Optimality and convexity conditions for piecewise smooth objective functions

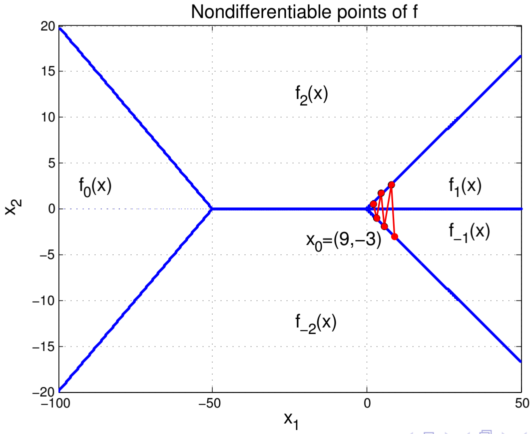
Andreas Griewank

EUROPT2016, Warsaw, 1. July 2016

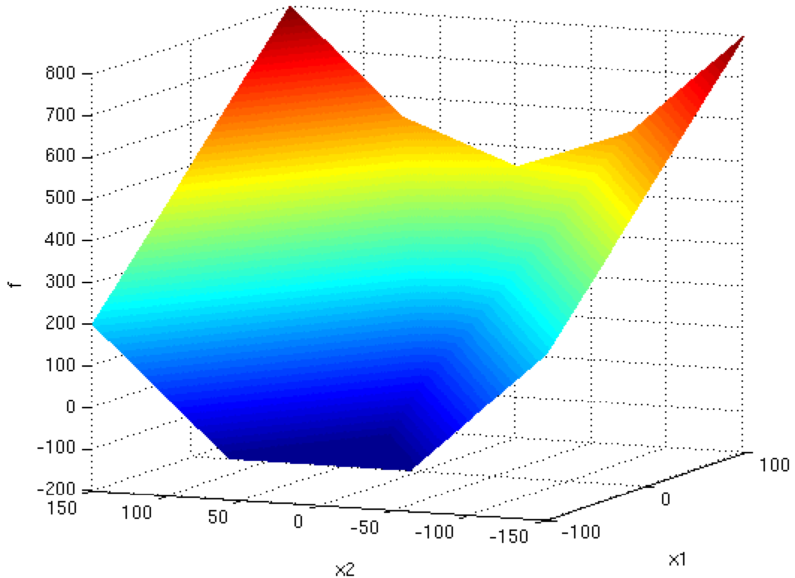
Yachay Tech University, Ecuador

Outline

- Steepest descent with exact line search may get stuck on convex piecewise linear (PL) f, due to Zenon effect = Zigzagging



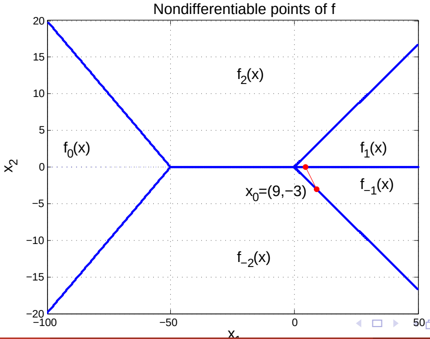
PL test function by H. U. & L.



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

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

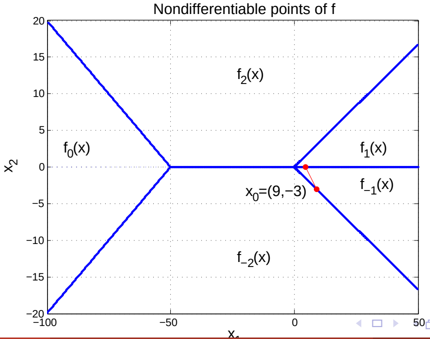
is in convex case unique solution of differential inclusion $\dot{x} \in -\partial f(x)$, which has stationary cluster points or limit x_* in minimal level set. Can be realized for PL f using abs-normal form and Zenon effect excluded.



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

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

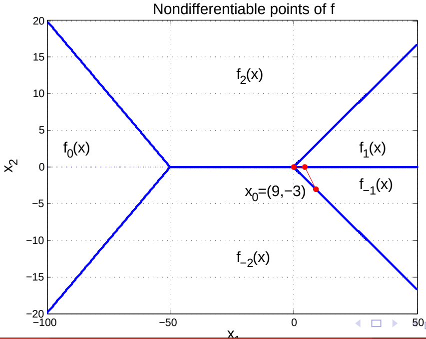
is in convex case unique solution of differential inclusion $\dot{x} \in -\partial f(x)$, which has stationary cluster points or limit x_* in minimal level set. Can be realized for PL f using abs-normal form and Zenon effect excluded.



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

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

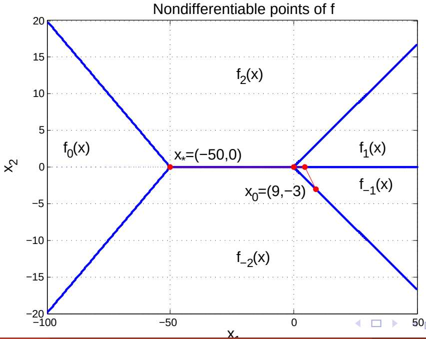
is in convex case unique solution of differential inclusion $\dot{x} \in -\partial f(x)$, which has stationary cluster points or limit x_* in minimal level set. Can be realized for PL f using abs-normal form and Zenon effect excluded.

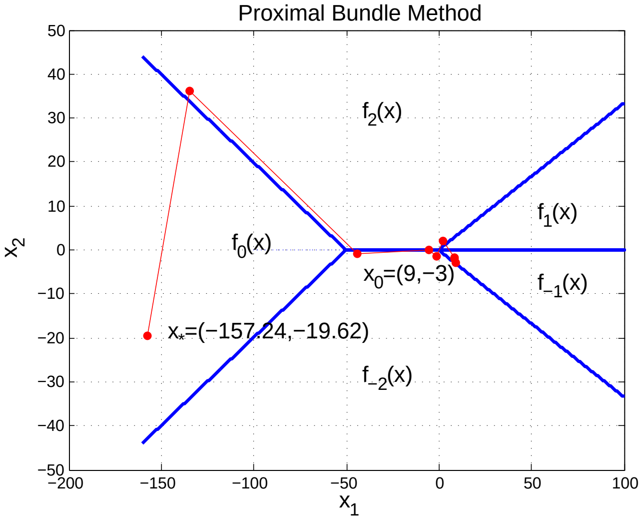


True steepest descent trajectory $x(t)$ defined by:

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

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





Proximal Bundle Method
50
40
30
20
10
0
−10
−20
−30
−40
−50
−200
−150
−100
−50
0
50
100
x2
x1
f2(x)
f1(x)
f0(x)
x0=(9,−3)
f−1(x)
x*=(−157.24,−19.62)
f−2(x)

Levels of Nonsmoothness

0. Straightline code with analytic elementals

Levels of Nonsmoothness

0. Straightline code with analytic elementals
1. Add abs, min and max to Level 0

Levels of Nonsmoothness

0. Straightline code with analytic elementals
1. Add abs, min and max to Level 0
2. Add Euclidean norm to Level 1

Levels of Nonsmoothness

0. Straightline code with analytic elementals
1. Add abs, min and max to Level 0
2. Add Euclidean norm to Level 1
3. Allow for sign and program branches

Levels of Nonsmoothness

0. Straightline code with analytic elementals
1. Add abs, min and max to Level 0
2. Add Euclidean norm to Level 1
3. Allow for sign and program branches

Assumed bounds on bounded domains

- Intermediate values and derivatives.
- Loop length and recursion depth.

