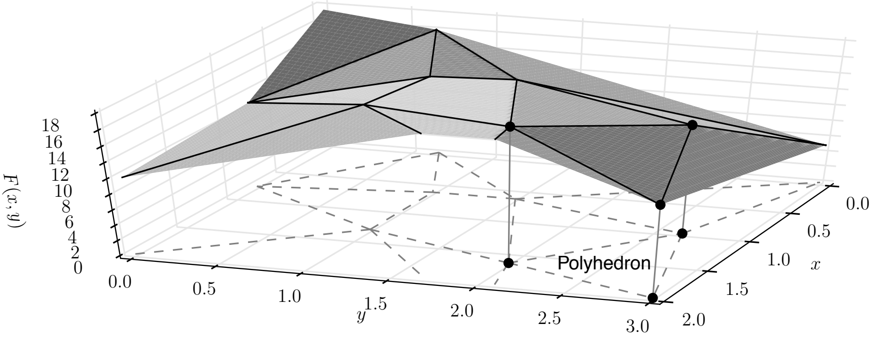
$f = \mathbb{1}_{\{y>x^2\}} \approx \mathbb{1}_{\{y>0\}} = \Delta f$ (at the origin)



$(x(t), y(t)) = t(\cos\varphi, \sin\varphi)$

$$f(x(t), y(t)) - \Delta f(x(t), y(t)) = \begin{cases} -1 & \text{if } t > t_* \\ 0 & \text{if } t < t_* \end{cases} \quad \text{where} \quad t_* = \frac{\tan(\varphi)}{\cos(\varphi)}$$

Polyhedral Decomposition



A simple $\mathbb{R}^2 \to \mathbb{R}^2$ example:

$$f_1 = x_1 + |x_1 - x_2| + |x_1 - |x_2||, f_2 = x_2$$

The switching variables z_i are the arguments of the abs-functions:

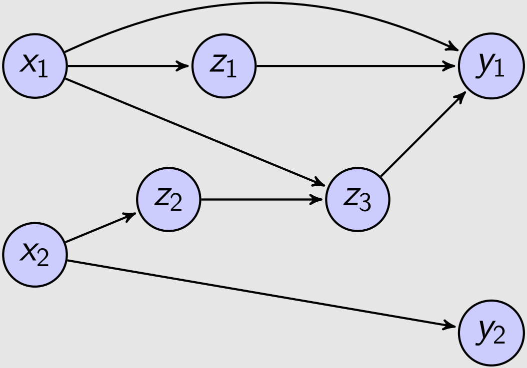
$$z_1 = x_1 - x_2, z_2 = x_2, z_3 = x_1 - |z_2| \Rightarrow f_1 = x_1 + |z_1| + |z_3|$$

In Abs-normal form:

$$\begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ y_1 \\ y_2 \end{bmatrix} = \left[\begin{array}{cc|ccc} \overbrace{\begin{matrix} 1 & -1 \end{matrix}}^{Z} & & \overbrace{\begin{matrix} 0 & 0 & 0 \end{matrix}}^{L} & & \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 \\ \hline 1 & 0 & 1 & 0 & 1 \\ \underbrace{\begin{matrix} 0 & 1 \end{matrix}}_{J} & & \underbrace{\begin{matrix} 0 & 0 & 0 \end{matrix}}_{Y} & & \end{array}\right] \begin{bmatrix} x_1 \\ x_2 \\ |z_1| \\ |z_2| \\ |z_3| \end{bmatrix}$$

Graph reprsentation and switching depth

Computational graph



Proposition (Griewank et al 2014)

Switching depth $\nu \equiv$ *length of directed path is bounded by* $\bar{\nu}(n) = 2\,n - 1$

Example: $\textbf{ordervalue}_k(z_1, \ldots, z_m) \;\equiv$ k$-$*th* largest z_i (N. Kreijic)

Abs-normal form for $y = F(x) : \mathbb{R}^n \to \mathbb{R}^m$

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

- $J \in \mathbb{R}^{m \times n}$ represents the smooth part of the function
- $L \in \mathbb{R}^{s \times s}$ is strictly lower triangular to yields $z = z(x)$ uniquely.
- Crossterms $Z \in \mathbb{R}^{s \times n}$ and $Y \in \mathbb{R}^{m \times s}$ link x to z and z to y.

Abs-normal form for $y = F(x) : \mathbb{R}^n \to \mathbb{R}^m$

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

- $J \in \mathbb{R}^{m \times n}$ represents the smooth part of the function
- $L \in \mathbb{R}^{s \times s}$ is strictly lower triangular to yields $z = z(x)$ uniquely.
- Crossterms $Z \in \mathbb{R}^{s \times n}$ and $Y \in \mathbb{R}^{m \times s}$ link x to z and z to y.

