Representation and analysis of continuous piecewise linear functions in abs-normal form

Tom Streubel
Andreas Griewank, Jens Uwe Bernt, Manuel Radons

Institute for Applied Mathematics, Humboldt University at Berlin, Germany
streubel@math.hu-berlin.de

14th European Workshop on Automatic Differentiation
09th & 10th December 2013 at University of Oxford.
Contents

1 Motivation

2 The abs-normal Form

3 Conditions for solvability

4 Original piecewise linear problem - OPL

5 Unfolded piecewise linear problem - UPL

6 Outlook
Example Consider an ODE and apply implicit Euler

\[
\dot{x}(t) = F(x(t)), \text{ where } F \text{ is piecewise smooth} \tag{1}
\]

solve \(x_+ - x = hF(x_+)\) for \(x_+\) \tag{2}
Example Consider an ODE and apply implicit Euler

\[
\dot{x}(t) = F(x(t)), \text{ where } F \text{ is piecewise smooth} \tag{1}
\]

solve \( x^{(i)}_+ - x = h \cdot \Box_{x^{(i-1)}_+} F(x^{(i)}_+) \) for \( x^{(i)}_+ \) repeatedly \tag{2}

\[\uparrow\]
\text{piecewise linearization}
Example Consider an ODE and apply implicit Euler

\[ \dot{x}(t) = F(x(t)), \text{ where } F \text{ is piecewise smooth} \]  
(1)

\[ \text{solve } x_{+}^{(i)} - x = h \cdot \Box_{x_{+}^{(i-1)}} F(x_{+}^{(i)}) \text{ for } x_{+}^{(i)} \text{ repeatedly} \]  
(2)

\[ \uparrow \text{piecewise linearization} \]

Example\(^1\) Let \( T \in \mathbb{R}^{n\times n} \) be an irreducible, symmetric, positive semidefinite matrix and \( x, e \in \mathbb{R}^{n} \) be vectors

\[ \text{solve } \max(0, x) + Tx = -e/2 \text{ for } x \]

\(^1\) *Iterative Solution Of Piecewise Linear Systems* - by Luigi Brugnano and Vincenzo Casulli
An introductory example: \( F(x_1, x_2) = \left[ \frac{x_1 + |x_1 - x_2| + |x_1 - |x_2||}{x_2} \right] \)

\[ z_1 = x_1 - x_2 \quad z_2 = x_2 \quad z_3 = x_1 - |z_2| \]
The computational graph

section 2 – The abs-normal Form

\[ F(x_1, x_2) = \begin{bmatrix} x_1 + |z_1| + |z_3| \\ x_2 \end{bmatrix} \]

\[ z_1 = x_1 - x_2 \quad z_2 = x_2 \quad z_3 = x_1 - |z_2| \]

\[
\begin{bmatrix}
  z_1 \\
  z_2 \\
  z_3 \\
  y_1 \\
  y_2
\end{bmatrix} = 
\begin{bmatrix}
  0 \\
  0 \\
  0 \\
  0 \\
  0
\end{bmatrix} + 
\begin{bmatrix}
  1 & -1 \\
  0 & 1 \\
  1 & 0 \\
\end{bmatrix} 
\begin{bmatrix}
  x_1 \\
  x_2
\end{bmatrix}
\]
The computational graph

section 2 – The abs-normal Form

\[ F(x_1, x_2) = \begin{bmatrix} x_1 + |z_1| + |z_3| \\ x_2 \end{bmatrix} \]

\[ z_1 = x_1 - x_2 \quad z_2 = x_2 \quad z_3 = x_1 - |z_2| \]

\[
\begin{bmatrix}
  z_1 \\
  z_2 \\
  z_3 \\
  y_1 \\
  y_2
\end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 \end{bmatrix} \begin{bmatrix}
  x_1 \\
  x_2 \\
  |z_1| \\
  |z_2| \\
  |z_3|
\end{bmatrix}
\]
The computational graph