Nesterov suggested $\varphi_\nu : \mathbb{R}^n \to \mathbb{R}$ given by

- $\varphi_0(x) = \frac{1}{4}(x_1 - 1)^2 + \sum_{i=1}^{n-1}(x_{i+1} - 2x_i^2 + 1)^2$ Level 0

  $\Rightarrow$ smooth and unimodal

- $\varphi_1(x) = \frac{1}{4}(x_1 - 1)^2 + \sum_{i=1}^{n-1}\left|x_{i+1} - 2x_i^2 + 1\right|$ Level 1

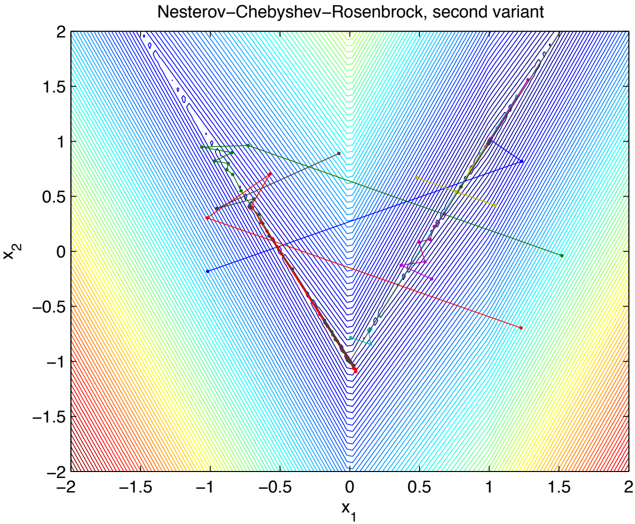
  $\Rightarrow$ nonsmooth, simply switched and unimodal

- $\varphi_2(x) = \frac{1}{4}|x_1 - 1| + \sum_{i=1}^{n-1}|x_{i+1} - 2|x_i| + 1|$ Level 1

  $\Rightarrow$ nonsmooth, multiply switched and multimodal

all have unique global minimizer $x_* = (1, 1, \cdots, 1)$
and bad start $(-1, 1, \ldots, (-1)^n)$.

BFGS on Second Nesterov Example



Popular standard: Clark stationarity

$$0 \in \partial^C \varphi(x) \;\equiv\; \textbf{conv}\{\lim g_k : g_k = \nabla\varphi(x_k) \text{ and } x_k \to x\}$$

A little stronger: Mordukhovich stationarity

$$0 \in \partial^M \varphi(x) \subset \partial^C \varphi(x)$$

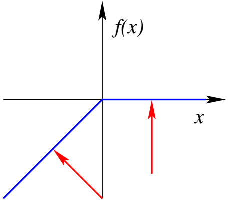
Glaring Example of Insufficiency

$$\varphi(x) = \min(x, 0) = \tfrac{1}{2}(x - |x|) \quad : \quad \mathbb{R} \to \mathbb{R}$$

$$\Rightarrow$$

$$0 \;\in\; \partial^M \varphi(0) = \{-1, 0\} \subset \partial^C \varphi(0) = [-1, 0]$$

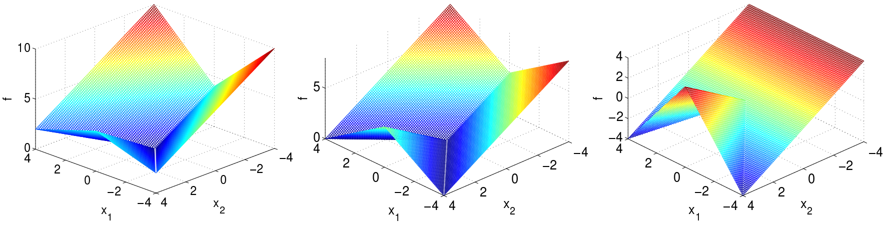
but φ is concave and unbounded below.



Has this function a minimum?

Gradient cube example with maximal switching depth

$$f(x) = |z_n| + \varepsilon \sum_{i=1}^{n-1} |z_i|, \quad z_1 = x_1, \quad \text{and} \quad z_i = x_i - |z_{i-1}|, \; i = 2, \ldots, n\,.$$



Note, no convexity whatsoever !

Level 1 Assumption:

Nonsmoothness cast in terms of abs() only!

Consequence

φ can be written in abs-normal form using switching variables

$$z_i, \quad i = 1, \ldots, s$$

as arguments of abs (.), i.e. $\varphi(x) = f(x, |z(x)|)$ where

$$\begin{aligned} z &= F(x, |z|) \quad \text{with} \quad F \in \mathcal{C}^2\left(\mathbb{R}^{n+s}, \mathbb{R}^s\right) \\ y &= f(x, |z|) \quad \text{with} \quad f \in \mathcal{C}^2\left(\mathbb{R}^{n+s}, \mathbb{R}\right) \end{aligned}$$