- The switching depth ν is the structural nilpotency degree of L.
- Abs-normal form is nonredundant and stable w.r.t. perturbations.
- ADOL-C can calculate $[b, c, Z, L, J, Y]$ after slight modification.

Abs-normal form for $y = F(x) : \mathbb{R}^n \to \mathbb{R}^m$

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

- $J \in \mathbb{R}^{m \times n}$ represents the smooth part of the function
- $L \in \mathbb{R}^{s \times s}$ is strictly lower triangular to yields $z = z(x)$ uniquely.
- Crossterms $Z \in \mathbb{R}^{s \times n}$ and $Y \in \mathbb{R}^{m \times s}$ link x to z and z to y.

- The switching depth ν is the structural nilpotency degree of L.
- Abs-normal form is nonredundant and stable w.r.t. perturbations.
- ADOL-C can calculate $[b, c, Z, L, J, Y]$ after slight modification.

- The sign vector $\sigma \equiv \mathbf{sign}(z(x))$ determines the control flow.
- $\Sigma \equiv \mathbf{diag}(\sigma) \implies |z| = \Sigma z$ is componentwise modulus.
- Relatively open $P_\sigma = \{x : \sigma(x) = \sigma\}$ form polyhedral skeleton.

Some other References and Traditions

Stern, Thomas Edwin (1956)

Piecewise-linear network theory

W. M. G. van Bokhoven (1981)

Piecewise-linear modelling and analysis

J. Rohn (2004)

A Theorem of the Alternatives for the equation $Ax + B|x| = b$

O. Mangasarian, and R. Meyer(2006)

Absolute value equations

L. Brugnano, and V. Casulli (2008)

Iterative solution of piecewise linear systems

A. Martin (2009)

Mixed Integer Nonlinear Programming

A. Hadjidimos, M. Lapidakis, M. and M. Tzoumas (2012)

On Iterative Solution for Linear Complementarity Problem with an H_+-Matrix

Limiting Jacobians/Gradients of PL function

Limiting (=Bouligand) Jacobian:

$$\partial^L F(x) \equiv \{J_\sigma | x \in \bar{P}_\sigma, \text{ with } P_\sigma \textbf{ open}\}$$

where

$$J_\sigma = J + Y\Sigma(I - L\Sigma)^{-1}Z$$

with

$$(I - L\Sigma)^{-1} = I + L\Sigma + (L\Sigma)^2 + \cdots + (L\Sigma)^{(\nu-1)}$$

Limiting Jacobians/Gradients of PL function

Limiting (=Bouligand) Jacobian:

$$\partial^L F(x) \;\equiv\; \{J_\sigma | x \in \bar{P}_\sigma, \text{ with } P_\sigma \textbf{ open}\}$$

where

$$J_\sigma = J + Y\Sigma(I - L\Sigma)^{-1} Z$$

with

$$(I - L\Sigma)^{-1} \;=\; I + L\Sigma + (L\Sigma)^2 + \cdots + (L\Sigma)^{(\nu-1)}$$

Polynomial escape in direction d_1

$$x(t) \;=\; x + \sum_{i=1}^{n} t^i d_i \;\in\; P_\sigma \textbf{ open}$$

when

$$\det(d_1 \ldots d_n) \neq 0 \quad \text{and} \quad 0 < t \approx 0$$

Limiting Jacobians/Gradients of PL function

Limiting (=Bouligand) Jacobian:

$$\partial^L F(x) \equiv \{J_\sigma | x \in \bar{P}_\sigma, \text{ with } P_\sigma \textbf{ open}\}$$

where

$$J_\sigma = J + Y\Sigma(I - L\Sigma)^{-1} Z$$

with

$$(I - L\Sigma)^{-1} = I + L\Sigma + (L\Sigma)^2 + \cdots + (L\Sigma)^{(\nu-1)}$$

Polynomial escape in direction d_1

$$x(t) = x + \sum_{i=1}^{n} t^i d_i \in P_\sigma \textbf{ open}$$

when

$$\det(d_1 \ldots d_n) \neq 0 \quad \text{and} \quad 0 < t \approx 0$$

Complexity range (also) utilizing reverse mode

$$3\min(m, n) \leq \texttt{OPS}(\texttt{J}_\sigma)/\texttt{OPS}(\texttt{F}) \leq 3\,\texttt{n}$$

Conical Jacobians of PS function

Proposition: Khan & Barton and A. G. 2013

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

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

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

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

Conical Jacobians of PS function

Proposition: Khan & Barton and A. G. 2013

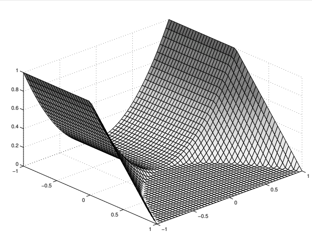
$$\partial^K F(x) \;\equiv\; \partial^L_{\Delta x} \Delta F(x;\Delta x)\Big|_{\Delta x = 0} \;\subset\; \partial^L F(x)$$