section 2 – The abs-normal Form

\[ F(x_1, x_2) = \begin{bmatrix} x_1 + |z_1| + |z_3| \\ x_2 \end{bmatrix} \]

\[ z_1 = x_1 - x_2 \quad z_2 = x_2 \quad z_3 = x_1 - |z_2| \]

\[
\begin{bmatrix}
  z_1 \\
  z_2 \\
  z_3 \\
  y_1 \\
  y_2
\end{bmatrix} = \begin{bmatrix}
  0 \\
  0 \\
  0 \\
  0 \\
  0
\end{bmatrix} + \begin{bmatrix}
  1 & -1 & 0 & 0 & 0 \\
  0 & 1 & 0 & 0 & 0 \\
  1 & 0 & 0 & -1 & 0 \\
  1 & 0 \\
  0 & 1
\end{bmatrix} \begin{bmatrix}
  x_1 \\
  x_2 \\
  |z_1| \\
  |z_2| \\
  |z_3|
\end{bmatrix}
\]
The computational graph

section 2 – The abs-normal Form

\[
F(x_1, x_2) = \begin{bmatrix}
  x_1 + |z_1| + |z_3| \\
  x_2
\end{bmatrix}
\]

\[
z_1 = x_1 - x_2 \quad z_2 = x_2 \quad z_3 = x_1 - |z_2|
\]

\[
\begin{bmatrix}
  z_1 \\
  z_2 \\
  z_3 \\
  y_1 \\
  y_2
\end{bmatrix} = \begin{bmatrix}
  0 \\
  0 \\
  0 \\
  0 \\
  0
\end{bmatrix} + \begin{bmatrix}
  1 & -1 & 0 & 0 & 0 \\
  0 & 1 & 0 & 0 & 0 \\
  1 & 0 & 0 & -1 & 0 \\
  1 & 0 & 1 & 0 & 1 \\
  0 & 1 & 0 & 0 & 0
\end{bmatrix} \begin{bmatrix}
  x_1 \\
  x_2 \\
  |z_1| \\
  |z_2| \\
  |z_3|
\end{bmatrix}
\]
Definition Abs-normal form for $F: \mathbb{R}^n \to \mathbb{R}^m$ PL

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

(3)

$Z \in \mathbb{R}^{s \times n}, L \in \mathbb{R}^{s \times s}, J \in \mathbb{R}^{n \times m}, Y \in \mathbb{R}^{n \times s}$, $c \in \mathbb{R}^s, b \in \mathbb{R}^m$
Definition: Abs-normal form for $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ PL

\[
\begin{bmatrix}
z \\
y
\end{bmatrix} = 
\begin{bmatrix}
c \\
b
\end{bmatrix} + 
\begin{bmatrix}
Z & L \\
J & Y
\end{bmatrix} 
\begin{bmatrix}
x \\
|z|
\end{bmatrix}
\] (3)

- $L$ is strictly lower triangular
- if $L = 0$ then (3) is said to be **simply switched** (KKT, LCP)

$Z \in \mathbb{R}^{s \times n}$, $L \in \mathbb{R}^{s \times s}$, $J \in \mathbb{R}^{n \times m}$, $Y \in \mathbb{R}^{n \times s}$, $c \in \mathbb{R}^s$, $b \in \mathbb{R}^m$
The abs-normal form

section 2 – The abs-normal Form

Definition Abs-normal form for $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ PL

\[
\begin{bmatrix}
z \\
y
\end{bmatrix} =
\begin{bmatrix}
c \\
b
\end{bmatrix} +
\begin{bmatrix}
Z & L \\
J & Y
\end{bmatrix}
\begin{bmatrix}
x \\
|z|
\end{bmatrix}
\] (3)

- $L$ is strictly lower triangular
- If $L = 0$ then (3) is said to be **simply switched** (KKT, LCP)
- Antecedents by Bokhoven et al in electrical engineering
- Representation stable and consistent w.r.t numerical perturbations