F and f or rather its relevant derivatives are obtainable by

Algorithmic Piecewise Differentiation APD

$$\begin{aligned}
\Delta y &\equiv \varphi(x+\Delta x) - \varphi(x) \\
&\approx a^\top \Delta x + b^\top(|z+\Delta z| - |z|) + \tfrac{1}{2}\Delta x^\top H \Delta x + O(\|\Delta x\|^2) \\
\Delta z &= Z\Delta x + L(|z+\Delta z| - |z|)
\end{aligned}$$

with

- $L = \dfrac{\partial F(x,|z|)}{\partial |z|} \in \mathbb{R}^{s\times s}$ strictly lower triangular of nilpotency $\nu \leq s$.
- $Z = \dfrac{\partial}{\partial x} F(x,|z|) \in \mathbb{R}^{s\times n}$
- $a = \dfrac{\partial}{\partial x} f(x,|z|) \in \mathbb{R}^{n}, \quad b = \dfrac{\partial}{\partial |z|} f(x,|z|) \in \mathbb{R}^{s}$
- $H = H(x,\lambda) \in \mathbb{R}^{n\times n} \;\equiv\;$ Hessian of suitable Lagrangian

$$\varphi_1(x) - \tfrac{1}{4}(x_1 - 1)^2 = \sum_{i=1}^{n-1} \left| x_{i+1} - 2x_i^2 + 1 \right| = \sum_{i=1}^{n-1} |z_i|$$

with

$$z_i = F_i(x, |z|) = x_{i+1} - 2x_i^2 + 1 \quad \text{for} \quad i = 1, \ldots, n-1$$

so that $L = 0 \in \mathbb{R}^{(n-1)\times(n-1)}$

$$Z(x) = \begin{bmatrix} -4\,x_1 & 1 & 0 & \cdots\cdots\cdots & 0 & 0 \\ 0 & -4\,x_2 & 1 & \cdots\cdots\cdots & 0 & 0 \\ 0 & 0 & -4\,x_3 & \cdots\cdots\cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots\cdots\cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots\cdots\cdots & 1 & 0 \\ 0 & 0 & 0 & \cdots\cdots\cdots & -4\,x_s & 1 \end{bmatrix} \in \mathbb{R}^{(n-1)\times n}$$

$a = \left(\tfrac{1}{2}(x_1 - 1), 0, \ldots, 0\right) \in \mathbb{R}^n$, and $b = (1, \ldots, 1) \in \mathbb{R}^{n-1}$

The signature vector

$$\sigma(x) = \mathrm{sgn}(z(x)) \;\in\; \{-1, 0, 1\}^s$$

and the corresponding diagonal matrix

$$\Sigma = \mathrm{diag}(\sigma) \;\in\; \{-1, 0, 1\}^{s \times s}$$

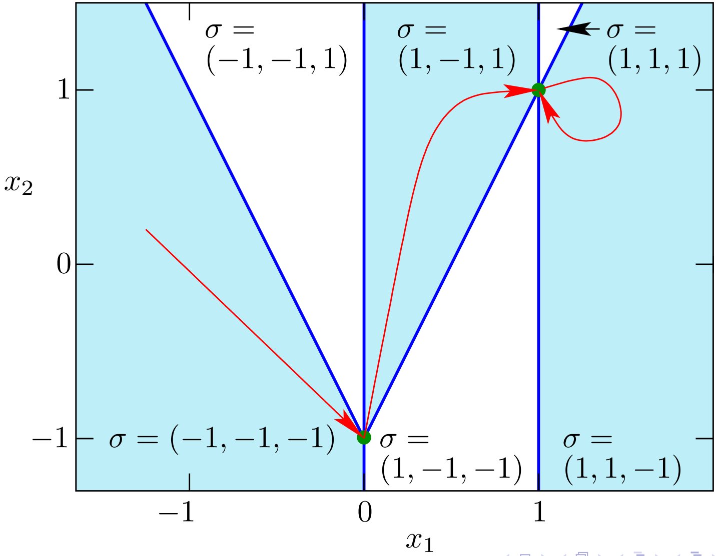
define active switch set

$$\alpha = \alpha(x) \equiv \{1 \le i \le s \;|\sigma_i(x) = 0\,\} \qquad |\alpha(x)| = s - |\sigma(x)|.$$

Furthermore, for fixed $\sigma(\Sigma)$

$$z = F(x, \Sigma z)$$

has unique solution z^σ with $\nabla z^\sigma = \frac{\partial}{\partial x} z^\sigma = (I - L\Sigma)^{-1} Z$.


σ = (−1, −1, 1)
σ = (1, −1, 1)
σ = (1, 1, 1)
σ = (−1, −1, −1)
σ = (1, −1, −1)
σ = (1, 1, −1)
1
0
−1
−1
0
1
x2
x1

Definition

We say that the linear independence kink qualification is satisfied at a point $x \in \mathbb{R}^n$ if for $\sigma = \sigma(x)$ 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 $|\sigma| \geq s - n$.

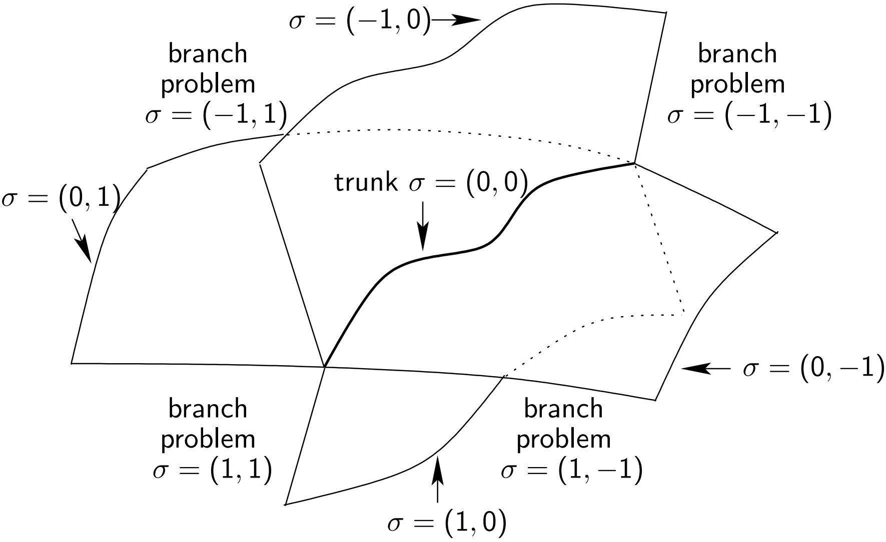
Lemma (Transversality of kink surfaces)

LIKQ implies that the sets $\{z_i(x) = 0\}$ form locally piecewise smooth hypersurfaces that are transversal whereever they intersect.

Lemma (LIKQ for Nesterov)

The functions φ_1 and φ_2 with their natural abs-normal forms satisfy LIKQ globally, i.e., throughout $\mathbb{R}^n$.