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

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

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

Kummer example for $\partial^L f \neq \partial^K f \equiv \partial^L \Delta f(x;\cdot)$



$$f(x,y) \;=\; (y^2 - x_+)_+ \;\text{ with }\; z_+ \equiv \max(0,z)$$

$$\Longrightarrow$$

$$\{0\} = \partial^K f(0) \;\not\ni\; (-1,0) \in \partial^L f(0) \subset \partial^C f(0)$$

Related perturbed semismooth system

$$F(x,y) \;=\; \begin{bmatrix} x/2 + (y^2 - x_+)_+ \\ y \end{bmatrix} \;=\; \begin{bmatrix} \delta \\ \varepsilon \end{bmatrix}$$

has for $x \leq 0$, $0 < x < y^2$, and $y^2 < x$, respectively

$$F'(x,y) = \begin{bmatrix} 1/2 & 2y \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} -1/2 & 2y \\ 0 & 1 \end{bmatrix} \text{ and } \begin{bmatrix} 1/2 & 0 \\ 0 & 1 \end{bmatrix}$$

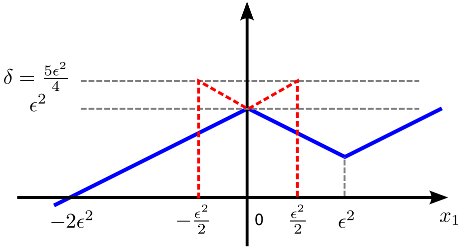


Figure: Cyclic behavior of Newton = Sucessive Linearization

Reformulations in the equation case $m = n$

Original Equation

$$0 = F(x) = b + J\,x + Y\,z(x) \quad \text{with} \quad z(x) = (I - L\Sigma)^{-1}(c + Z\,x)$$

Reformulations in the equation case $m = n$

Original Equation

$$0 \; = \; F(x) \; = \; b + J\,x + Y\,z(x) \quad \text{with} \quad z(x) = (I - L\Sigma)^{-1}(c + Z\,x)$$

Schur complement and complementary system

$\det(J) \neq 0$ (achievable using $v \equiv |v + |v|| - |v|$) ensures existence

$$S \equiv L - ZJ^{-1}Y \in \mathbb{R}^{s \times s} \qquad \Longrightarrow$$

$$z = S|z| + \hat{c} \quad \text{with} \quad \hat{c} = c - Z\,J^{-}1b$$

Reformulations in the equation case $m = n$

Original Equation

$$0 \;=\; F(x) \;=\; b + J\,x + Y\,z(x) \quad \text{with} \quad z(x) = (I - L\Sigma)^{-1}(c + Z\,x)$$

Schur complement and complementary system

$\det(J) \neq 0$ (achievable using $v \equiv |v + |v|| - |v|$) ensures existence

$$S \equiv L - ZJ^{-1}Y \in \mathbb{R}^{s\times s} \quad \Longrightarrow$$

$$z = S|z| + \hat{c} \quad \text{with} \quad \hat{c} = c - Z\,J^{-}1b$$

Corresponding Linear Complementarity Problem

$$z = u - w,\; 0 \leq u \perp w \geq 0 \quad \Longrightarrow$$

$$u \perp Mu + q \geq 0 \quad \text{with} \quad M = (I - S)^{-1}(I + S)$$

Conditions for solvability and convergence

General Implication chain for PL functions

$$
\begin{array}{rccc}
 & \textbf{Bijectivity} & \Leftrightarrow & \textbf{Injectivity} \\
\Rightarrow & \textbf{Openness} & \Leftrightarrow & \textbf{Coherent Orientation} \\
\Rightarrow & \textbf{Surjectivity} & &
\end{array}
$$

Conditions for solvability and convergence

General Implication chain for PL functions

$$
\begin{array}{rccc}
 & \textbf{Bijectivity} & \Leftrightarrow & \textbf{Injectivity} \\
\Rightarrow & \textbf{Openness} & \Leftrightarrow & \textbf{Coherent Orientation} \\
\Rightarrow & \textbf{Surjectivity} & &
\end{array}
$$

Griewank et al 2013:

For equation $0 = F(x) = \min(x, Mx + q)$ and other simply switched system coherent orientation (i.e. M is P-Matrix) implies already bijectivity.