$Z \in \mathbb{R}^{s \times n}, L \in \mathbb{R}^{s \times s}, J \in \mathbb{R}^{n \times n}, Y \in \mathbb{R}^{n \times s}, c \in \mathbb{R}^s, b \in \mathbb{R}^n$
The abs-normal form

section 2 – The abs-normal Form

Definition Abs-normal form for $F : \mathbb{R}^n \to \mathbb{R}^n$ PL

\[
\begin{bmatrix}
z \\
y
\end{bmatrix} = \begin{bmatrix}
c \\
b
\end{bmatrix} + \begin{bmatrix}
Z & L \\
J & Y
\end{bmatrix} \begin{bmatrix}
x \\
|z|
\end{bmatrix}
\] (3)

- The sign vector $\sigma(x) \equiv \text{sign}(z(x)), \sigma \in \{-1, 0, 1\}^s$
- The control flow as $\Sigma \equiv \text{diag}(\sigma), \Sigma_x \equiv \text{diag}(\sigma(x))$
- Each sign pattern $\sigma_x$ labels a polyhedron of the decomposition
- Our interest here: square case $m = n$

$Z \in \mathbb{R}^{s \times n}, L \in \mathbb{R}^{s \times s}, J \in \mathbb{R}^{n \times n}, Y \in \mathbb{R}^{n \times s}, c \in \mathbb{R}^s, b \in \mathbb{R}^n$
### Definition

Abs-normal form for $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ PL

\[
\begin{bmatrix}
z \\ y
\end{bmatrix} =
\begin{bmatrix}
c \\ b
\end{bmatrix} +
\begin{bmatrix}
Z & L \\ J & Y
\end{bmatrix}
\begin{bmatrix}
x \\ \Sigma x \cdot z
\end{bmatrix}
\]

Let $y$ be fixed. Take the second row, solve for $x$ and plug into the 1st

\[
z = c - ZJ^{-1}(b - y) + (L - ZJ^{-1}Y)\Sigma z \iff (I - S\Sigma)z - \hat{c} = 0
\]

Take the first row, solve for $z$ and plug into the 2nd

\[
F(x) \equiv y = b + Y\Sigma(I - L\Sigma)^{-1}c + (J + Y\Sigma(I - L\Sigma)^{-1}Z)x \equiv b_{\sigma(x)} + J_{\sigma(x)}
\]
Consider the piecewise linear function $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ with

$$F(x) = b_{\sigma(x)} + J_{\sigma(x)} \cdot x$$

By the Sherman-Morrison-Woodbury formula we have

$$\det(J_{\sigma(x)}) = \det(J) \det(I - S\Sigma_x)$$

The PL function $F$ is said to be *coherently oriented* (c.o.) iff all arising $J_\sigma$ have the same nonzero determinant sign, which can be assumed w.l.o.g. to be positive

$$\forall x \in \mathbb{R}^n \det(I - S\Sigma_x) > 0 \iff F \text{ is c.o.}$$
The real spectral radius\(^2\)

\[ \rho_0(S) = \max\{|\lambda| \mid \lambda \text{ is a real eigenvalue of } S\} \]

and the (continuous) sign real spectral radius\(^2\)

\[ \rho_s^0(S) = \max \{ \rho_0(S\Sigma) \mid \text{for } \Sigma = \text{diag}(\sigma), \text{ with } \sigma \in \{-1, 0, 1\}^s \} \]

**Theorem by Rump\(^2\)**

\[ \forall \sigma \in \{-1, 0, 1\}^s : \det(I - S\Sigma) > 0 \iff \rho_s^0(S) < 1 \iff (S - I)^{-1}(S + I) \text{ is a P-Matrix} \]