Trunk and branches for $n = 3, s = 2$



Assumptions

Full activity, i.e., $s = |\alpha| \leq n$ and
LIKQ, i.e. $J(x) = Z$ has full rank with $Z^\top V = 0$ for $V \in \mathbb{R}^{n\times(n-s)}$

Consequence 1

$$\min \varphi(x) \equiv f(x, 0) \quad s.t. \quad z = F(x, 0) = 0$$

satisfies LICQ and minimality requires

Tangential Stationarity

$$a^T + \lambda^T Z \; = \; 0 \; \in \mathbb{R}^n \quad \text{with} \quad \lambda \in \mathbb{R}^s$$

Positive Curvature

$$V^\top H V \succeq 0 \text{ with } H(x, \lambda) \equiv f(x, 0)_{xx} + \left(\lambda^\top F(x, 0)\right)_{xx} \in \mathbb{R}^{n\times n}$$

$$x \text{ local minimizer of } \varphi(x)$$

$$\Longleftrightarrow$$

$$x \text{ local minimizer of the smooth problems}$$

$$f(x, \Sigma z) \text{ s.t. } z = F(x, \Sigma z),\ \Sigma z \geq 0$$

$$\text{for any } \Sigma = \text{diag}(\sigma) \text{ with } \sigma \in \{-1, 1\}^s$$

Normal Growth

$$b^T + \lambda^T(L - \Sigma) \equiv \mu \geq 0 \quad \Longleftrightarrow \quad b^T \geq |\lambda^\top| - \lambda^T L \quad \Longleftrightarrow$$

$$b^T \geq \lambda^\top \Sigma - \lambda^T L \quad \text{and} \quad b^T \geq \lambda^\top(-\Sigma) - \lambda^T L \quad \text{for some } \Sigma$$

Lemma (Sufficient conditions in linear case)

If F and f are linear then a point x where LIKQ holds is a local minimizer $\Longleftrightarrow$

$$a^\top + \lambda^\top Z \;=\; 0 \quad \textit{and} \quad b^\top \;\geq\; |\lambda|^\top - \lambda^\top L \,.$$

i.e. the tangential stationarity and normal growth conditions are satisfied.

Corollary (Second order Sufficiency)

For general F, f the point x must be strict local minimizer if the normal growth condition holds strictly, i.e. $b > |\lambda| - L^\top \lambda$ and $V^T H V \succ 0$.

Lemma (Relation to stationarity concepts)

Tangential Stationarity and Normal Growth $\Longrightarrow$
Mordukhovich $\Longrightarrow$ *Clarke* $\Longrightarrow$ *Tangential Stationarity.*

Elimination of inactive kinks

Inactive switching variables $\hat{z} \equiv (\sigma_i z_i)_{i \notin \alpha}$ keep their sign in neighborhood and can be expressed as functions of x and critical switches $\check{z} = (z_i)_{i \in \alpha}$ i.e.

$$\hat{z} = \hat{z}(x, |\check{z}|) \in \mathbb{R}^{|\sigma|}$$

$$\Longleftrightarrow$$

$$\hat{z} = \hat{F}(x, |\check{z}|, \hat{z}) \equiv (\sigma_i F_i(x, |\check{z}|, \hat{z}))_{i \notin \alpha} \in \mathbb{R}^{|\sigma|}$$

Resulting reduced problem

$$\begin{aligned} \check{z} &= \check{F}(x, |\check{z}|, \hat{z}(x, |\check{z}|)) \in \mathbb{R}^{|\alpha|} \\ y &= f(x, |\check{z}|, \hat{z}(x, |\check{z}|)) \in \mathbb{R} i \end{aligned}$$

is fully active at reference point and LIKQ is maintained.

Tangential stationarity:

$$[f_x, f_{\hat{z}}] \;=\; -[\,\check{\lambda}^\top \;\hat{\lambda}^\top] \begin{bmatrix} \check{F}_x & \check{F}_{\hat{z}} \\ \hat{F}_x & \hat{F}_{\hat{z}} - I \end{bmatrix} \;\in\; \mathbb{R}^{n+|\sigma|}$$

Tangential stationarity:

$$[f_x, f_{\hat{z}}] \;=\; -[\,\check{\lambda}^\top\; \hat{\lambda}^\top] \begin{bmatrix} \check{F}_x & \check{F}_{\hat{z}} \\ \hat{F}_x & \hat{F}_{\hat{z}} - I \end{bmatrix} \;\in\; \mathbb{R}^{n+|\sigma|}$$

Normal growth:

$$f_{\bar{z}} \;\geq\; |\,\check{\lambda}^\top| \;-\; [\,\check{\lambda}^\top\; \hat{\lambda}^\top] \begin{bmatrix} \check{F}_{\bar{z}} \\ \hat{F}_{\bar{z}} \end{bmatrix} \;\in\; \mathbb{R}^{|\alpha|}$$

Optimality conditions in general case

Tangential stationarity:

$$[f_x, f_{\hat{z}}] \;=\; -[\check{\lambda}^\top \; \hat{\lambda}^\top] \begin{bmatrix} \check{F}_x & \check{F}_{\hat{z}} \\ \hat{F}_x & \hat{F}_{\hat{z}} - I \end{bmatrix} \;\in\; \mathbb{R}^{n+|\sigma|}$$

Normal growth:

$$f_{\bar{z}} \;\geq\; |\check{\lambda}^\top| \;-\; [\check{\lambda}^\top \; \hat{\lambda}^\top] \begin{bmatrix} \check{F}_{\bar{z}} \\ \hat{F}_{\bar{z}} \end{bmatrix} \;\in\; \mathbb{R}^{|\alpha|}$$

Positive Curvature:

$$0 \;\preceq\; \check{V}^\top \check{H} \check{V} \;\in \mathbb{R}^{(n-|\alpha|)\times(n-|\alpha|)}$$

Tangential stationarity:

$$[f_x, f_{\hat{z}}] \; = \; -[\check{\lambda}^\top \; \hat{\lambda}^\top] \begin{bmatrix} \check{F}_x & \check{F}_{\hat{z}} \\ \hat{F}_x & \hat{F}_{\hat{z}} - I \end{bmatrix} \; \in \; \mathbb{R}^{n+|\sigma|}$$

Normal growth:

$$f_{\bar{z}} \;\; \geq \;\; |\check{\lambda}^\top| \; - \; [\check{\lambda}^\top \; \hat{\lambda}^\top] \begin{bmatrix} \check{F}_{\bar{z}} \\ \hat{F}_{\bar{z}} \end{bmatrix} \; \in \; \mathbb{R}^{|\alpha|}$$

Positive Curvature:

$$0 \;\; \preceq \;\; \check{V}^\top \check{H} \check{V} \; \in \mathbb{R}^{(n-|\alpha|)\times(n-|\alpha|)}$$

- Becomes sufficient if normal growth strict and $\check{V}^\top \check{H} \check{V}$ is nonsingular.

Optimality conditions in general case

Tangential stationarity:

$$[f_x, f_{\hat{z}}] \;=\; -[\check{\lambda}^\top \; \hat{\lambda}^\top] \begin{bmatrix} \check{F}_x & \check{F}_{\hat{z}} \\ \hat{F}_x & \hat{F}_{\hat{z}} - I \end{bmatrix} \;\in\; \mathbb{R}^{n+|\sigma|}$$

Normal growth:

$$f_{\bar{z}} \;\geq\; |\check{\lambda}^\top| - [\check{\lambda}^\top \; \hat{\lambda}^\top] \begin{bmatrix} \check{F}_{\bar{z}} \\ \hat{F}_{\bar{z}} \end{bmatrix} \;\in\; \mathbb{R}^{|\alpha|}$$

Positive Curvature:

$$0 \;\preceq\; \check{V}^\top \check{H} \check{V} \;\in \mathbb{R}^{(n-|\alpha|)\times(n-|\alpha|)}$$

- Becomes sufficient if normal growth strict and $\check{V}^\top \check{H} \check{V}$ is nonsingular.
- Violation of any necessary condition yields parabolas of descent.