Conditions for solvability and convergence

General Implication chain for PL functions

$$\begin{array}{rccc} & \textbf{Bijectivity} & \Leftrightarrow & \textbf{Injectivity} \\ \Rightarrow & \textbf{Openness} & \Leftrightarrow & \textbf{Coherent Orientation} \\ \Rightarrow & \textbf{Surjectivity} & & \end{array}$$

Griewank et al 2013:

For equation $0 = F(x) = \min(x, Mx + q)$ and other simply switched system coherent orientation (i.e. M is P-Matrix) implies already bijectivity.

Sherman-Morrison-Woodbury yields

$$\det(J_\sigma) = \det(J)\det(I - S\,\Sigma)$$

Hence $det(I - S\Sigma) > 0$ for all Σ $\Rightarrow$ **Coherent Orientation**

Equivalent and related conditions

Rump showed

With *sign real spectral radius* $\rho_0^s(S) = \max(|\lambda|)$ over $\mathbb{R} \ni \lambda \in \mathbf{spect}(S)$

$$\det(I - S\,\Sigma) > 0 \;\Leftrightarrow\; \rho_0^s(S) < 1 \quad \Leftrightarrow \quad (S - I)^{-1}(S + I) \text{ is } P$$

also equivalent to non-expansiveness (Rohn).

$$x \geq 0 \implies |S\,x| \not\geq |x| \quad \text{componentwise}$$

Implication chain

Difficulty: Test for above property is NP-Hard.

$$\rho(|S|) \;<\; 1 \quad \Rightarrow \quad \|D^{-1}SD\|_p \;<\; 1 \quad \Rightarrow \quad \rho_0^s(S) < 1$$

(Absolute contractivity) (Smooth dominance) (Coherent orientation.)

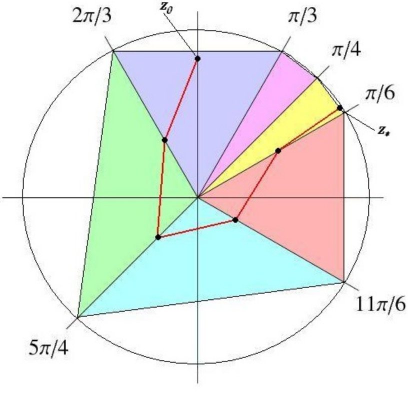
Solvers for PL systems of equations in original abs-normal or complementary form.

Method	Convergence condition	Rate	Effort
Generalized Newton on OPL	$2\,\hat{\rho} < (1 - \|L\|_p - \hat{\rho}/2)^2$	finite	$I - S\Sigma, J$
Generalized Newton on CPL	$\|S\|_p < 1/3$	finite	$I - S\Sigma$
Signed Gauss on CPL	$\rho(\|S\|) \leq 1/2$	finite	$I - S\Sigma$ once
Block Seidel on CPL	$\|S - L\|_p + \|L\|_p < 1$	linear	$I - L\Sigma, J$
Modulus Iteration on CPL	$\|S\|_p < 1$	linear	J
Piecewise Newton on OPL	coherent orient. of F	finite	$I - S\Sigma, J$
Piecewise Newton on CPL	$\rho_0^s(S) < 1$	finite	$I - S\Sigma$

Numerical results on Murty's LCP example

n	Anzahl der Iterationen		
	PLN (1. Startvektor)	PLN (2. Startvektor)	Lemke
2	2	2	$4 = 2^2$
4	4	6	$16 = 2^4$
6	10	14	$64 = 2^6$
8	24	32	$256 = 2^8$
10	54	66	$1024 = 2^{10}$
12	116	136	$4096 = 2^{12}$
14	246	276	$16384 = 2^{14}$
16	512	558	2^{16}
18	1048	1110	2^{18}
20	2126	2220	2^{20}

Piecewise linear Newton on Rosette Example



Optimization with quadratic overestimation

Under our assumptions on compact domains

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

Consequence: Bundle type iteration

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

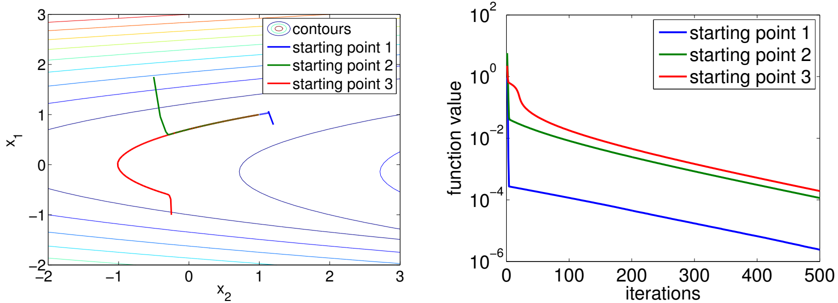
is guaranteed to converge from within bounded level set.