---

\(^2\)Theorems of Perron-Frobenius type for matrices without sign restrictions - by Sigfried Rump
Smooth dominance

section 3 – Conditions for solvability

- Checking $S$ for $\rho_0^s(S) < 1$ is NP-hard

**Definition: Smooth dominance**

$F : \mathbb{R}^n \to \mathbb{R}^n$ in abs-normal form is called smooth dominant, if for some nonsingular diagonal matrix $D$ and a $p \in [1, \infty]$

$$\|DSD^{-1}\|_p < 1$$
Checking $S$ for $\rho_0^s(S) < 1$ is NP-hard

**Definition: Smooth dominance**

$F : \mathbb{R}^n \to \mathbb{R}^n$ in abs-normal form is called smooth dominant, if for some nonsingular diagonal matrix $D$ and a $p \in [1, \infty]$

$$\| DSD^{-1} \|_p < 1$$

If $|S|$ is not permuted block triangular, then there is a positive eigenvector $d$ for $\rho(|S|)$, such that, with $D = \text{diag}(d)$

$$\| D^{-1}SD \|_\infty = \rho(|S|) \text{ (Perron-Frobenius scaling)}$$

Thus: $\rho(|S|) < 1 \implies S$ smooth dominant
Definition Linear independence kink qualification

A simply switched \((L = 0)\) abs-normal System satisfies LIKQ, if the normals of all hyperplanes (defined by \(0 = z_i(x) \equiv c_i + e_i^\top Zx\)) intersecting at some point are linearly independent.
Definition Linear independence kink qualification

A simply switched \((L = 0)\) abs-normal System satisfies LIKQ, if the normals of all hyperplanes (defined by \(0 = z_i(x) \equiv c_i + e_i^\top Zx\)) intersecting at some point are linearly independent.

\[
f = \begin{bmatrix} x_1 + |z_1| + |z_2| - |z_3| \end{bmatrix}
\]

\[
\begin{align*}
z_1 &= x_1 - x_2 \\
z_2 &= x_1 \\
z_3 &= x_2
\end{align*}
\]
**Definition Linear independence kink qualification**

A simply switched ($L = 0$) abs-normal System satisfies LIKQ, if the normals of all hyperplanes (defined by $0 = z_i(x) \equiv c_i + e_i^\top Zx$) intersecting at some point are linearly independent.

\[
f = \begin{bmatrix} x_1 + |z_1| + |z_2| - |z_3| \\ x_2 \end{bmatrix}
\]

\[
\begin{align*}
z_1 &= x_1 - x_2 \\
z_2 &= x_1 \\
z_3 &= x_2
\end{align*}
\]
Definition Linear independence kink qualification

A simply switched ($L = 0$) abs-normal System satisfies LIKQ, if the normals of all hyperplanes (defined by $0 = z_i(x) \equiv c_i + e_i^\top Zx$) intersecting at some point are linearly independent.

$$f = \begin{bmatrix} x_1 + |z_1| + |z_2| - |z_3| \\ x_2 \\ z_1 = x_1 - x_2 \\ z_2 = x_1 \\ z_3 = x_2 \end{bmatrix}$$
**Definition Linear independence kink qualification**

A simply switched ($L = 0$) abs-normal System satisfies LIKQ, if the normals of all hyperplanes (defined by $0 = z_i(x) \equiv c_i + e_i^\top Zx$) intersecting at some point are linearly independent.
**Definition Linear independence kink qualification**

A simply switched \((L = 0)\) abs-normal System satisfies LIKQ, if the normals of all hyperplanes (defined by \(0 = z_i(x) \equiv c_i + e_i^\top Zx\)) intersecting at some point are linearly independent.

\[
\tilde{f} = \begin{bmatrix}
x_1 + |z_1| + |z_2| - |z_3| \\
x_2
\end{bmatrix}
\]

\[
z_1 = x_1 - x_2 - 1 \\
z_2 = x_1 \\
z_3 = x_2
\]

**Note** \(H(z) = (I - S\Sigma)z - \hat{c}\) satisfies LIKQ
From Scholtes we know for $F$ PL

$F$ homeom. $\iff$ $F$ injective $\iff$ $F$ c.o. $\iff$ $F$ surjective

$F$ open
From Scholtes we know for $F$ PL

$F$ homeom. $\iff$ $F$ injective $\not\implies$ $F$ c.o. $\not\implies$ $F$ surjective

