On Nonsmooth Optimization Based on Abs-Linearization

Andrea Walther
Institut für Mathematik
Humboldt-Universität zu Berlin

in memory of

Andreas Griewank, Humboldt-Universität zu Berlin

GdR MOA 2022

October 11, 2022

Andreas Griewank (1950 – 2021)

Outline

joint work with

Franz Bethke, Sabrina Fiege, Andreas Griewank, Timo Kreimeier, ...

50 Ways of Nonsmoothness ...

Smooth	$\Rightarrow$	Tight optimality conditions, (super-)linear convergence to roots of equation systems via linearization

Convex	$\Rightarrow$	Beautiful optimality and duality theory, but generally sublinear convergence to unconstrained minima

50 Ways of Nonsmoothness ...

Smooth	$\Rightarrow$	Tight optimality conditions, (super-)linear convergence to roots of equation systems via linearization
$\Downarrow$		
Abs-Smooth	$\Rightarrow$	Global piecewise linear approximation with uniform second order error

Convex	$\Rightarrow$	Beautiful optimality and duality theory, but generally sublinear convergence to unconstrained minima

50 Ways of Nonsmoothness ...

Smooth	$\Rightarrow$	Tight optimality conditions, (super-)linear convergence to roots of equation systems via linearization
$\Downarrow$		
Abs-Smooth	$\Rightarrow$	Global piecewise linear approximation with uniform second order error
$\Downarrow$		
Piecewise Smooth	$\Rightarrow$	Local piecewise linear approximation with very local second order error

Convex	$\Rightarrow$	Beautiful optimality and duality theory, but generally sublinear convergence to unconstrained minima

50 Ways of Nonsmoothness ...

Smooth	$\Rightarrow$	Tight optimality conditions, (super-)linear convergence to roots of equation systems via linearization
$\Downarrow$		
Abs-Smooth	$\Rightarrow$	Global piecewise linear approximation with uniform second order error
$\Downarrow$		
Piecewise Smooth	$\Rightarrow$	Local piecewise linear approximation with very local second order error
$\Downarrow$		
Sub-Analytic	$\Rightarrow$	Function finitely defined with Poussilieux expansion

Convex	$\Rightarrow$	Beautiful optimality and duality theory, but generally sublinear convergence to unconstrained minima

50 Ways of Nonsmoothness ...

Smooth	$\Rightarrow$	Tight optimality conditions, (super-)linear convergence to roots of equation systems via linearization
$\Downarrow$		
Abs-Smooth	$\Rightarrow$	Global piecewise linear approximation with uniform second order error
$\Downarrow$		
Piecewise Smooth	$\Rightarrow$	Local piecewise linear approximation with very local second order error
$\Downarrow$		
Sub-Analytic	$\Rightarrow$	Function finitely defined with Poussilieux expansion
$\Downarrow$		
Semi-Smooth	$\Rightarrow$	Very local superlinear convergence under rank conditions, combinatorics must be resolved for first iterate!
$\Uparrow$		
Convex	$\Rightarrow$	Beautiful optimality and duality theory, but generally sublinear convergence to unconstrained minima

... and Optimality Conditions

Given: $f : \mathbb{R}^n \mapsto \mathbb{R}$ not diff'able everywhere but with suitable properties

... and Optimality Conditions

Given: $f : \mathbb{R}^n \mapsto \mathbb{R}$ not diff'able everywhere but with suitable properties

Generalized derivative concepts required:

- directional derivatives

$$f'(x;d) := \lim_{t \to 0_+} \frac{f(x+td) - f(x)}{t} \in \mathbb{R} \qquad \forall d \in \mathbb{R}^n$$

... and Optimality Conditions

Given: $f : \mathbb{R}^n \mapsto \mathbb{R}$ not diff'able everywhere but with suitable properties

Generalized derivative concepts required:

- directional derivatives
  necessary optimality condition $f'(x; d) \geq 0$ for all $d \in \mathbb{R}^n$

... and Optimality Conditions

Given: $f : \mathbb{R}^n \mapsto \mathbb{R}$ not diff'able everywhere but with suitable properties

Generalized derivative concepts required:

- directional derivatives
  necessary optimality condition $f'(x; d) \geq 0$ for all $d \in \mathbb{R}^n$

- Clarke generalized gradient

  F. Clarke: Optimization and Nonsmooth Analysis, SIAM, 1990

$$\partial^C f(x) := \operatorname{conv}\Big\{ \lim_{i\to\infty} \nabla f(x_i) \,\Big|\, x_i \mapsto x, \nabla f(x_i) \text{ exists}\Big\} = \operatorname{conv}\{\partial^L f(x)\}$$

... and Optimality Conditions

Given: $f : \mathbb{R}^n \mapsto \mathbb{R}$ not diff'able everywhere but with suitable properties

Generalized derivative concepts required:

- directional derivatives
  necessary optimality condition $f'(x; d) \geq 0$ for all $d \in \mathbb{R}^n$

- Clarke generalized gradient
  F. Clarke: Optimization and Nonsmooth Analysis, SIAM, 1990
  necessary optimality condition $0 \in \partial^C f(x)$

... and Optimality Conditions

Given: $f : \mathbb{R}^n \mapsto \mathbb{R}$ not diff'able everywhere but with suitable properties

Generalized derivative concepts required:

- directional derivatives
  necessary optimality condition $f'(x; d) \geq 0$ for all $d \in \mathbb{R}^n$

- Clarke generalized gradient
  F. Clarke: Optimization and Nonsmooth Analysis, SIAM, 1990
  necessary optimality condition $0 \in \partial^C f(x)$

- Mordukhovich subgradient $\partial^M f(x)$
  T. Rockafellar, R. Wets: Variational Analysis, Springer, 1998

$$\partial^M f(x) = \left\{ g \in \mathbb{R}^n \,\middle|\, x_k \to x, f(x_k) \to f(x), g_k \in \widehat{\partial}^M f(x_k), g_k \to g \right\}$$

... and Optimality Conditions

Given: $f : \mathbb{R}^n \mapsto \mathbb{R}$ not diff'able everywhere but with suitable properties

Generalized derivative concepts required:

- directional derivatives
  necessary optimality condition $f'(x; d) \geq 0$ for all $d \in \mathbb{R}^n$

- Clarke generalized gradient
  F. Clarke: Optimization and Nonsmooth Analysis, SIAM, 1990
  necessary optimality condition $0 \in \partial^C f(x)$

- Mordukhovich subgradient $\partial^M f(x)$
  T. Rockafellar, R. Wets: Variational Analysis, Springer, 1998
  necessary optimality condition $0 \in \partial^M f(x)$

... and Optimality Conditions

Given: $f : \mathbb{R}^n \mapsto \mathbb{R}$ not diff'able everywhere but with suitable properties

Generalized derivative concepts required:

- directional derivatives
  necessary optimality condition $f'(x; d) \geq 0$ for all $d \in \mathbb{R}^n$

- Clarke generalized gradient
  F. Clarke: Optimization and Nonsmooth Analysis, SIAM, 1990
  necessary optimality condition $0 \in \partial^C f(x)$

- Mordukhovich subgradient $\partial^M f(x)$
  T. Rockafellar, R. Wets: Variational Analysis, Springer, 1998
  necessary optimality condition $0 \in \partial^M f(x)$

... and Optimality Conditions

Given: $f : \mathbb{R}^n \mapsto \mathbb{R}$ not diff'able everywhere but with suitable properties

Generalized derivative concepts required:

- directional derivatives
  necessary optimality condition $f'(x; d) \geq 0$ for all $d \in \mathbb{R}^n$

- Clarke generalized gradient
  F. Clarke: Optimization and Nonsmooth Analysis, SIAM, 1990
  necessary optimality condition $0 \in \partial^C f(x)$ $\quad \partial^C(|0|) = \partial^C(-|0|)$!

- Mordukhovich subgradient $\partial^M f(x)$
  T. Rockafellar, R. Wets: Variational Analysis, Springer, 1998
  necessary optimality condition $0 \in \partial^M f(x)$

Is This Really Necessary?

Theorem (Rademacher)

If the function $f : \mathbb{R}^n \mapsto \mathbb{R}$ is locally Lipschitz continuous then f is almost everywhere differentiable.

Is This Really Necessary?

Theorem (Rademacher)

If the function $f : \mathbb{R}^n \mapsto \mathbb{R}$ is locally Lipschitz continuous then f is almost everywhere differentiable.

Obvious idea: Use smooth optimization methods!