Application to Rosenbrock á la Nesterov

$$f(x_1, x_2) = \tfrac{1}{4}(x_1 - 1)^2 + \left| x_2 - 2\,x_1^2 + 1 \right| \ .$$

yields piecewise linearization

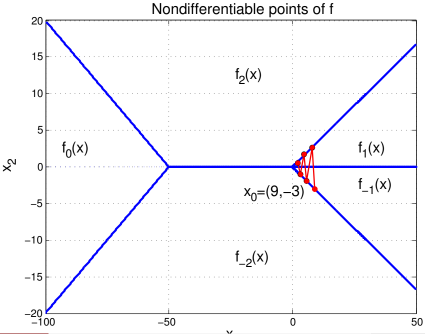
$$f(x_1, x_2) + \Delta f(x_1, x_2; \Delta x_1, \Delta x_2) \ =$$
$$\tfrac{1}{4}(x_1 - 1)^2 + \tfrac{1}{2}(x_1 - 1)\Delta x_1 + \left| x_2 + \Delta x_2 - 2\,x_1^2 - 4x_1 \Delta x_1 + 1 \right| \ .$$



Local = Inner Problem

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

- At least, global minimization is NP-hard ($\leftarrow$ SAT3)
- Bad News going back to Hirriart-Urruty & Lemarechal: Steepest descent with exact line search may fail on convex PL f.
- Challenge is to avoid Zenon effect = Zigzagging

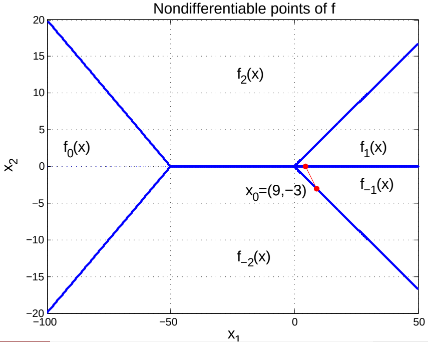


Good News by H. U. & L and Griewank et al:

True steepest descent trajectory $x(t)$ defined by:

$$-\frac{d\,x(t)}{d\,t_+} = -d(x) \equiv \textbf{short}(\partial f(x)) \equiv \textbf{argmin}\{\|g\| : g \in \partial f(x)\}$$

is in convex case unique solution of differential inclusion $\dot{x} \in -\partial f(x)$, which has stationary cluster points or limit x_* in initial level set. Can be realized using abs-normal form an and Zenon effect excluded.

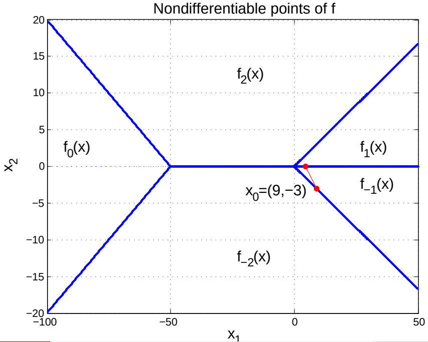


Good News by H. U. & L and Griewank et al:

True steepest descent trajectory $x(t)$ defined by:

$$-\frac{d\,x(t)}{d\,t_+} \;=\; -d(x) \equiv \textbf{short}(\partial f(x)) \equiv \textbf{argmin}\{\|g\| : g \in \partial f(x)\}$$

is in convex case unique solution of differential inclusion $\dot{x} \in -\partial f(x)$, which has stationary cluster points or limit x_* in initial level set. Can be realized using abs-normal form an and Zenon effect excluded.

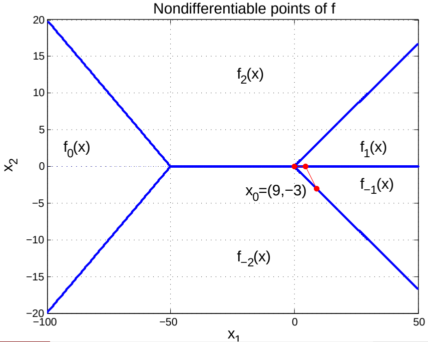


Good News by H. U. & L and Griewank et al:

True steepest descent trajectory $x(t)$ defined by:

$$-\frac{d\,x(t)}{d\,t_+} = -d(x) \equiv \mathbf{short}(\partial f(x)) \equiv \mathbf{argmin}\{\|g\| : g \in \partial f(x)\}$$

is in convex case unique solution of differential inclusion $\dot{x} \in -\partial f(x)$, which has stationary cluster points or limit x_* in initial level set.
Can be realized using abs-normal form an and Zenon effect excluded.