φ_0 : Conditions reduce to $\nabla\varphi_0 = 0$ and $H = \nabla^2\varphi_0 \succ 0$ which hold only at x_* with $\det(H) > 0$.

φ_1 : Tangential stationarity is only satisfied at x_*, which also exhibits strict normal growth and SSC.

φ_2 : All 2^{n-1} Clarke stationary points satisfy tangential stationarity but only x_* exhibits normal growth and that in strict form.

0 Initialize $\sigma \in \{-1,1\}^s$ and corresponding $\Sigma = \mathrm{diag}(\sigma)$

1 Compute local minimizer x^* of branch problem

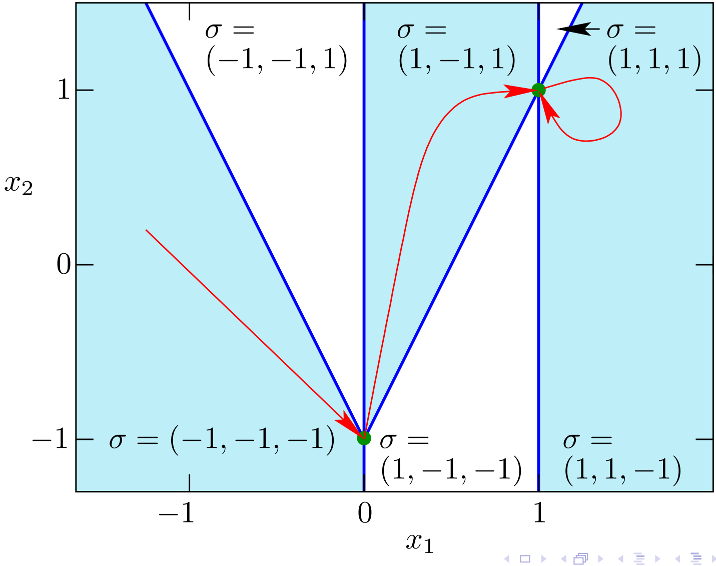
$$\min f(x, \Sigma z) \quad s.t. \quad z = F(x, \Sigma z) \quad \text{and} \quad \Sigma z \geq 0$$

2 Terminate if same x^* was obtained in previous iteration.

3 With $\sigma^* = \mathrm{sgn}(z(x^*))$ flip signs of σ_i for which $\sigma_i^* = 0$ and goto 1.

Lemma (Finite Convergnce)

Reflection algorithm reaches local minimizer if all NLOPs solvable, LIKQ holds everywhere and φ is bounded below.


$\sigma =$ $(-1,-1,1)$
$\sigma =$ $(1,-1,1)$
$\sigma =$ $(1,1,1)$
$\sigma = (-1,-1,-1)$
$\sigma =$ $(1,-1,-1)$
$\sigma =$ $(1,1,-1)$
1
0
−1
x_2
−1
0
1
x_1

0. Initialize starting point x and approximate Lagrangian Hessian H
1. Evaluate L, Z, a, b at x by AD update H by secant formula or exactly
2. Compute Δx by solving via bundle or reflection

$$\begin{aligned} \min \Delta y &\equiv a^\top \Delta x + b^\top(|z + \Delta z| - |z|) + \tfrac{1}{2}\Delta x^\top H \Delta x \\ s.t.\ \Delta z &= Z\Delta x + L(|z+\Delta z| - |z|) \end{aligned}$$

3. Set $x \mathrel{+}= \Delta x$ if $\varphi(x+\Delta x) < \varphi(x)$
4. Unless Δx or gain $\varphi(x) - \varphi(x+\Delta x)$ too small goto 1

Lemma (Conjecture:)

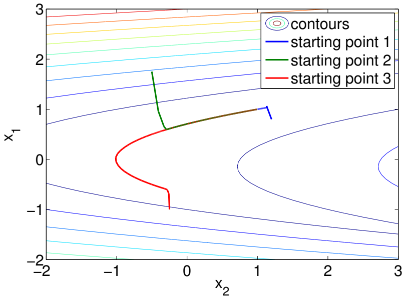
Local convergence with linear, superlinear, or quadratic rate depending on whether H is constant, secant updated, or evaluated, respectively.

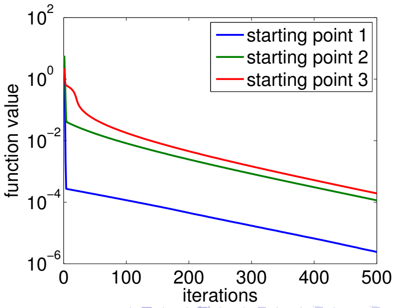
$$f(x_1, x_2) = \tfrac{1}{4}(x_1 - 1)^2 + \left| x_2 - 2\, x_1^2 + 1 \right| \; .$$

yields piecewise linearization

$$f(x_1, x_2) + \Delta f(x_1, x_2; \Delta x_1, \Delta x_2) \; = $$

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





One linear and one nonlinear test

We considered the scalable L1hilb function

$$f : \mathbb{R}^n \mapsto \mathbb{R}, \quad f(x) = \sum_{i=1}^{n} \left| \sum_{i=1}^{n} \frac{x_j}{i+j-1} \right| .$$

and the nonlinear MAXQ function

$$f(x) = \max_{1 \le i \le 5} \left(x^\top A^i x - x^\top b^i \right)$$

$$\begin{aligned} A^i_{kj} &= A^i_{jk} = e^{j/k} \cos(jk) \sin(i), \quad \text{for} \quad j < k, \quad j,k = 1, \dots, 10 \\ A^i_{jj} &= \tfrac{j}{10} |\sin(i)| + \textstyle\sum_{k \ne j} \left| A^i_{jk} \right| , \\ b^i_j &= e^{j/i} \sin(ij), \\ x^0_i &= 0, \quad \text{for all} \quad i = 1, \dots, 10. \end{aligned}$$