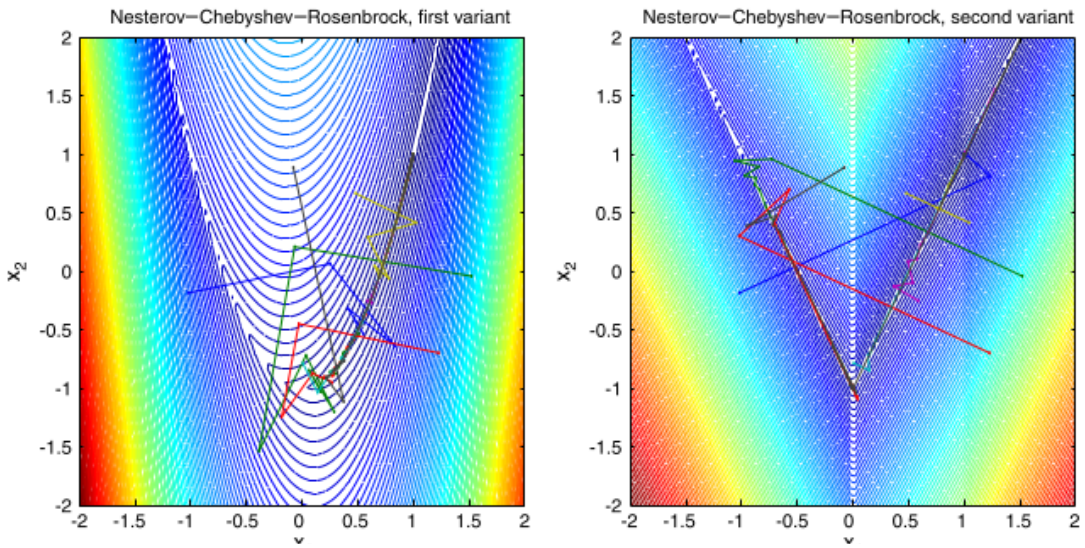
Is This Really Necessary?

Theorem (Rademacher)

If the function $f : \mathbb{R}^n \mapsto \mathbb{R}$ is locally Lipschitz continuous then f is almost everywhere differentiable.

Obvious idea: Use smooth optimization methods, e.g., quasi-Newton

M. Gürbüzbalaban, M.L. Overton / Nonlinear Analysis 75 (2012) 1282–1289

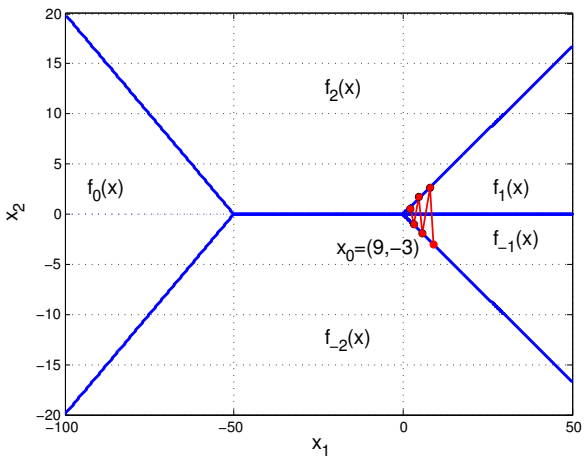


Is This Really Necessary?

Theorem (Rademacher)

If the function $f : \mathbb{R}^n \mapsto \mathbb{R}$ is locally Lipschitz continuous then f is almost everywhere differentiable.

Obvious idea: Use smooth optimization methods, e.g., steepest descent



Example by Hiriart-Urruty and Lemaréchal

Current (= Black Box) Approaches for Nonsmooth Optimization

- Use methods for smooth problems
  may fail, see slide before, no convergence theory
- Subgradient method
  very (!) slow convergence
- Bundle methods
  lots of parameters, erratic convergence behaviour
  involves oracle
- Semi-smooth Newton methods
  only local convergence
- Derivative-free methods
  no structure exploitation,
  difficult when number of optimization variables large

Lessons learned

- various concepts for nonsmoothness
- various concepts for generalized derivatives
- resulting in various optimality conditions,
  usually difficult to verify
- many important applications, e.g., machine learning!!
- no out-of-the-shelf solution algorithm
  structure exploitation indispensable

Where are we?

Smooth	$\Rightarrow$	Tight optimality conditions, (super-)linear convergence to roots of equation systems via linearization
$\Downarrow$		
Abs-Smooth	$\Rightarrow$	Global piecewise linear approximation with uniform second order error
$\Downarrow$		
Piecewise Smooth	$\Rightarrow$	Local piecewise linear approximation with very local second order error
$\Downarrow$		
Sub-Analytic	$\Rightarrow$	Function finitely defined with Poussilieux expansion
$\Downarrow$		
Semi-Smooth	$\Rightarrow$	Very local superlinear convergence under rank conditions, combinatorics must be resolved for first iterate!
$\Uparrow$		
Convex	$\Rightarrow$	Beautiful optimality and duality theory, but generally sublinear convergence to unconstrained minima

Our Class of Functions

Definition ($\mathcal{C}^d_{\text{abs}}(\mathbb{R}^n)$ Functions)

For any $d \in \mathbb{N}$, the set of functions $f : \mathbb{R}^n \mapsto \mathbb{R}, y = f(x)$, defined by an abs-smooth form

$$\begin{aligned} z &= F(x, z, |z|) \,, \\ y &= \varphi(x, z), \end{aligned}$$

with $F \in \mathcal{C}^d(\mathbb{R}^{n+s+s}, \mathbb{R}^s)$ and $\varphi \in \mathcal{C}^d(\mathbb{R}^{n+s}, \mathbb{R})$, such that z_i is determined only by the values of z_j, $1 \leq j < i$, is denoted by $\mathcal{C}^d_{\text{abs}}(\mathbb{R}^n)$.

Our Class of Functions

Definition ($\mathcal{C}^d_{\text{abs}}(\mathbb{R}^n)$ Functions)

For any $d \in \mathbb{N}$, the set of functions $f : \mathbb{R}^n \mapsto \mathbb{R}, y = f(x)$, defined by an abs-smooth form

$$\begin{aligned} z &= F(x, z, |z|) \,, \\ y &= \varphi(x, z), \end{aligned}$$

with $F \in \mathcal{C}^d(\mathbb{R}^{n+s+s}, \mathbb{R}^s)$ and $\varphi \in \mathcal{C}^d(\mathbb{R}^{n+s}, \mathbb{R})$, such that z_i is determined only by the values of z_j, $1 \leq j < i$, is denoted by $\mathcal{C}^d_{\text{abs}}(\mathbb{R}^n)$.

An element $f \in \mathcal{C}^d(\mathbb{R}^{n+s+s}, \mathbb{R}^s)$ for $d \geq 1$ is called abs-smooth.

Our Class of Functions

Definition ($\mathcal{C}^d_{\text{abs}}(\mathbb{R}^n)$ Functions)

For any $d \in \mathbb{N}$, the set of functions $f : \mathbb{R}^n \mapsto \mathbb{R}, y = f(x)$, defined by an abs-smooth form

$$\begin{aligned} z &= F(x, z, |z|) \,, \\ y &= \varphi(x, z), \end{aligned}$$

with $F \in \mathcal{C}^d(\mathbb{R}^{n+s+s}, \mathbb{R}^s)$ and $\varphi \in \mathcal{C}^d(\mathbb{R}^{n+s}, \mathbb{R})$, such that z_i is determined only by the values of z_j, $1 \leq j < i$, is denoted by $\mathcal{C}^d_{\text{abs}}(\mathbb{R}^n)$.

An element $f \in \mathcal{C}^d(\mathbb{R}^{n+s+s}, \mathbb{R}^s)$ for $d \geq 1$ is called abs-smooth.

The components z_i, $1 \leq i \leq s$, of z are called switching variables.

Abs-smooth Example Problems

Exact ℓ_1 penalty functions

Reformulation of constrained optimization problem

$$\min_x f(x) \quad \text{s.t.} \quad c_i(x) = 0,\ i \in \mathcal{E}, \quad c_i(x) \geq 0,\ i \in \mathcal{I}$$

as unconstrained optimization problem with ℓ_1-penalty

$$\phi(x;\mu) = f(x) + \mu \sum_{i \in \mathcal{E}} |c_i(x)| + \mu \sum_{i \in \mathcal{I}} \max\{0, -c_i(x)\}$$

Abs-smooth Example Problems

Exact ℓ_1 penalty functions

Reformulation of constrained optimization problem

$$\min_x f(x) \quad \text{s.t.} \quad c_i(x) = 0,\ i \in \mathcal{E}, \quad c_i(x) \geq 0,\ i \in \mathcal{I}$$

as unconstrained optimization problem with ℓ_1-penalty

$$\phi(x;\mu) = f(x) + \mu \sum_{i \in \mathcal{E}} |c_i(x)| + \mu \sum_{i \in \mathcal{I}} \max\{0, -c_i(x)\}$$

Robust optimization often formulated as min-max problems

Abs-smooth Example Problems

Exact ℓ_1 penalty functions

Reformulation of constrained optimization problem

$$\min_x f(x) \quad \text{s.t.} \quad c_i(x) = 0,\ i \in \mathcal{E}, \quad c_i(x) \geq 0,\ i \in \mathcal{I}$$

as unconstrained optimization problem with ℓ_1-penalty

$$\phi(x;\mu) = f(x) + \mu \sum_{i \in \mathcal{E}} |c_i(x)| + \mu \sum_{i \in \mathcal{I}} \max\{0, -c_i(x)\}$$