$F$ open
... remember

coherent orientation (c.o.)  smooth dominance (s.d.)  LIKQ

\[
\begin{align*}
F \text{ homeom.} & \iff F \text{ injective} & \iff F \text{ c.o.} & \iff F \text{ surjective} \\
& & \iff F \text{ open}
\end{align*}
\]
... remember

**coherent orientation** (c.o.)  **smooth dominance** (s.d.)  LIKQ

\[
\begin{align*}
S \text{ s.d.} & \iff H \text{ bijective} & \iff H \text{ c.o.} \\
F \text{ homeom.} & \iff F \text{ injective} & \not\iff F \text{ c.o.} & \not\iff F \text{ surjective} \\
& & \iff F \text{ open}
\end{align*}
\]
1 Motivation
2 The abs-normal Form
   - The computational graph
   - Block eliminations
3 Conditions for solvability
   - Coherent Orientation & Sign real spectral radius
   - Smooth dominance & LIKQ
4 Original piecewise linear problem - OPL
   - Newton variants on OPL
   - Alternating block Seidel iteration
5 Unfolded piecewise linear problem - UPL
   - Modulus Algorithm on UPL
   - Newton variants on UPL
6 Outlook
- All PL functions are semi smooth
All PL functions are semi smooth

A limiting Jacobian can be evaluated from abs-normal form using polynomial escape
All PL functions are semi smooth
A limiting Jacobian can be evaluated from abs-normal form using polynomial escape
Hence semi smooth Newton a la Qi & Sun can be applied with their local convergence result holding trivially
- All PL functions are semi smooth
- A limiting Jacobian can be evaluated from abs-normal form using polynomial escape
- Hence semi smooth Newton a la Qi & Sun can be applied with their local convergence result holding trivially
- In terms of the abs-normal form, we formulated sufficient conditions for global full step convergence in finitely many steps
All PL functions are semi smooth

A limiting Jacobian can be evaluated from abs-normal form using polynomial escape

Hence semi smooth Newton a la Qi & Sun can be applied with their local convergence result holding trivially

In terms of the abs-normal form, we formulated sufficient conditions for global full step convergence in finitely many steps

Alternatively one may apply piecewise Newton which only requires coherent orientation, rather than strong contractivity
Let be $y = 0$ fixed. From elimination of $x$ (unfolding) we know

$$z : \mathbb{R}^n \rightarrow \mathbb{R}^s \quad z(x) = (I - L\Sigma_x)^{-1}(c + Zx)$$

and from the lower half of the abs-normal form

$$x : \mathbb{R}^s \rightarrow \mathbb{R}^n \quad x(z) = J^{-1}(b + Y\Sigma_xz)$$

So we can define a fixed point iteration

$$z_+ = z(x(z))$$

which converges from everywhere towards the unique solution $z^*$, if

$$\|S\|_p \leq \|L\|_p + \|ZJ^{-1}Y\|_p < 1$$
\[ 0 \triangleright H(z) = (I - S\Sigma)z - \hat{c} \iff z = S|z| - \hat{c} \]

**modulus algorithm (Bokhoven)**

\[ z_+ = S|z| - \hat{c} \]

- With Banach FP theorem, it converges in smooth dominant case

\[ \|H(z) - H(\tilde{z})\|_p = \|S(|z| - |\tilde{z}|)\|_p \leq \|S\|_p \|z - \tilde{z}\|_p \]

- **BUT**: Generally yields linear rather than finite convergence
Definition: Generalized Newton on UPL: $H(z) \overset{!}{=} 0$

\[ z_+ = z - (I - S\Sigma)^{-1}H(z) = (I - S\Sigma)^{-1}\hat{c} \]
Definition: Generalized Newton on UPL: \( H(z) = 0 \)

\[
z_+ = z - (I - S\Sigma)^{-1}H(z) = (I - S\Sigma)^{-1}\hat{c}
\]