	n	q^0	f^*	$\#f$	$\#\nabla f$	Iter
LiPsMin	2	0	2.2e-16	3	3	1
	5	0	2.3e-16	3	3	1
	10	0	6.4e-16	3	3	1
	20	0	8.8e-11	3	6	1
	50	0	3.8e-10	3	6	1
	100	0	7.5e-15	3	3	1
HANSO	2	–	1.6e-2	10191	10191	5 + 3GS
	5	–	5.7e-3	11678	11678	4 + 3GS
	10	–	8.8e-3	14320	14320	2 + 3GS
	20	–	1.2e-1	17953	17953	3 + 3GS
	50	–	1.8e-1	26841	26841	3 + 3GS
	100	–	4.4e-2	38484	38484	3 + 3GS
MPBNGC	2	–	4.1e-15	40	40	37
	5	–	1.4e-1	10000	10000	103
	10	–	1.5e-3	10000	10000	3347
	20	–	1.2e-2	10000	10000	5010
	50	–	3.3e-1	10000	10000	3338
	100	–	4.0e-1	10000	10000	3338

	n	q^0	f^*	$\#f$	$\#\nabla f$	Iter
LiPsMin	2	0.1	5.6e-9	37	63	15
	5	0.1	1.4e-9	47	132	22
	10	0.1	4.2e-9	68	309	33
	20	0.1	2.9e-9	74	642	36
	50	0.01	3.8e-9	131	2109	64
	100	0.01	5.0e-10	166	4562	79
HANSO	2	–	3.2e-19	18	18	16 (*9)
	5	–	3.0e-19	242	242	116 (*47)
	10	–	6.2e-17	787	787	352 (*88)
	20	–	1.1e-16	1362	1362	637 (*221)
	50	–	2.1e-16	4409	4409	1906 (*494)
	100	–	3.0e-16	8922	8922	3991 (*1023)
MPBNGC	2	–	7.6e-9	15	15	14
	5	–	3.1e-9	60	60	49
	10	–	3.4e-9	126	126	34
	20	–	2.6e-9	244	244	222
	50	–	3.8e-9	577	577	549
	100	–	4.5e-9	1118	1118	1083

Supporting hyperplane condition

$$\begin{aligned} & \phi(x) \geq \phi(x_*) + g^\top (x - x_*) - o(\|x - x_*\|) \\ \iff \quad & \phi(x) - g^\top (x - x_*) \text{ first order minimal} \\ \iff \quad & b^\top \geq |\lambda|^\top - \lambda^\top L = |\lambda|^\top (I - \Sigma L) \end{aligned}$$

represents normal growth for some λ, which can be found by LOP.

Full local convexity condition

$$b^\top (I - DL)^{-1} \ \geq \ 0 \quad \text{if} \quad |D| \leq 1 \, ,$$

where D ranges over all diagonal matrices, requiring exponential test.

Intermediate Summary and Conclusion

- Piecewise smooth functions typically arise in abs-normal form.

Intermediate Summary and Conclusion

- Piecewise smooth functions typically arise in abs-normal form.
- Then Linear Independence Kink Qualification (LIKQ) is generic and makes local combinatorial structure manageable.

- Piecewise smooth functions typically arise in abs-normal form.
- Then Linear Independence Kink Qualification (LIKQ) is generic and makes local combinatorial structure manageable.
- Tight first and second order minimality conditions are obtained in terms of vectors and matrices computable by an extension of AD.

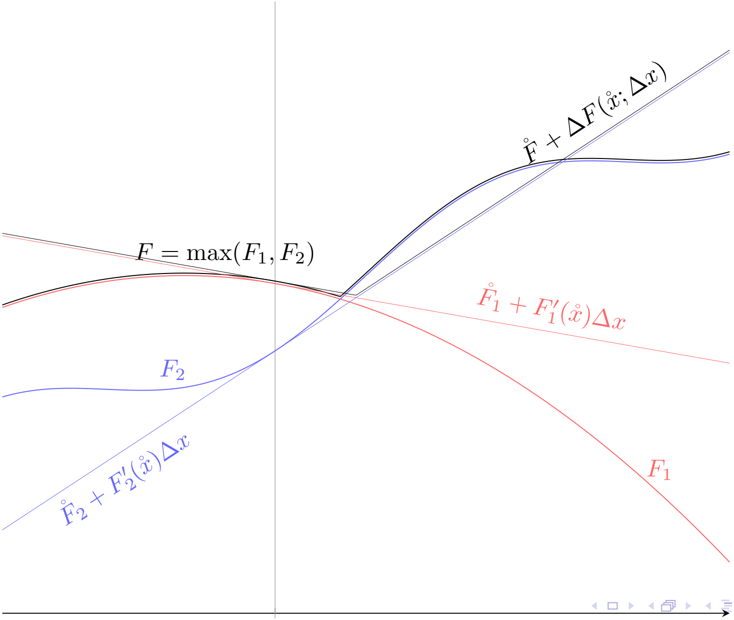
Intermediate Summary and Conclusion

- Piecewise smooth functions typically arise in abs-normal form.
- Then Linear Independence Kink Qualification (LIKQ) is generic and makes local combinatorial structure manageable.
- Tight first and second order minimality conditions are obtained in terms of vectors and matrices computable by an extension of AD.
- Violations of the necessary conditions yield descent directions or parabolas constructively.

Intermediate Summary and Conclusion

- Piecewise smooth functions typically arise in abs-normal form.
- Then Linear Independence Kink Qualification (LIKQ) is generic and makes local combinatorial structure manageable.
- Tight first and second order minimality conditions are obtained in terms of vectors and matrices computable by an extension of AD.
- Violations of the necessary conditions yield descent directions or parabolas constructively.
- Various algorithmic approaches promise linear, superlinear or quadratic convergence under LIKQ.
- Under additional nonredundancy condition $\mathcal{V} = (Z^\top)$ for $\mathcal{VU}$ decomposition.

Basic idea of tangent linearization:



Smoothies

$$\begin{aligned} \Delta v &= \Delta u + \Delta w \quad \text{or} \quad \Delta v = u\,\Delta w + w\,\Delta u, \\ \Delta v &= \varphi'(u)\,\Delta u \quad \text{if} \quad v = \varphi(u). \end{aligned}$$

Smoothies

$$\begin{aligned}\Delta v &= \Delta u + \Delta w \quad \text{or} \quad \Delta v = u\,\Delta w + w\,\Delta u,\\ \Delta v &= \varphi'(u)\,\Delta u \quad \text{if} \quad v = \varphi(u).\end{aligned}$$

Kinks

$$v + \Delta v = \text{abs}(u + \Delta u) \quad \Leftrightarrow \quad \Delta v = \text{abs}(u + \Delta u) - \text{abs}(u)$$

Smoothies

$$\Delta v = \Delta u + \Delta w \quad \text{or} \quad \Delta v = u\,\Delta w + w\,\Delta u,$$
$$\Delta v = \varphi'(u)\,\Delta u \quad \text{if} \quad v = \varphi(u).$$

Kinks

$$v + \Delta v = \text{abs}(u + \Delta u) \quad \Leftrightarrow \quad \Delta v = \text{abs}(u + \Delta u) - \text{abs}(u)$$

Jumps

$$\text{if } (\text{firstsign}(u + \Delta u, \nabla u) < 0) \quad \text{goto} \quad (\text{statement } \#\text{m})$$

where ∇u is lexicographic derivative of u and firstsign(a) is sign of first nonzero component of a vector $a \in \mathbb{R}^{1+n}$.