Robust optimization often formulated as min-max problems

Train timetabling yields piecewise linear optimization problem

F. Fischer, C. Helmberg: Dynamic Graph Generation and Dynamic Rolling Horizon Techniques in Large Scale Train Timetabling, 2010

Abs-smooth Example Problems

Exact ℓ_1 penalty functions

Reformulation of constrained optimization problem

$$\min_x f(x) \quad \text{s.t.} \quad c_i(x) = 0,\ i \in \mathcal{E}, \quad c_i(x) \geq 0,\ i \in \mathcal{I}$$

as unconstrained optimization problem with ℓ_1-penalty

$$\phi(x;\mu) = f(x) + \mu \sum_{i \in \mathcal{E}} |c_i(x)| + \mu \sum_{i \in \mathcal{I}} \max\{0, -c_i(x)\}$$

Robust optimization often formulated as min-max problems

Train timetabling yields piecewise linear optimization problem

F. Fischer, C. Helmberg: Dynamic Graph Generation and Dynamic Rolling Horizon Techniques in Large Scale Train Timetabling, 2010

Local models may yield piecewise linear optimization problem

F. Liers, M. Merkert: Structural Investigation of Piecewise Linearized Flow Problems. 2016

Abs-smooth Example Problems

Exact ℓ_1 penalty functions

Reformulation of constrained optimization problem

$$\min_x f(x) \quad \text{s.t.} \quad c_i(x) = 0,\ i \in \mathcal{E}, \quad c_i(x) \geq 0,\ i \in \mathcal{I}$$

as unconstrained optimization problem with ℓ_1-penalty

$$\phi(x;\mu) = f(x) + \mu \sum_{i \in \mathcal{E}} |c_i(x)| + \mu \sum_{i \in \mathcal{I}} \max\{0, -c_i(x)\}$$

Robust optimization often formulated as min-max problems

Train timetabling yields piecewise linear optimization problem

F. Fischer, C. Helmberg: Dynamic Graph Generation and Dynamic Rolling Horizon Techniques in Large Scale Train Timetabling, 2010

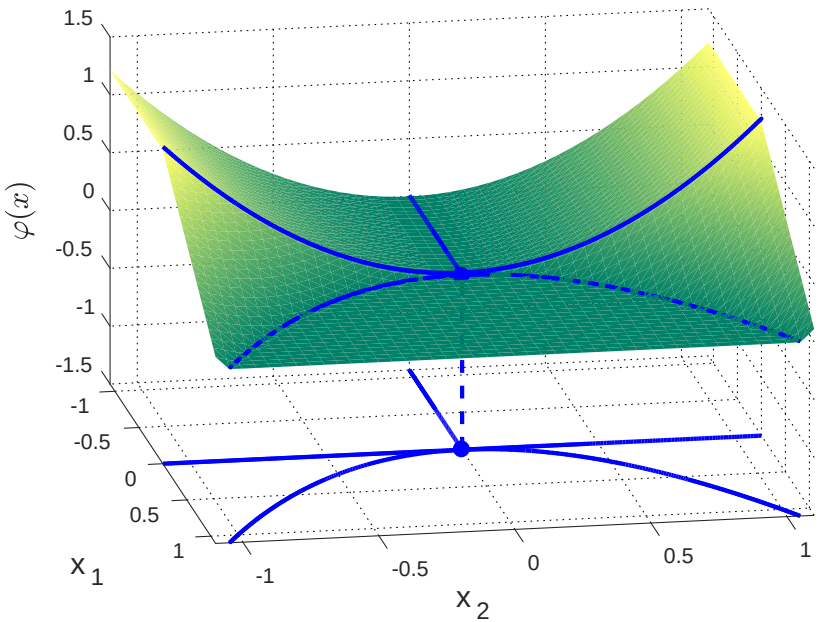
Local models may yield piecewise linear optimization problem

F. Liers, M. Merkert: Structural Investigation of Piecewise Linearized Flow Problems. 2016

Machine Learning nonsmooth activation functions like ReLu

The Half-pipe Function

$$f : \mathbb{R}^2 \mapsto \mathbb{R},\ f(x_1, x_2) = \max(x_2^2 - \max(x_1, 0), 0)$$
$$= \tfrac{1}{2}(x_2^2 - \tfrac{1}{2}\left(x_1 + |x_1|\right) + |x_2^2 - \tfrac{1}{2}\left(x_1 + |x_1|\right)|)$$



The Half-pipe Function

$$f : \mathbb{R}^2 \mapsto \mathbb{R},\ f(x_1, x_2) = \max(x_2^2 - \max(x_1, 0), 0)$$
$$= \tfrac{1}{2}(x_2^2 - \tfrac{1}{2}(x_1 + |x_1|) + |x_2^2 - \tfrac{1}{2}(x_1 + |x_1|)|)$$

has the abs-smooth form

$$\left.\begin{array}{lcl} z_1 & = & x_1 \\ z_2 & = & x_2^2 - \frac{1}{2}(x_1 + |z_1|) \\ z_3 & = & |z_2| \end{array}\right\} \qquad \text{i.e.,} \qquad z = F(x, z, |z|)$$
$$y = \varphi(x, z) = \tfrac{1}{2}(z_2 + z_3)$$

Definition (Piecewise Smooth (PS), Piecewise Linear (PL))

Let $\mathcal{D} \subseteq \mathbb{R}^n$ be open and $f_i : \mathcal{D} \to \mathbb{R}^m, i = 1, \ldots, k$ with $k \in \mathbb{N}$ be given.

- $f : \mathcal{D} \to \mathbb{R}^m$ is called *continuous selection* of the collection $f_1, \ldots, f_k$ on the set $U \subseteq \mathcal{D}$ if f is continuous on U and

$$f(x) \in \{f_1(x), \ldots, f_k(x)\} \qquad \forall x \in U.$$

S. Scholtes: Introduction to Piecewise Differentiable Equations, Springer, 2012

Definition (Piecewise Smooth (PS), Piecewise Linear (PL))

Let $\mathcal{D} \subseteq \mathbb{R}^n$ be open and $f_i : \mathcal{D} \to \mathbb{R}^m, i = 1, \ldots, k$ with $k \in \mathbb{N}$ be given.

- $f : \mathcal{D} \to \mathbb{R}^m$ is called *continuous selection* of the collection $f_1, \ldots, f_k$ on the set $U \subseteq \mathcal{D}$ if f is continuous on U and
$$f(x) \in \{f_1(x), \ldots, f_k(x)\} \qquad \forall x \in U.$$
- $f : \mathcal{D} \to \mathbb{R}^m$ is called *PC^r-function* with $r \in \mathbb{N} \cup \{\infty\}$ if for every $x \in \mathcal{D}$ there exists an open neighboorhood $U \subseteq \mathcal{D}$, such that f is a continuous selection of $f_1, \ldots, f_k$ on U and $f_i \in C^r(\mathcal{D}), 1 \leq i \leq k$.

S. Scholtes: Introduction to Piecewise Differentiable Equations, Springer, 2012

Definition (Piecewise Smooth (PS), Piecewise Linear (PL))

Let $\mathcal{D} \subseteq \mathbb{R}^n$ be open and $f_i : \mathcal{D} \to \mathbb{R}^m, i = 1, \ldots, k$ with $k \in \mathbb{N}$ be given.

- $f : \mathcal{D} \to \mathbb{R}^m$ is called *continuous selection* of the collection $f_1, \ldots, f_k$ on the set $U \subseteq \mathcal{D}$ if f is continuous on U and
$$f(x) \in \{f_1(x), \ldots, f_k(x)\} \qquad \forall x \in U.$$
- $f : \mathcal{D} \to \mathbb{R}^m$ is called *PC^r-function* with $r \in \mathbb{N} \cup \{\infty\}$ if for every $x \in \mathcal{D}$ there exists an open neighboorhood $U \subseteq \mathcal{D}$, such that f is a continuous selection of $f_1, \ldots, f_k$ on U and $f_i \in C^r(\mathcal{D}), 1 \leq i \leq k$.
- A PC^r-function is also called *piecewise smooth*.

S. Scholtes: Introduction to Piecewise Differentiable Equations, Springer, 2012

Definition (Piecewise Smooth (PS), Piecewise Linear (PL))

Let $\mathcal{D} \subseteq \mathbb{R}^n$ be open and $f_i : \mathcal{D} \to \mathbb{R}^m, i = 1, \ldots, k$ with $k \in \mathbb{N}$ be given.

- $f : \mathcal{D} \to \mathbb{R}^m$ is called *continuous selection* of the collection $f_1, \ldots, f_k$ on the set $U \subseteq \mathcal{D}$ if f is continuous on U and
$$f(x) \in \{f_1(x), \ldots, f_k(x)\} \qquad \forall x \in U.$$
- $f : \mathcal{D} \to \mathbb{R}^m$ is called *PC^r-function* with $r \in \mathbb{N} \cup \{\infty\}$ if for every $x \in \mathcal{D}$ there exists an open neighboorhood $U \subseteq \mathcal{D}$, such that f is a continuous selection of $f_1, \ldots, f_k$ on U and $f_i \in C^r(\mathcal{D}), 1 \leq i \leq k$.
- A PC^r-function is also called *piecewise smooth*.
- A continuous selection $f : U \to \mathbb{R}^m$ is called *piecewise linear* if all elements of the collection $f_1, \ldots, f_k$ are affine functions.