Note that the new iterative only depends on the old flow control \( \Sigma \)

- Since only finitely many sign combinations are possible, either
  - The iteration cycles
  - Finite convergence occurs
- The iterative process can be represented as a graph of sign combinations \( \sigma \), as vertices
- Branches arise when ever an entry of \( \sigma \) is 0 and must be replaced by either 1 or \(-1\) (does not happen for generic \( \hat{c} \))
Consider the following matrix-vector pair

\[
S = \begin{bmatrix}
0 & 0 & \frac{1}{3} \\
\frac{1}{3} & 0 & 0 \\
0 & \frac{1}{3} & 0 \\
\end{bmatrix}, \quad \hat{c} = \begin{bmatrix}
7 \\
1 \\
1 \\
\end{bmatrix}
\]

which uniquely defines a UPL

\[
H(z) = (I - S\Sigma)z - \hat{c} = 0
\]

The graph generated by generalized Newton
If $\rho(|S|) \leq \frac{1}{2}$, at least one additional component of $z$ attains its final sign in each semi smooth Newton iteration.

Then the iteration terminates in at most $s$ steps and can be arranged s.t. the total computational effort is exactly $\frac{1}{3}s^3$ fused multiply-adds.

The Condition $\rho(|S|) \leq \frac{1}{2}$ is rather sharp, since divergence can occur when $\rho(|S|) > \frac{1}{2} + \frac{1}{2s}$ as shown by a cyclic example.
Consider

\[
S = \begin{bmatrix}
0 & a \\
a \cdot I_{s-1} & 0
\end{bmatrix}, \quad \hat{c} = \begin{bmatrix}
1 \\
\vdots \\
1
\end{bmatrix}
\]

where the gen. Newton does not converge for

\[
\frac{1}{2} + \frac{1}{2^s} \leq a \leq \frac{1}{\sqrt{2}}
\]

and a suitable starting vector, with one negative and else positive entries.

The graph generated by generalized Newton for \( s = 3 \)
Consider the second motivation example from Brugnano & Casulli

\[ \max(0, x) + Tx = -e/2 \iff (2T + I)x + |x| + b = 0 \]

\[ T = \text{tridiag}(-1, 2, -1) \]

and let \(-e/2\) be such \(x^* = \left(\exp\left[6 \frac{i-1}{n-1} - 5\right]\right)_{i=1}^{n}\) is the solution.

Then \(S = -(2T + I)^{-1}\), with \(\rho(S) \approx 1 - \frac{10}{n^2}\)

**Converging table for \(n = 1000\) and \(x_0 = (1, \ldots, 1)\)\(^\top\)**

<table>
<thead>
<tr>
<th>Method</th>
<th>Steps</th>
<th>Until (|H|_\infty &lt; \text{eps})</th>
<th>Steps</th>
<th>Until (|H|_\infty &lt; 1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ssN-UPL</td>
<td>5</td>
<td></td>
<td>5</td>
<td>10,000</td>
</tr>
</tbody>
</table>
Efficient stable numerical linear algebra (PLAN-C)
Piecewise Newton implementation
Comparative testing and tuning
Embedding in outer nonlinear loop
Branin variant for incoherently oriented problems
The abs-normal form of a PL function $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$

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

With the trivial identity $x = |x + |x|| - |x|$ applied on every component of $J = J \pm I = |J - I + |J - I|| - |J - I| + I$ we achieve a new equivalent form

$$
\begin{bmatrix}
  z \\
  z_1 \\
  z_2 \\
  y
\end{bmatrix} = \begin{bmatrix}
  c \\
  0 \\
  0 \\
  b
\end{bmatrix} + \begin{bmatrix}
  Z & L & 0 & 0 \\
  J - I & 0 & 0 & 0 \\
  J - I & 0 & I & 0 \\
  I & Y & -I & I
\end{bmatrix} \begin{bmatrix}
  x \\
  |z| \\
  |z_1| \\
  |z_2|
\end{bmatrix}
$$