Smoothies

$$\begin{aligned} \Delta v &= \Delta u + \Delta w \quad \text{or} \quad \Delta v = u\,\Delta w + w\,\Delta u, \\ \Delta v &= \varphi'(u)\,\Delta u \quad \text{if} \quad v = \varphi(u). \end{aligned}$$

Kinks

$$v + \Delta v = \text{abs}(u + \Delta u) \quad \Leftrightarrow \quad \Delta v = \text{abs}(u + \Delta u) - \text{abs}(u)$$

Jumps

$$\text{if (firstsign}(u + \Delta u, \nabla u) < 0) \quad \text{goto} \quad (\text{statement } \#\text{m})$$

where ∇u is lexicographic derivative of u and firstsign(a) is sign of first nonzero component of a vector $a \in \mathbb{R}^{1+n}$.

Resulting Mapping

$$\Delta x \mapsto \Delta y \text{ for fixed } x \text{ denoted by } \Delta y = \Delta F(x; \Delta x)$$

Linearity and Product Rule

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

$$\Longrightarrow$$

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

Resulting Mapping

$$\Delta x \mapsto \Delta y \text{ for fixed } x \text{ denoted by } \Delta y = \Delta F(x; \Delta x)$$

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 (Level 1 case)

If F is Level 1 on a closed convex domain $\mathcal{K} \subset \mathbb{R}^n$ then there exists a constant γ such that for all triplets $\hat{x}, \check{x}, x \in \mathcal{K}$

$$\|[F(\hat{x}) + \Delta F(\hat{x}; x - \hat{x})] - [F(\check{x}) + \Delta F(\check{x}; x - \check{x})]\|$$
$$\leq \tfrac{1}{2}\gamma\|\hat{x} - \check{x}\|[\|x - \hat{x}\| + \|x - \check{x}\|]$$

Proposition (Level 1 case)

If F is Level 1 on a closed convex domain $\mathcal{K} \subset \mathbb{R}^n$ then there exists a constant γ such that for all triplets $\hat{x}, \check{x}, x \in \mathcal{K}$

$$\|[F(\hat{x}) + \Delta F(\hat{x}; x - \hat{x})] - [F(\check{x}) + \Delta F(\check{x}; x - \check{x})]\|$$
$$\leq \tfrac{1}{2}\gamma\|\hat{x} - \check{x}\|[\|x - \hat{x}\| + \|x - \check{x}\|]$$

$$\Longrightarrow \quad \|[F(x) - [F(\mathring{x}) + \Delta F(\mathring{x}; x - \mathring{x})]\| \leq \tfrac{1}{2}\gamma\|x - \mathring{x}\|^2$$

Proposition (Level 1 case)

If F is Level 1 on a closed convex domain $\mathcal{K} \subset \mathbb{R}^n$ then there exists a constant γ such that for all triplets $\hat{x}, \check{x}, x \in \mathcal{K}$

$$\|[F(\hat{x}) + \Delta F(\hat{x}; x - \hat{x})] - [F(\check{x}) + \Delta F(\check{x}; x - \check{x})]\| \\ \leq \tfrac{1}{2}\gamma\|\hat{x} - \check{x}\|[\|x - \hat{x}\| + \|x - \check{x}\|]$$

$$\implies \quad \|[F(x) - [F(\mathring{x}) + \Delta F(\mathring{x}; x - \mathring{x})]\| \leq \tfrac{1}{2}\gamma\|x - \mathring{x}\|^2$$

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

Proposition (Level 1 case)

If F is Level 1 on a closed convex domain $\mathcal{K} \subset \mathbb{R}^n$ then there exists a constant γ such that for all triplets $\hat{x}, \check{x}, x \in \mathcal{K}$

$$\|[F(\hat{x}) + \Delta F(\hat{x}; x - \hat{x})] - [F(\check{x}) + \Delta F(\check{x}; x - \check{x})]\| \\ \leq \tfrac{1}{2}\gamma\|\hat{x} - \check{x}\|[\|x - \hat{x}\| + \|x - \check{x}\|]$$

$$\Longrightarrow \quad \|[F(x) - [F(\mathring{x}) + \Delta F(\mathring{x}; x - \mathring{x})]\| \leq \tfrac{1}{2}\gamma\|x - \mathring{x}\|^2$$

$$\Longrightarrow \quad (\|\Delta F(\hat{x}; \Delta x) - \Delta F(\check{x}; \Delta x)\|)/(1 + \|\Delta x\|) \leq \gamma\|\hat{x} - \check{x}\|$$

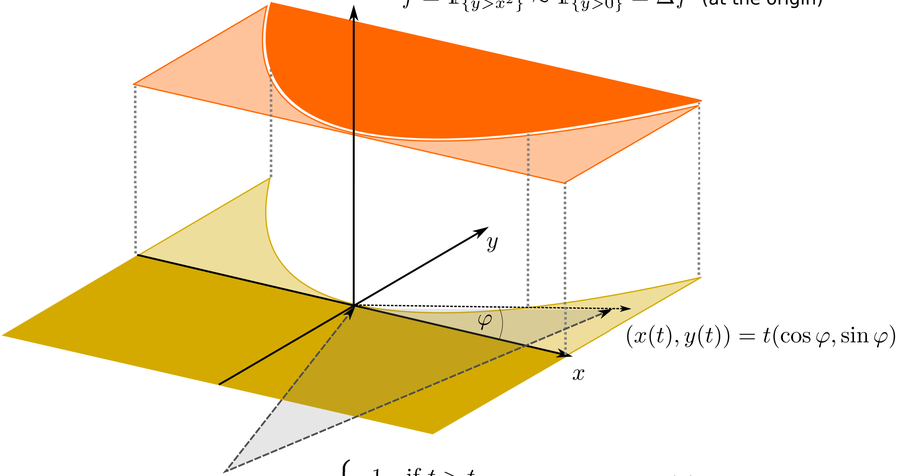
Finally there is a continuous radius $\rho(x)$ such that