S. Scholtes: Introduction to Piecewise Differentiable Equations, Springer, 2012

Definition (Piecewise Smooth (PS), Piecewise Linear (PL))

Let $\mathcal{D} \subseteq \mathbb{R}^n$ be open and $f_i : \mathcal{D} \to \mathbb{R}^m, i = 1, \ldots, k$ with $k \in \mathbb{N}$ be given.

- $f : \mathcal{D} \to \mathbb{R}^m$ is called *continuous selection* of the collection $f_1, \ldots, f_k$ on the set $U \subseteq \mathcal{D}$ if f is continuous on U and
$$f(x) \in \{f_1(x), \ldots, f_k(x)\} \qquad \forall x \in U.$$
- $f : \mathcal{D} \to \mathbb{R}^m$ is called *PC^r-function* with $r \in \mathbb{N} \cup \{\infty\}$ if for every $x \in \mathcal{D}$ there exists an open neighboorhood $U \subseteq \mathcal{D}$, such that f is a continuous selection of $f_1, \ldots, f_k$ on U and $f_i \in C^r(\mathcal{D}), 1 \leq i \leq k$.
- A PC^r-function is also called *piecewise smooth*.
- A continuous selection $f : U \to \mathbb{R}^m$ is called *piecewise linear* if all elements of the collection $f_1, \ldots, f_k$ are affine functions.

S. Scholtes: Introduction to Piecewise Differentiable Equations, Springer, 2012

One can show: $\mathcal{C}^d_{\text{abs}}(\mathbb{R}^n)$ is a proper subset of the PS functions!

Information Gained from $\mathcal{C}^d_{\text{abs}}(\mathbb{R}^n)$ Functions

For $d \geq 1$ the following matrices and vectors are well defined:

$$
\begin{aligned}
Z &= \tfrac{\partial}{\partial x} F(x, z, |z|) \in \mathbb{R}^{s \times n} \\
M &= \tfrac{\partial}{\partial z} F(x, z, |z|) \in \mathbb{R}^{s \times s} && \text{strictly lower triangular} \\
L &= \tfrac{\partial}{\partial |z|} F(x, z, |z|) \in \mathbb{R}^{s \times s} && \text{strictly lower triangular} \\
a &= \tfrac{\partial}{\partial x} \varphi(x, z) \in \mathbb{R}^n, && b = \tfrac{\partial}{\partial z} \varphi(x, z) \in \mathbb{R}^s
\end{aligned}
$$

Information Gained from $\mathcal{C}^d_{\text{abs}}(\mathbb{R}^n)$ Functions

For $d \geq 1$ the following matrices and vectors are well defined:

$$\begin{aligned}
Z &= \tfrac{\partial}{\partial x} F(x, z, |z|) \in \mathbb{R}^{s \times n} \\
M &= \tfrac{\partial}{\partial z} F(x, z, |z|) \in \mathbb{R}^{s \times s} && \text{strictly lower triangular} \\
L &= \tfrac{\partial}{\partial |z|} F(x, z, |z|) \in \mathbb{R}^{s \times s} && \text{strictly lower triangular} \\
a &= \tfrac{\partial}{\partial x} \varphi(x, z) \in \mathbb{R}^n, && b = \tfrac{\partial}{\partial z} \varphi(x, z) \in \mathbb{R}^s
\end{aligned}$$

The signature vector and the corresponding diagonal matrix given by

$$\sigma(x) = \text{sign}(z(x)) \quad \text{and} \quad \Sigma = \text{diag}(\sigma(x))$$

define the active switch set

$$\alpha = \alpha(x) \equiv \{1 \leq i < s \,|\sigma_i(x) = 0\} \,.$$

Information Gained from $\mathcal{C}^d_{\text{abs}}(\mathbb{R}^n)$ Functions

For $d \geq 1$ the following matrices and vectors are well defined:

$$\begin{aligned}
Z &= \tfrac{\partial}{\partial x} F(x, z, |z|) \in \mathbb{R}^{s \times n} \\
M &= \tfrac{\partial}{\partial z} F(x, z, |z|) \in \mathbb{R}^{s \times s} && \text{strictly lower triangular} \\
L &= \tfrac{\partial}{\partial |z|} F(x, z, |z|) \in \mathbb{R}^{s \times s} && \text{strictly lower triangular} \\
a &= \tfrac{\partial}{\partial x} \varphi(x, z) \in \mathbb{R}^n, && b = \tfrac{\partial}{\partial z} \varphi(x, z) \in \mathbb{R}^s
\end{aligned}$$

The signature vector and the corresponding diagonal matrix given by

$$\sigma(x) = \text{sign}(z(x)) \quad \text{and} \quad \Sigma = \text{diag}(\sigma(x))$$

define the active switch set

$$\alpha = \alpha(x) \equiv \{1 \leq i < s \,|\sigma_i(x) = 0\} \,.$$

Required derivatives by extended algorithmic differentiation (AD)!

Once More the Half-pipe Function

$$f : \mathbb{R}^2 \mapsto \mathbb{R},\ f(x_1, x_2) = \max(x_2^2 - \max(x_1, 0), 0)$$
$$= \tfrac{1}{2}\left(x_2^2 - \tfrac{1}{2}\left(x_1 + |x_1|\right) + \left|x_2^2 - \tfrac{1}{2}\left(x_1 + |x_1|\right)\right|\right)$$

Once More the Half-pipe Function

$$f : \mathbb{R}^2 \mapsto \mathbb{R},\ f(x_1, x_2) = \max(x_2^2 - \max(x_1, 0), 0)$$
$$= \tfrac{1}{2}\left(x_2^2 - \tfrac{1}{2}\left(x_1 + |x_1|\right) + \left|x_2^2 - \tfrac{1}{2}\left(x_1 + |x_1|\right)\right|\right)$$

with

$$\left.\begin{array}{rcl} z_1 & = & x_1 \\ z_2 & = & x_2^2 - \frac{1}{2}\left(x_1 + |z_1|\right) \\ z_3 & = & |z_2| \end{array}\right\} \Rightarrow Z = \begin{bmatrix} 1 & 0 \\ -\frac{1}{2} & 2x_2 \\ 0 & 0 \end{bmatrix},\ M = 0,\ L = \begin{bmatrix} 0 & 0 & 0 \\ -\frac{1}{2} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

$$y = \varphi(x, z) = \tfrac{1}{2}\left(z_2 + z_3\right) \Rightarrow a = 0 \text{ and } b = (0,\ 0.5,\ 0.5)^\top .$$

A Local PL Model for Abs-smooth Functions

Definition (Abs-linear form of abs-smooth $f : \mathbb{R}^n \to \mathbb{R}$ at $\mathring{x}$)

The abs-linear form of f at $\mathring{x}$ is defined by $\Delta f(\mathring{x}; .) : \mathbb{R}^n \mapsto \mathbb{R}$,

$$\begin{bmatrix} z \\ y \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} Z & M & L \\ a & b & 0 \end{bmatrix} \begin{bmatrix} x \\ z \\ |z| \end{bmatrix} .$$

A Local PL Model for Abs-smooth Functions

Definition (Abs-linear form of abs-smooth $f : \mathbb{R}^n \to \mathbb{R}$ at $\mathring{x}$)

The abs-linear form of f at $\mathring{x}$ is defined by $\Delta f(\mathring{x}; .) : \mathbb{R}^n \mapsto \mathbb{R}$,

$$\begin{bmatrix} z \\ y \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} Z & M & L \\ a & b & 0 \end{bmatrix} \begin{bmatrix} x \\ z \\ |z| \end{bmatrix} .$$

Theorem

Suppose f is abs-smooth on $D \subset K \subset \mathbb{R}^n$, D open, K closed and convex. Then there exists $\gamma > 0$ such that for all $x, \mathring{x} \in K$

$$\|f(x) - \Delta f(\mathring{x}; x - \mathring{x})\| = \gamma \|x - \mathring{x}\|^2$$

A Local PL Model for Abs-smooth Functions

Definition (Abs-linear form of abs-smooth $f : \mathbb{R}^n \to \mathbb{R}$ at $\mathring{x}$)

The abs-linear form of f at $\mathring{x}$ is defined by $\Delta f(\mathring{x}; .) : \mathbb{R}^n \mapsto \mathbb{R}$,

$$\begin{bmatrix} z \\ y \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} Z & M & L \\ a & b & 0 \end{bmatrix} \begin{bmatrix} x \\ z \\ |z| \end{bmatrix} .$$

Theorem

Suppose f is abs-smooth on $D \subset K \subset \mathbb{R}^n$, D open, K closed and convex. Then there exists $\gamma > 0$ such that for all $x, \mathring{x} \in K$

$$\|f(x) - \Delta f(\mathring{x}; x - \mathring{x})\| = \gamma \|x - \mathring{x}\|^2$$