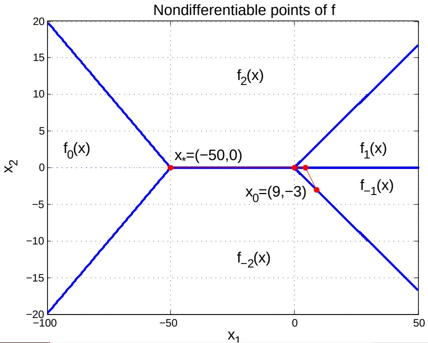


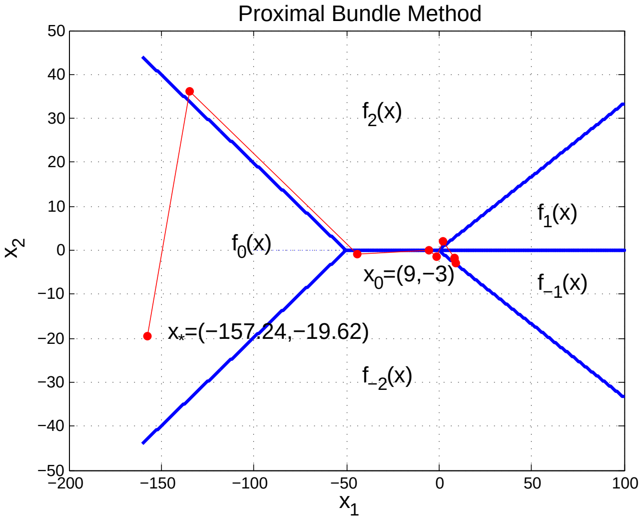
Good News by H. U. & L and Griewank et al:

True steepest descent trajectory $x(t)$ defined by:

$$-\frac{d\,x(t)}{d\,t_+} = -d(x) \equiv \textbf{short}(\partial f(x)) \equiv \textbf{argmin}\{\|g\| : g \in \partial f(x)\}$$

is in convex case unique solution of differential inclusion $\dot{x} \in -\partial f(x)$, which has stationary cluster points or limit x_* in initial level set. Can be realized using abs-normal form an and Zenon effect excluded.




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_1
x_2
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:

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

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 time step h

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

yields local third order truncation and globally second order convergence.

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 time step h

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

yields local third order truncation and globally second order convergence.

Properties in special cases

When F is PL GMP coincides with Average Vector Field Method (Quispel). Thus exact energy preservation if $F = J\nabla f$ Hamiltonian. Generally GMP is in contrast to IMP not (nonsmooth) symplectic.