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

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

Piecewise Linearization of Discontinuous f

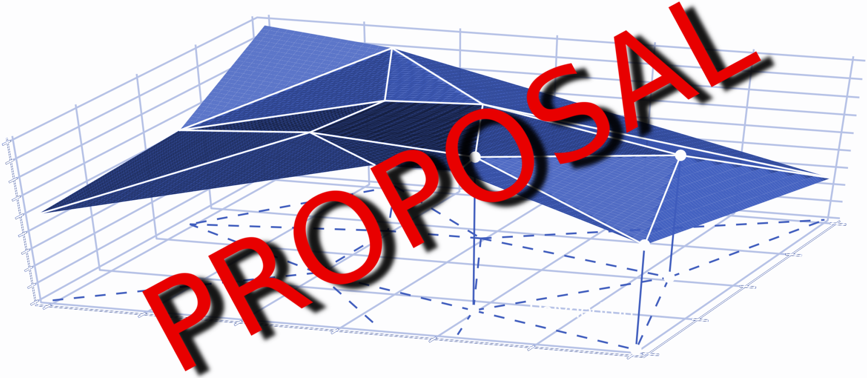
$f = \mathbb{1}_{\{y>x^2\}} \approx \mathbb{1}_{\{y>0\}} = \Delta f$ (at the origin)



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

Function Space:	Diff.Op.:	Model Space:	Discrepancy:
Level 0	$\overset{\partial\|_{\mathring{x}}}{\longmapsto}$ Lip	$L =$ linear	uniform
$\cap$		$\cap$	
Level 1	$\overset{\Delta\|_{\mathring{x}}}{\longmapsto}$ Lip	$PL =$ Piecewise L	uniform
$\cap$		$\downarrow \ \partial^B\|_{\mathring{x}}$	
Level 2	$\overset{\partial^B\|_{\mathring{x}}}{\longmapsto}$???	$PL_h =$ homog. PL	nonuniform
$\cap$		$\cap$	
Level 3	$\overset{\Delta\|_{\mathring{x}}}{\longmapsto}$???	$DPL =$ discont. PL	nonuniform

Numerical Methods for Nonsmooth Problems,
Applications of Algorithmic Piecewise Differentiation



A.Griewank, A. Walther, T.Bosse, T.Munson

Benefits of piecewise linearization

- Continuous piecewise smooth (PS) functions can be locally approximated with second order error by piecewise linearization (PL).

Benefits of piecewise linearization

- Continuous piecewise smooth (PS) functions can be locally approximated with second order error by piecewise linearization (PL).
- PL model varies Lipschitz continuously with development point !!!

Benefits of piecewise linearization

- Continuous piecewise smooth (PS) functions can be locally approximated with second order error by piecewise linearization (PL).
- PL model varies Lipschitz continuously with development point !!!
- PL model can be generated effectively by minor modification of algorithmic differentiation tools allowing for abs, min and max.

Benefits of piecewise linearization

- Continuous piecewise smooth (PS) functions can be locally approximated with second order error by piecewise linearization (PL).
- PL model varies Lipschitz continuously with development point !!!
- PL model can be generated effectively by minor modification of algorithmic differentiation tools allowing for abs, min and max.
- Generalized Jacobians of PL models can be computed and represent (conically active) generalized Jacobians of underlying PS function.

Benefits of piecewise linearization

- Continuous piecewise smooth (PS) functions can be locally approximated with second order error by piecewise linearization (PL).
- PL model varies Lipschitz continuously with development point !!!
- PL model can be generated effectively by minor modification of algorithmic differentiation tools allowing for abs, min and max.
- Generalized Jacobians of PL models can be computed and represent (conically active) generalized Jacobians of underlying PS function.
- Prospects for use of PL models for equation solving, (un)constrained optimization and integration of Lipschitzian ODEs under way.

- Continuous piecewise smooth (PS) functions can be locally approximated with second order error by piecewise linearization (PL).
- PL model varies Lipschitz continuously with development point !!!
- PL model can be generated effectively by minor modification of algorithmic differentiation tools allowing for abs, min and max.
- Generalized Jacobians of PL models can be computed and represent (conically active) generalized Jacobians of underlying PS function.
- Prospects for use of PL models for equation solving, (un)constrained optimization and integration of Lipschitzian ODEs under way.
- Combintorial structure can be dealt with at the piecewise linear level.

- Continuous piecewise smooth (PS) functions can be locally approximated with second order error by piecewise linearization (PL).
- PL model varies Lipschitz continuously with development point !!!
- PL model can be generated effectively by minor modification of algorithmic differentiation tools allowing for abs, min and max.
- Generalized Jacobians of PL models can be computed and represent (conically active) generalized Jacobians of underlying PS function.
- Prospects for use of PL models for equation solving, (un)constrained optimization and integration of Lipschitzian ODEs under way.
- Combintorial structure can be dealt with at the piecewise linear level.
- Discontinuous PS functions lead to algebraic and differential inclusions and thus automatic event handling in ODE case.

References

- A.G.: On stable piecewise linearization and generalized algorithmic differentiation. *Optimization Methods and Software*, (2013).

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 ODE's, *Mongolian Journal of Mathematics* (2014).

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 ODE's, *Mongolian Journal of Mathematics* (2014).
- A.G., M. Radons, T. Streubel, J.U. Bernd: On piecewise linear equations in abs-normal form. *Linear Algebra and Applications* (2015).

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 ODE's, *Mongolian Journal of Mathematics* (2014).
- A.G., M. Radons, T. Streubel, J.U. Bernd: On piecewise linear equations in abs-normal form. *Linear Algebra and Applications* (2015).
- A.G., A. Walther, S. Fiege, T. Bosse: On Lipschitz optimization based on gray-box piecewise linearization *to appear in Mathematical Programming, (2016)*

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 ODE's, *Mongolian Journal of Mathematics* (2014).
- A.G., M. Radons, T. Streubel, J.U. Bernd: On piecewise linear equations in abs-normal form. *Linear Algebra and Applications* (2015).
- A.G., A. Walther, S. Fiege, T. Bosse: On Lipschitz optimization based on gray-box piecewise linearization *to appear in Mathematical Programming, (2016)*
- A.G., A. Walther, S. Fiege, K. Kulshreshta: An Algorithm for Nonsmooth Optimization by Successive Piecewise Linearization, *to be submitted to Mathematical Programming, (2016)*

- 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 ODE's, *Mongolian Journal of Mathematics* (2014).
- A.G., M. Radons, T. Streubel, J.U. Bernd: On piecewise linear equations in abs-normal form. *Linear Algebra and Applications* (2015).
- A.G., A. Walther, S. Fiege, T. Bosse: On Lipschitz optimization based on gray-box piecewise linearization *to appear in Mathematical Programming, (2016)*
- A.G., A. Walther, S. Fiege, K. Kulshreshta: An Algorithm for Nonsmooth Optimization by Successive Piecewise Linearization, *to be submitted to Mathematical Programming, (2016)*
- A.G., A. Walther: First and second order optimality conditions for piecewise smooth functions. *to appear in OMS, (2017)*