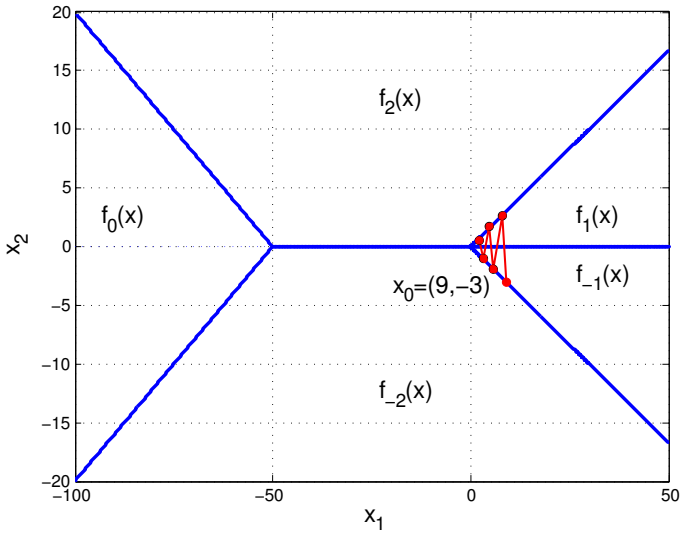
$\Rightarrow$ $\Delta f(\mathring{x}; .)$ is local piecewise linear model of second order!

A. Griewank. On stable piecewise linearization and generalized AD, OMS, 2013

Observations

Even min $\varphi(x)$ with PL convex φ not easy:

- Global minimization is NP-hard
- Steepest descent with exact line search may fail
- Zeno behaviour possible,
  i.e., solution trajectory with infinite number of direction changes in a finite amount
  of time



J.-B. Hiriart-Urruty, C. Lemaréchal: Convex Analysis and Minimization Algorithms I, Springer, 1993

Representations of PL Functions

There are many choices

Representations of PL Functions

There are many choices, e.g.,

Theorem (Max-Min representation of PL functions)

For each PL $f : \mathbb{R}^n \mapsto \mathbb{R}$ with selection functions $f_j(x) = a_j^\top x + b_j$, $1 \le j \le k$, there exist index sets $M_i \subset \{1, \ldots, k\}$, $1 \le i \le l$, such that

$$f(x) = \max_{1 \le i \le l} \min_{j \in M_i} a_j^\top x + b_j \ .$$

e.g., S. Scholtes: Introduction to Piecewise Differentiable Equations, Springer, 2012

Representations of PL Functions

There are many choices, e.g.,

Theorem (Max-Min representation of PL functions)

For each PL $f : \mathbb{R}^n \mapsto \mathbb{R}$ with selection functions $f_j(x) = a_j^\top x + b_j$, $1 \le j \le k$, there exist index sets $M_i \subset \{1, \ldots, k\}$, $1 \le i \le l$, such that

$$f(x) = \max_{1 \le i \le l} \min_{j \in M_i} a_j^\top x + b_j \ .$$

e.g., S. Scholtes: Introduction to Piecewise Differentiable Equations, Springer, 2012

Then, it follows from

$$\min(v, u) = (v + u - |v - u|)/2 \quad \text{and}$$
$$\max(v, u) = (v + u + |v - u|)/2,$$

that each PL function is in $\mathcal{C}^d_{\text{abs}}(\mathbb{R}^n)$ and has an abs-linear form!

Representations of PL Functions

There are many choices, e.g.,

Theorem (Max-Min representation of PL functions)

For each PL $f : \mathbb{R}^n \mapsto \mathbb{R}$ with selection functions $f_j(x) = a_j^\top x + b_j$, $1 \le j \le k$, there exist index sets $M_i \subset \{1, \ldots, k\}$, $1 \le i \le l$, such that

$$f(x) = \max_{1 \le i \le l} \min_{j \in M_i} a_j^\top x + b_j \ .$$

e.g., S. Scholtes: Introduction to Piecewise Differentiable Equations, Springer, 2012

Then, it follows from

$$\min(v, u) = (v + u - |v - u|)/2 \quad \text{and}$$
$$\max(v, u) = (v + u + |v - u|)/2,$$

that each PL function is in $\mathcal{C}^d_{\text{abs}}(\mathbb{R}^n)$ and has an abs-linear form!

$\Rightarrow$ Exploit abs-linear form for optimization!

Signature Domaines

Definition ((Extended) Signature domain)

For a fixed $\sigma \in \{-1, 0, 1\}^s$ and $f \in \mathcal{C}^d_{\text{abs}}(\mathbb{R}^n)$, we define

$$\mathcal{P}_\sigma \equiv \{x \in \mathbb{R}^n \mid \mathbf{sgn}(z(x)) = \sigma\} \subset \overline{\mathcal{P}}_\sigma \equiv \{x \in \mathbb{R}^n \mid \Sigma z(x) = |z(x)|\} \,.$$

$\mathcal{P}_\sigma$ is called *signature domain* and $\overline{\mathcal{P}}_\sigma$ *extended signature domain*.

Signature Domaines

Definition ((Extended) Signature domain)

For a fixed $\sigma \in \{-1, 0, 1\}^s$ and $f \in \mathcal{C}^d_{\text{abs}}(\mathbb{R}^n)$, we define

$$\mathcal{P}_\sigma \equiv \{x \in \mathbb{R}^n \mid \mathbf{sgn}(z(x)) = \sigma\} \subset \overline{\mathcal{P}}_\sigma \equiv \{x \in \mathbb{R}^n \mid \Sigma z(x) = |z(x)|\} \,.$$

$\mathcal{P}_\sigma$ is called *signature domain* and $\overline{\mathcal{P}}_\sigma$ *extended signature domain*.

- the signature domains form a disjoint decomposition of $\mathbb{R}^n$
- for a PL function f
  - each signature domain $\mathcal{P}_\sigma$ is a polyhedron and
  - f is linear on $\mathcal{P}_\sigma$

Signature Domaines

Definition ((Extended) Signature domain)

For a fixed $\sigma \in \{-1, 0, 1\}^s$ and $f \in \mathcal{C}^d_{\text{abs}}(\mathbb{R}^n)$, we define

$$\mathcal{P}_\sigma \equiv \{x \in \mathbb{R}^n \mid \mathbf{sgn}(z(x)) = \sigma\} \subset \overline{\mathcal{P}}_\sigma \equiv \{x \in \mathbb{R}^n \mid \Sigma z(x) = |z(x)|\} \,.$$

$\mathcal{P}_\sigma$ is called *signature domain* and $\overline{\mathcal{P}}_\sigma$ *extended signature domain*.

- the signature domains form a disjoint decomposition of $\mathbb{R}^n$
- for a PL function f
  - each signature domain $\mathcal{P}_\sigma$ is a polyhedron and
  - f is linear on $\mathcal{P}_\sigma$

Algorihmic idea: Minimize PL function on $\mathcal{P}_\sigma$, choose next $\mathcal{P}_{\tilde{\sigma}}$ carefully

Example: A Nesterov-Rosenbrock function

The Nesterov-Rosenbrock function

$$f : \mathbb{R}^n \mapsto \mathbb{R}, \quad f(x) = \tfrac{1}{4}\,|x_1 - 1| \; + \sum_{i=1}^{n-1} |x_{i+1} - 2|x_i| + 1|$$

has 2^{n-1} Clarke-stationary points!

M. Gürbüzbalaban, M. Overton, On Nesterov's nonsmooth Chebyshev–Rosenbrock functions, Nonlinear Anal: Theory, 2012

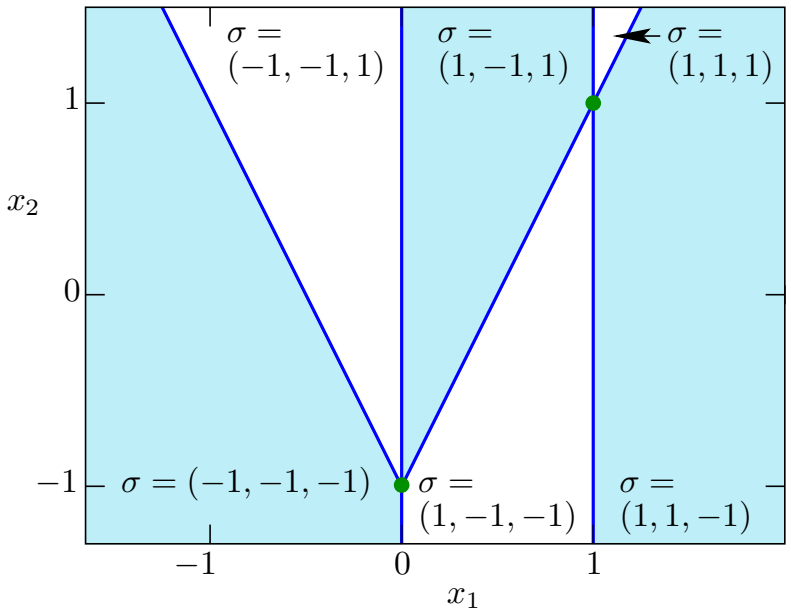
Example: A Nesterov-Rosenbrock function

The Nesterov-Rosenbrock function

$$f : \mathbb{R}^n \mapsto \mathbb{R}, \quad f(x) = \tfrac{1}{4}\,|x_1 - 1| \; + \sum_{i=1}^{n-1} |x_{i+1} - 2|x_i| + 1|$$

has 2^{n-1} Clarke-stationary points!