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

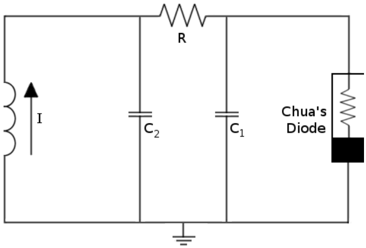


Figure: Chua circuit

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

- constants are $\alpha = 15.6, \beta = 28, m_0 = -1.143, m_1 = -0.714$

Chua circuit

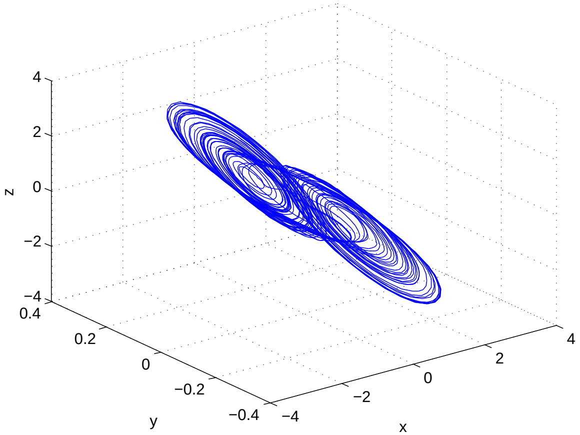


Figure: Chua Circuit - The Double Scroll

Convergence

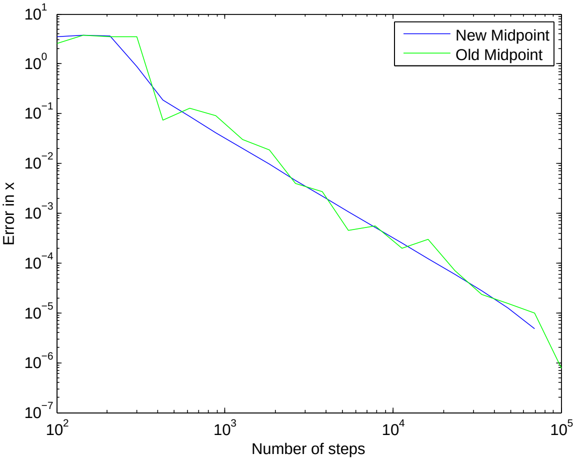
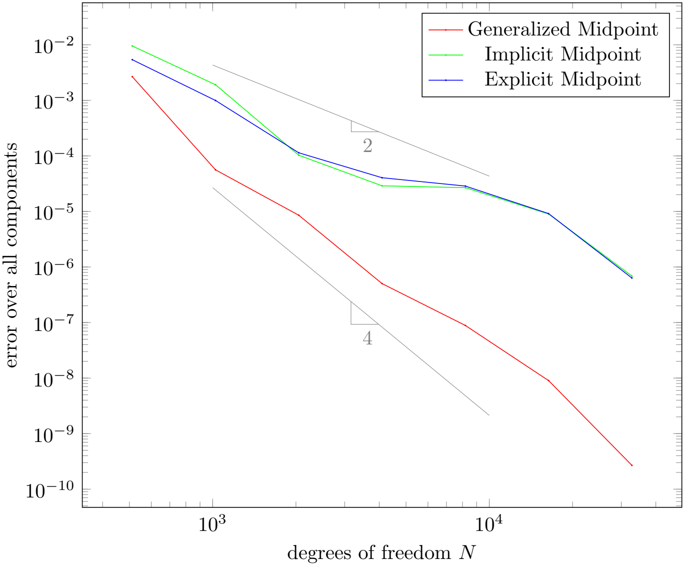
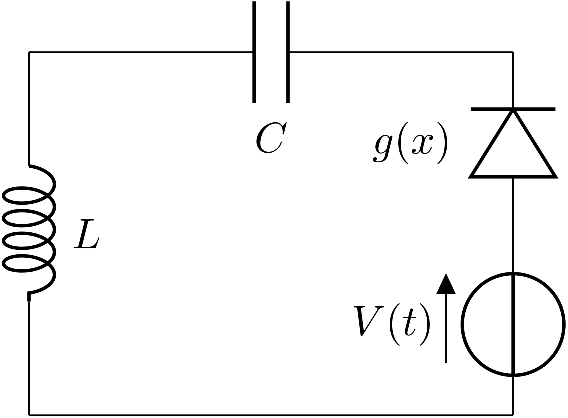


Figure: Error compared to fine grid solution

First level Richardson Extrapolation yields



Electrical circuit with diode



$$I(t) = z(t)$$

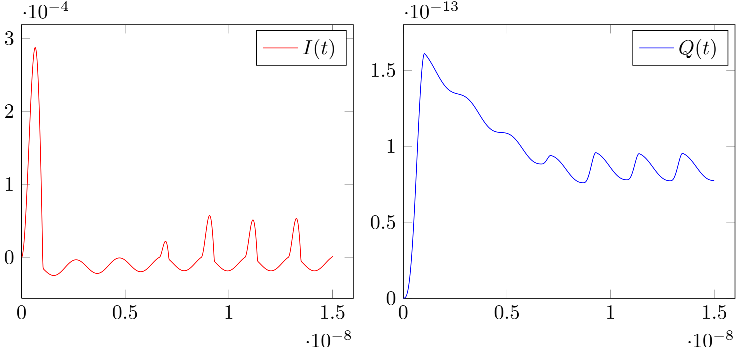
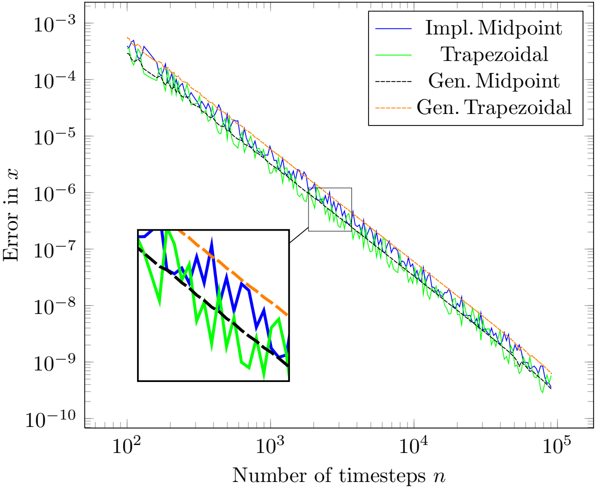
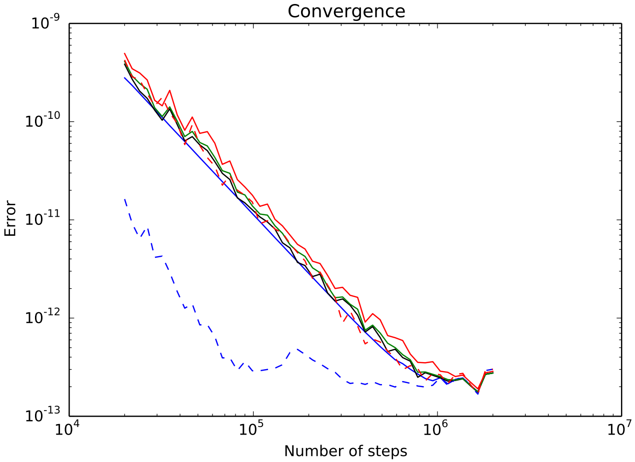