M. Gürbüzbalaban, M. Overton, On Nesterov's nonsmooth Chebyshev–Rosenbrock functions, Nonlinear Anal: Theory, 2012



Signature Optimal Point

Hence, consider for fixed $\sigma \in \{-1, 0, 1\}^s$

$$\begin{aligned} \min_{x \in \mathbb{R}^n, z \in \mathbb{R}^s} \quad & a^\top x + b^\top z \\ \text{s.t.} \quad & z = c + Zx + Mz + L\Sigma z \ , \\ & 0 = (I_s - |\Sigma|)z \ , \quad 0 \leq \Sigma z \ , \end{aligned}$$

for the signature matrix $\Sigma = \operatorname{diag}(\sigma)$.

Signature Optimal Point

Hence, consider for fixed $\sigma \in \{-1, 0, 1\}^s$

$$\begin{aligned} \min_{x \in \mathbb{R}^n, z \in \mathbb{R}^s} \quad & a^\top x + b^\top z + \frac{1}{2} x^\top Q x \\ \text{s.t.} \quad & z = c + Zx + Mz + L\Sigma z \ , \\ & 0 = (I_s - |\Sigma|) z \ , \quad 0 \leq \Sigma z \ , \end{aligned}$$

for the signature matrix $\Sigma = \operatorname{diag}(\sigma)$ and a positive definite matrix Q.

Signature Optimal Point

Hence, consider for fixed $\sigma \in \{-1, 0, 1\}^s$

$$\begin{aligned} \min_{x \in \mathbb{R}^n, z \in \mathbb{R}^s} \quad & a^\top x + b^\top z + \frac{1}{2} x^\top Q x \\ \text{s.t.} \quad & z = c + Zx + Mz + L\Sigma z \,, \\ & 0 = (I_s - |\Sigma|) z \,, \quad 0 \leq \Sigma z \,, \end{aligned}$$

for the signature matrix $\Sigma = \operatorname{diag}(\sigma)$ and a positive definite matrix Q.

Definition (Signature optimal point)

Consider a fixed signature vector $\sigma \in \{-1, 0, 1\}^s$. A minimizer $x_\sigma \in \mathcal{P}_\sigma$ of the original optimization problem restricted to $\mathcal{P}_\sigma$ is called *signature optimal point* of the original optimization problem.

Required Regularity Condition

Definition (LIKQ)

Let a PL function $f : \mathbb{R}^n \mapsto \mathbb{R}$ and a signature vector $\sigma \in \{-1, 0, 1\}^s$ be given. We say that the linear independence kink qualification (LIKQ) is satisfied at a point $x_\sigma \in \mathbb{R}^n$ if the active Jacobian

$$J(x) \equiv \nabla z_\alpha^\sigma(x) \equiv \left(e_i^\top \nabla z^\sigma(x)\right)_{i \in \alpha} \in \mathbb{R}^{|\alpha| \times n}$$

has full row rank $|\alpha|$, which requires in particular that $|\sigma| \geq s - n$.

Necessary and sufficient optimality condition

Theorem

Let a PL function $f : \mathbb{R}^n \mapsto \mathbb{R}$ and a signature vector $\sigma \in \{-1, 0, 1\}^s$ be given. Assume that x_σ is signature optimal and that LIKQ holds at x_σ. Then x_σ is a local minimizer of f if and only if there exist Lagrange multipliers $\lambda \in \mathbb{R}^s$, such that

$$0 = a^\top + b^\top |\Sigma| \mathring{Z} - \lambda^\top P_\alpha^\top P_\alpha \mathring{Z} \qquad \text{and} \qquad |P_\alpha(b + \lambda)| \leq -P_\alpha \mathring{L}^\top \lambda$$

with $\mathring{Z} = (I_s - M - L\Sigma)^{-1} Z, \mathring{L} = (I_s - M - L\Sigma)^{-1} L$ and $P_\alpha \equiv (e_i^\top)_{i \in \alpha}$.

Necessary and sufficient optimality condition

Theorem

Let a PL function $f : \mathbb{R}^n \mapsto \mathbb{R}$ and a signature vector $\sigma \in \{-1, 0, 1\}^s$ be given. Assume that x_σ is signature optimal and that LIKQ holds at x_σ. Then x_σ is a local minimizer of f if and only if there exist Lagrange multipliers $\lambda \in \mathbb{R}^s$, such that

$$0 = a^\top + b^\top |\Sigma| \mathring{Z} - \lambda^\top P_\alpha^\top P_\alpha \mathring{Z} \qquad \text{and} \qquad |P_\alpha(b + \lambda)| \leq -P_\alpha \mathring{L}^\top \lambda$$

with $\mathring{Z} = (I_s - M - L\Sigma)^{-1} Z, \mathring{L} = (I_s - M - L\Sigma)^{-1} L$ and $P_\alpha \equiv (e_i^\top)_{i \in \alpha}$.

No combinatorics involved, can be verified in polynomial time!

A. Griewank, A. Walther: Finite convergence of an active signature method to local minima of piecewise linear functions. OMS, 2019

Active Signature Method (ASM)

= Optimization of unconstrained, piecewise linear functions

- minimization over a sequence of polyhedra
- new optimality conditions that can be verified in polynomial time
- corresponding adapted QP solver on each polyhedron
- convergence in finitely many steps

For the first time convergence to local minimizers!!

Active Signature Method (ASM)

= Optimization of unconstrained, piecewise linear functions

- minimization over a sequence of polyhedra
- new optimality conditions that can be verified in polynomial time
- corresponding adapted QP solver on each polyhedron
- convergence in finitely many steps

For the first time convergence to local minimizers!!

Example: Nesterov-Rosenbrock function (2^{n-1} Clarke-stationary points!)

$$\varphi_2 : \mathbb{R}^n \mapsto \mathbb{R}, \quad \varphi(x) = \tfrac{1}{4}\,|x_1 - 1| \; + \sum\nolimits_{i=1,\ldots,n-1} |x_{i+1} - 2|x_i| + 1|$$

Active Signature Method (ASM)

= Optimization of unconstrained, piecewise linear functions

- minimization over a sequence of polyhedra
- new optimality conditions that can be verified in polynomial time
- corresponding adapted QP solver on each polyhedron
- convergence in finitely many steps

For the first time convergence to local minimizers!!

Example: Nesterov-Rosenbrock function (2^{n-1} Clarke-stationary points!)

$$\varphi_2 : \mathbb{R}^n \mapsto \mathbb{R}, \quad \varphi(x) = \tfrac{1}{4}\,|x_1 - 1| + \sum\nolimits_{i=1,\dots,n-1} |x_{i+1} - 2|x_i| + 1|$$

Iterations numbers:

n	1	2	3	4	5	6	7	8	9	10
ASM+QP	2	4	8	16	32	64	128	256	512	1024
HANSO	3	61	494*	1341*	2521*	329*	357*	326*	307*	515*
MPBNGC	3	52	9859	9978*	3561*	4166*	2547*	1959*	9420*	9807*

* = stop at non-optimal, stationary point

A. Griewank, A. Walther: Finite convergence of an active signature method to local minima of piecewise linear functions. OMS, 2019

The Constrained Case I

First, we consider PL constraints, i.e.,

$$\begin{aligned} \min_{x\in\mathbb{R}^n, z\in\mathbb{R}^s} \quad & a^\top x + b^\top z \\ \text{s.t.} \quad & 0 = g + Ax + Bz + C|z| \ , \\ & 0 \geq h + Dx + Ez + F|z| \ , \\ & z = c + Zx + Mz + L|z| \ , \end{aligned}$$

Hence, target function might still be unbounded.

The Constrained Case I

First, we consider PL constraints, i.e.,

$$\begin{aligned} \min_{x\in\mathbb{R}^n, z\in\mathbb{R}^s} \quad & a^\top x + b^\top z \\ \text{s.t.} \quad & 0 = g + Ax + Bz + C|z| \,, \\ & 0 \geq h + Dx + Ez + F|z| \,, \\ & z = c + Zx + Mz + L|z| \,, \end{aligned}$$

Hence, target function might still be unbounded.

- generalization of LIKQ and optimality conditions possible yields Constrained Active Signature Method (CASM)
- same convergence results

PhD thesis of T. Kreimeier
T. Kreimeier, A. Walther und A. Griewank: An active signature method for constrained abs-linear minimization. In revision.

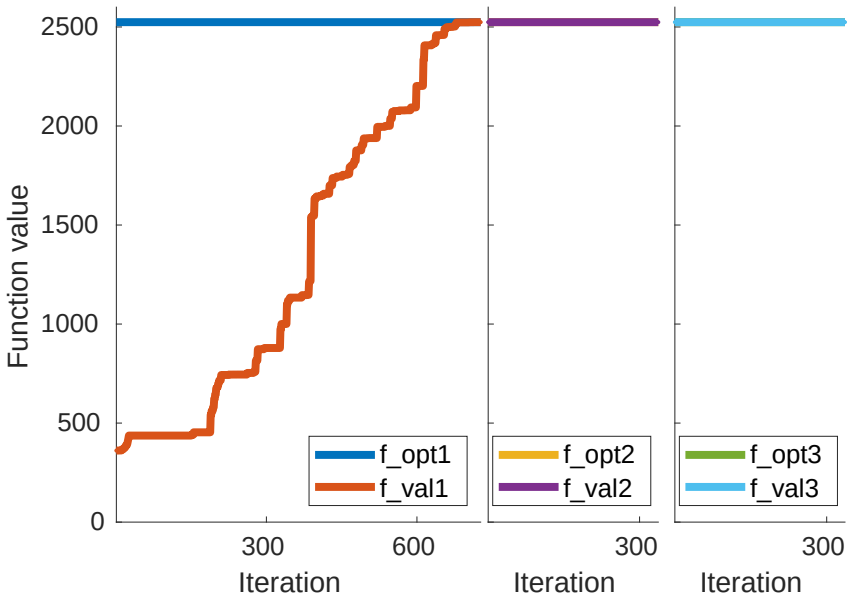
Robust gas network optimization

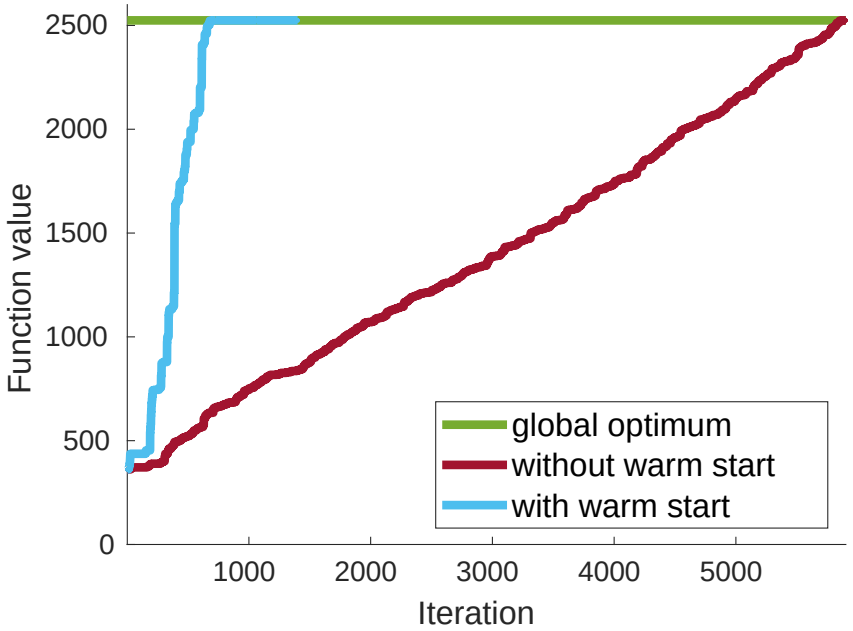
Here: Uncertainties in demand and in the physical parameters
Leads to PL constrained problem in inner loop of bundle method

Test instance GasLib134, i.e., the Greek gas network

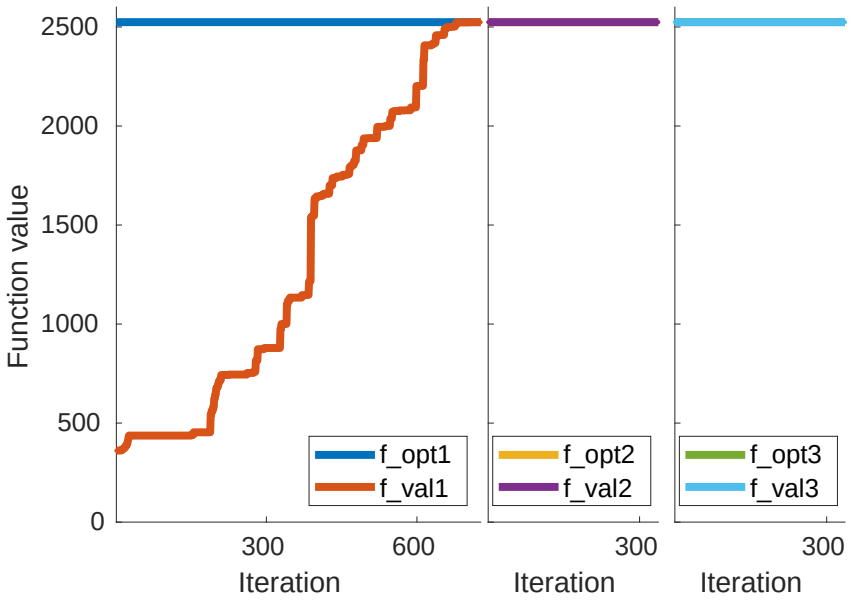
T. Kreimeier, M. Kuchlbauer, F. Liers, M. Stingl, A. Walther: Towards the Solution of Robust Gas Network Optimization Problems Using the Constrained Active Signature Method, INOC 2022

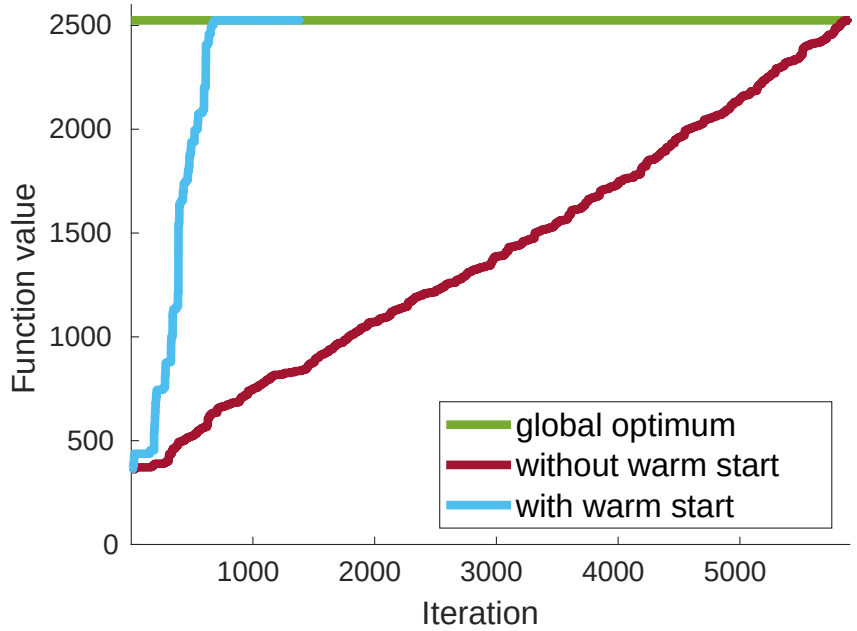
Results for Subproblem of GasLib-134





Results for Subproblem of GasLib-134





relaxation	1.	2.	3.
variables n		534	
equal. const. m		230	
inequal. const. p		784	
switching variables s	737	1107	1985
rows/columns of eq. system	3022	3762	5518

The Constrained Case II

Second, we consider

$$\min_{x \in C, z \in \mathbb{R}^s} \quad a^\top x + b^\top z$$
$$\text{s.t.} z = c + Zx + Mz + L|z| \ ,$$

for a convex, closed, and polyhedral feasible set C.

The Constrained Case II

Second, we consider

$$\min_{x \in C, z \in \mathbb{R}^s} \quad a^\top x + b^\top z$$
$$\text{s.t.} z = c + Zx + Mz + L|z| \ ,$$

for a convex, closed, and polyhedral feasible set C.

Existence on minimizers guaranteed!

The Constrained Case II

Second, we consider

$$\min_{x \in C, z \in \mathbb{R}^s} \quad a^\top x + b^\top z$$
$$\text{s.t.} z = c + Zx + Mz + L|z| \ ,$$

for a convex, closed, and polyhedral feasible set C.

Existence on minimizers guaranteed!

$\Rightarrow$ Use adapted version of ASM!