Figure: Solution of the ODE System

Log-log plot of convergence



Richardson/Romberg Extrapolation



Convergence Orders

Midpoint/Trapezoidal	Classical	Generalized	Extrapolated
General Position	$O(h)$	$O(h^2)$	$O(h^2)$
Transversal Position	$O(h^2)$	$ch^2 + O(h^3)$	$O(h^3)$
PL+smooth forcing	$O(h^2)$	$ch^2 + O(h^4)$	$O(h^4)$

Conjecture:

On discontinuous right hand sides all orders reduced by 1.

Challenge: Algebraic inclusion solving

In discontinuous case PL based discretization requires solution of $F(x) \ni 0$.
Special case: piecewise constant $F = \nabla f$ for continuous PL objective f.

Summary and Conclusion

- Differentiation of composite piecewise smooth functions yields PL function in abs-normal form.

Summary and Conclusion

- Differentiation of composite piecewise smooth functions yields PL function in abs-normal form.
- Local PL model has uniform second order fit and varies Lipschitz continuously with reference point.

Summary and Conclusion

- Differentiation of composite piecewise smooth functions yields PL function in abs-normal form.

- Local PL model has uniform second order fit and varies Lipschitz continuously with reference point.

- Limiting Jacobians/gradients of PL model can be computed and are conical derivatives of CPS function.

Summary and Conclusion

- Differentiation of composite piecewise smooth functions yields PL function in abs-normal form.
- Local PL model has uniform second order fit and varies Lipschitz continuously with reference point.
- Limiting Jacobians/gradients of PL model can be computed and are conical derivatives of CPS function.
- When $m = n$ the PL model system can be solved directly or after rewriting as complementary problem. In the transformation openness = coherent orientation = path continuation may be lost.

Summary and Conclusion

- Differentiation of composite piecewise smooth functions yields PL function in abs-normal form.
- Local PL model has uniform second order fit and varies Lipschitz continuously with reference point.
- Limiting Jacobians/gradients of PL model can be computed and are conical derivatives of CPS function.
- When $m = n$ the PL model system can be solved directly or after rewriting as complementary problem. In the transformation openness = coherent orientation = path continuation may be lost.
- Abs-normal form provides grey box information for systematic optimization of PL function by bundle type method.

Summary and Conclusion

- Differentiation of composite piecewise smooth functions yields PL function in abs-normal form.
- Local PL model has uniform second order fit and varies Lipschitz continuously with reference point.
- Limiting Jacobians/gradients of PL model can be computed and are conical derivatives of CPS function.
- When $m = n$ the PL model system can be solved directly or after rewriting as complementary problem. In the transformation openness = coherent orientation = path continuation may be lost.
- Abs-normal form provides grey box information for systematic optimization of PL function by bundle type method.
- Lipschitzian dynamical systems can be integrated with generalized and extrapolated midpoint/trapezoidal rule of global order 2, 3 or 4.

Summary and Conclusion

- Differentiation of composite piecewise smooth functions yields PL function in abs-normal form.
- Local PL model has uniform second order fit and varies Lipschitz continuously with reference point.
- Limiting Jacobians/gradients of PL model can be computed and are conical derivatives of CPS function.
- When $m = n$ the PL model system can be solved directly or after rewriting as complementary problem. In the transformation openness $=$ coherent orientation $=$ path continuation may be lost.
- Abs-normal form provides grey box information for systematic optimization of PL function by bundle type method.
- Lipschitzian dynamical systems can be integrated with generalized and extrapolated midpoint/trapezoidal rule of global order 2, 3 or 4.

References

- A.G.: On stable piecewise linearization and generalized algorithmic differentiation. *Optimization Methods and Software*, (2013).
- P. Boeck, B. Gompil, A. Griewank, R. Hasenfelder, N. Strogies: Experiments with Generalized Midpoint and Trapezoidal Rules on two Nonsmooth ODEs, *Mongolian Journal of Mathematics* (2014).
- A.G., M. Radons, T. Streubel, J.U. Bernd: On piecewise linear equations in abs-normal form. *Revised submission to Linear Algebra and Applications* (2014).
- A.G., A. Walther, S. Fiege, T. Bosse: On Lipschitz optimization based on gray-box piecewise linearization *submitted to Mathematical Programming, (2014)*
- A.G. et al: Plan-C, Piecewise Linear functions in Abs Normal form, software under development.