- again minimization over sequence of polyhedra
  now incorporating additional constraints and $Q = 0$
- optimality conditions like ASM,
  can be again verified in polynomial time
- LP solver on each polyhedron,
  here: HiGHS as solver

Again: The Nesterov-Rosenbrock function

We had for the Nesterov-Rosenbrock function

n	1	2	3	4	5	6	7	8	9	10
ASM+QP	2	4	8	16	32	64	128	256	512	1024
HANSO	3	61	494*	1341*	2521*	329*	357*	326*	307*	515*
MPBNGC	3	52	9859	9978*	3561*	4166*	2547*	1959*	9420*	9807*

Again: The Nesterov-Rosenbrock function

We had for the Nesterov-Rosenbrock function

n	1	2	3	4	5	6	7	8	9	10
ASM+QP	2	4	8	16	32	64	128	256	512	1024
HANSO	3	61	494*	1341*	2521*	329*	357*	326*	307*	515*
MPBNGC	3	52	9859	9978*	3561*	4166*	2547*	1959*	9420*	9807*

Introducing additional bounds not interfering with the minimizer, we get

n	1	2	3	4	5	6	7	8	9	10
polyh.	1	8	32	128	512	2048	8192	32768	131072	524288
aASM	1	2	4	8	16	32	64	128	256	512
splx	0	0	0	0	0	0	0	0	0	0

n	11	12	13	14	15	16	17	18	19	20
aASM	1024	2048	4096	8192	16384	32768	65536	131072	262144	524288
splx	0	0	0	0	0	0	0	0	0	0

The Unconstrained Case

The local PL model allows the optimization approach

$$x_{k+1} = x_k + \arg\min_{\Delta x} \left\{ \Delta f(x_k; \Delta x) + \tfrac{q}{2} \|\Delta x\|^2 \right\}$$

= **S**uccessive **A**bs-**L**inear **MIN**imization with a proximal term = SALMIN

The Unconstrained Case

The local PL model allows the optimization approach

$$x_{k+1} = x_k + \arg\min_{\Delta x} \left\{ \Delta f(x_k; \Delta x) + \tfrac{q}{2} \|\Delta x\|^2 \right\}$$

= **S**uccessive **A**bs-**L**inear **MIN**imization with a proximal term = SALMIN

Pros:

- piecewise linear local model can be generated by AD
- convergence theory and convergence rates (Griewank, Walther 2019)
- optimality can be verified in polynomial time using optimality conditions for the abs-smooth case, see Griewank, Walther (2016)

The Unconstrained Case

The local PL model allows the optimization approach

$$x_{k+1} = x_k + \arg\min_{\Delta x} \left\{ \Delta f(x_k; \Delta x) + \tfrac{q}{2} \|\Delta x\|^2 \right\}$$

= **S**uccessive **A**bs-**L**inear **MIN**imization with a proximal term = SALMIN

Pros:

- piecewise linear local model can be generated by AD
- convergence theory and convergence rates (Griewank, Walther 2019)
- optimality can be verified in polynomial time using optimality conditions for the abs-smooth case, see Griewank, Walther (2016)

Cons: For large-scale problems (large s!)

- computing the abs-linear form is expensive
  matrices are usually sparse, but sparsity ignored so far
- optimization process is slow since inner loop to compute $\arg\min_{\Delta x}(\ldots)$ stops at every kink

The Unconstrained Case

The local PL model allows the optimization approach

$$x_{k+1} = x_k + \arg\min_{\Delta x} \left\{ \Delta f(x_k; \Delta x) + \tfrac{q}{2} \|\Delta x\|^2 \right\}$$

= **S**uccessive **A**bs-**L**inear **MIN**imization with a proximal term = SALMIN

Pros:

- piecewise linear local model can be generated by AD
- convergence theory and convergence rates (Griewank, Walther 2019)
- optimality can be verified in polynomial time using optimality conditions for the abs-smooth case, see Griewank, Walther (2016)

Cons: For large-scale problems (large s!)

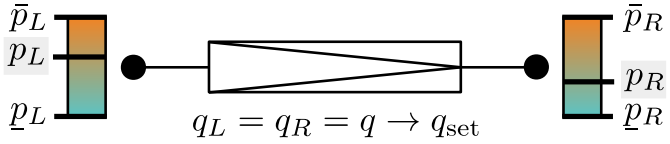
- computing the abs-linear form is expensive
  matrices are usually sparse, but sparsity ignored so far
- optimization process is slow since inner loop to compute $\arg\min_{\Delta x}(\ldots)$ stops at every kink

$\Rightarrow$ Developement of nonsmooth CG method (PhD topic of Franz Bethke)

Simulation of Gas Networks

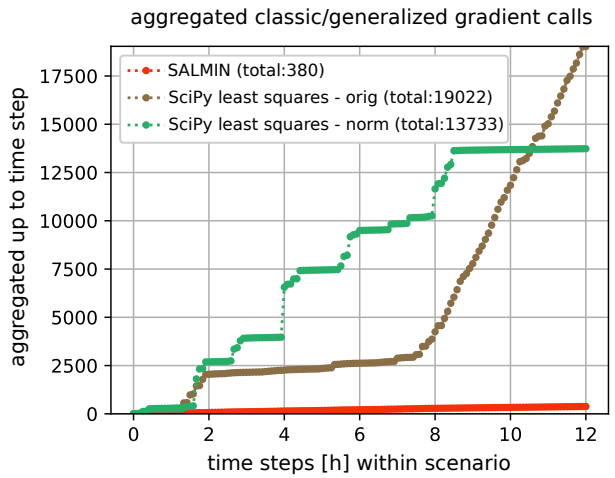
by combining Least-Squares Collocation and SALMIN

Control valve



$$\dot{q} = \frac{1}{\gamma}\max(-1, \min(1, \max(-\gamma \cdot q, \min(p_L - \underline{p}_L, \min(\bar{p}_R, p_L) - p_R, \max(\gamma \cdot (q_{\text{set}} - q), p_L - \bar{p}_L, \underline{p}_R - p_R)))))$$

Circuit symbol, set-point values and nonsmooth model

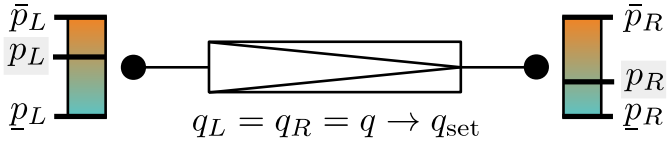


Comparison SALMIN vs. SciPy solver (transient control valve)

Simulation of Gas Networks

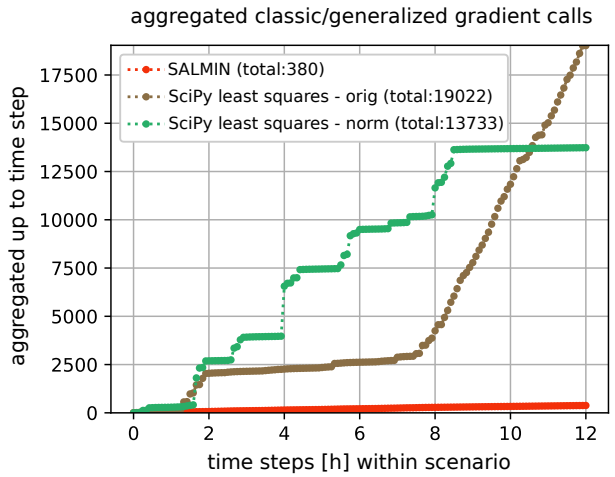
by combining Least-Squares Collocation and SALMIN

Control valve



$$\dot{q} = \frac{1}{\gamma}\max(-1, \min(1, \max(-\gamma \cdot q, \min(p_L - \underline{p}_L, \min(\bar{p}_R, p_L) - p_R, \max(\gamma \cdot (q_{\text{set}} - q), p_L - \bar{p}_L, \underline{p}_R - p_R)))))$$

Circuit symbol, set-point values and nonsmooth model



Comparison SALMIN vs. SciPy solver (transient control valve)

- faster convergence/better numerical stability than state-of-art solvers
- applied also to a 70 node extended network derived from GasLib40 (now including a close to real world compressor station)

T. Kreimeier, H. Sauter, T. Streubel, C. Tischendorf, A. Walther: Solving Least-Squares Collocated Differential Algebraic Equations by Successive Abs-Linear Minimization - A Case Study on Gas Network Simulation. TRR 154 preprint, in review

Conclusion and Outlook

- classes of nonsmooth optimization problems and their properties

Conclusion and Outlook

- classes of nonsmooth optimization problems and their properties
- abs-smooth functions and the local PL model via abs-linearization

Conclusion and Outlook

- classes of nonsmooth optimization problems and their properties
- abs-smooth functions and the local PL model via abs-linearization
- optimization of PL functions
  - algorithmic idea
  - convergence results
  - numerical results

  serves as work horse for nonsmooth optimization

Conclusion and Outlook

- classes of nonsmooth optimization problems and their properties
- abs-smooth functions and the local PL model via abs-linearization
- optimization of PL functions
  - algorithmic idea
  - convergence results
  - numerical results

  serves as work horse for nonsmooth optimization
- optimization of abs-smooth functions without constraints

Conclusion and Outlook

- classes of nonsmooth optimization problems and their properties
- abs-smooth functions and the local PL model via abs-linearization
- optimization of PL functions
  - algorithmic idea
  - convergence results
  - numerical results

  serves as work horse for nonsmooth optimization
- optimization of abs-smooth functions without constraints
- future work:
  optimization of abs-smooth functions with constraints

Thanks to TRR 154
Berlin Mathematics Research Center MATH+