

Piecewise Linear Secant Approximation via Algorithmic Piecewise Differentiation

Andreas Griewank¹, Tom Streubel^{2,3}, Lutz Lehmann³, Manuel
Radons⁴, and Richard Hasenfelder³

¹School of Mathematical Sciences and Information Technology,
Ecuador

²Zuse Institute Berlin, Germany

³Humboldt University of Berlin, Germany

⁴Technical University of Berlin, Germany

September 18, 2018

Abstract

It is shown how piecewise differentiable functions $F : \mathbb{R}^n \mapsto \mathbb{R}^m$ that are defined by evaluation programs can be approximated locally by a piecewise linear model based on a pair of sample points \tilde{x} and \hat{x} . We show that the discrepancy between function and model at any point x is of the bilinear order $\mathcal{O}(\|x - \tilde{x}\| \|x - \hat{x}\|)$. As an application of the piecewise linearization procedure we devise a generalized Newton's method based on successive piecewise linearization and prove for it sufficient conditions for convergence and convergence rates equaling those of semismooth Newton. We conclude with the derivation of formulas for the numerically stable implementation of the aforedeveloped piecewise linearization methods.

Keywords Automatic differentiation, Stable piecewise linearization, Generalized Newton's method, Lipschitz continuity, Generalized Hermite interpolation, ADOL-C

MSC2010 65D25, 65K10, 49J52

1 Introduction

In this paper we refine and extend the theory of piecewise linearizations¹ of piecewise differentiable functions $F : \mathbb{R}^n \mapsto \mathbb{R}^m$ that was introduced in [Gri13].

¹Our notion of linearity includes nonhomogeneous functions, where the adjective *affine* or perhaps *polyhedral* would be more precise. However, such mathematical terminology might be less appealing to computational practitioners and to the best of our knowledge there are no good nouns corresponding to *linearity* and *linearization* for the adjectives affine and polyhedral.

Throughout, we assume that any such function is defined by an evaluation procedure consisting of a sequence of elemental functions $v_i = \varphi_i(v_j)_{j \prec i}$, where φ_i is either smooth or the absolute value function. The data dependence relation \prec generates a partial ordering, which yields a directed acyclic graph. In other words, we assume that we have a straight line program without any loops or jumps in the control flow.

In the above reference a piecewise linearization $\Delta F(\hat{x}; \Delta x)$ of F about the point \hat{x} is constructed such that, for arbitrary $x \in \mathbb{R}^n$ with $\hat{F} = F(\hat{x})$, it holds

$$F(x) = \hat{F} + \Delta F(\hat{x}; x - \hat{x}) + \mathcal{O}(\|x - \hat{x}\|^2).$$

Since $\Delta F(\hat{x}; \Delta x)$ is constructed by replacing smooth elementals by their tangent lines at \hat{x} , we will refer to it as **piecewise tangent linearization**. An important property of these piecewise linearizations is that they vary continuously with respect to the sample point or points at which they are developed.

As a generalization of $\Delta F(\hat{x}; \Delta x)$ we construct a **piecewise secant linearization** $\Delta F(\check{x}, \hat{x}; \Delta x)$ of F at the pair \check{x} and \hat{x} such that for the midpoints $\hat{x} = (\check{x} + \hat{x})/2$ and $\hat{F} = (\check{F} + \hat{F})/2$ with $\check{F} = F(\check{x})$ and $\hat{F} = F(\hat{x})$ it holds

$$F(x) = \hat{F} + \Delta F(\check{x}, \hat{x}; x - \hat{x}) + \mathcal{O}(\|x - \check{x}\| \|x - \hat{x}\|).$$

The new $\Delta F(\check{x}, \hat{x}; \Delta x)$ reduces to $\Delta F(\hat{x}; \Delta x)$ when the two sample points \check{x} and \hat{x} coalesce at \hat{x} .

Our results in this article are twofold:

- Firstly, we present approximation properties and Lipschitz continuity estimates. Here the two major points are that the piecewise linearization is a second order approximation to the underlying function. Moreover, we do not only prove the existence of Lipschitz constants, but provide explicit estimates. These results immediately yield additional statement on conditions for perturbation-stable surjectivity of a piecewise linear model.
- Secondly, we develop an application for the piecewise linearization. Namely, we introduce, for each linearization mode, a Newton's method based on successive piecewise linearization and give sufficient conditions for convergence, as well as statements on the convergence rates.

In an appended section we moreover provide formulas for singularity free and thus numerically stable implementations of the piecewise secant linearizations.

One significant advantage of the piecewise linearization-approach is that, once the piecewise linearization is generated, we also know where its kinks are located. This means that we are liberated of the complications of event handling that the nondifferentiabilities of piecewise smooth functions usually bring about, which was one of the motivations for the development of the techniques presented in [Gri13]. In said reference it is proved that the variations of the piecewise linear model are Lipschitz continuous with respect to perturbations of the development point. The results mentioned in the first bullet point represent a significant improvement over those given in [Gri13] in so far as they are not only sharper, but also provide explicit bounds for the Lipschitz constants for the variations of the model with respect to perturbations of the base point.

Content and Structure

To provide some more background, we will, in the second section, elaborate on the setting outlined above both from the mathematical and the implementation perspective. In Section 2 we will introduce the piecewise linearization framework that is investigated thereafter. In Section 3 we derive the approximation and stability properties of said framework. In the subsequent section the generalized Newton's methods are derived for both linearization modes. In Section 5 we show that the secant linearization can be computed in a division free, centered form such that it reduces continuously to the tangent mode when the reference points \tilde{x} and \hat{x} coalesce. The latter is followed by a numerical example in Section 6. The article is concluded by some final remarks.

Preliminary Remarks

Let $\tilde{\Phi}$ be a library of elemental functions that conforms to the condition of elemental differentiability as described in [GW08]. In the following we will consider piecewise differentiable functions that are defined by an evaluation procedure consisting of a sequence of such elemental functions $v_i = \varphi_i(v_j)_{j < i}$, where the φ_i are contained in a library

$$\Phi := \tilde{\Phi} \cup \{\mathbf{abs}()\} .$$

We remark that the inclusion of the absolute value function into the library means that we can also evaluate $\min()$ and $\max()$ via the identities

$$\max(u, w) = (u+w+\mathbf{abs}(u-w))/2 , \quad \min(u, w) = (u+w-\mathbf{abs}(u-w))/2 . \quad (1)$$

There is a slight implicit restriction, namely, we assume that whenever \min or \max are evaluated both their arguments have well defined finite values so that the same is true for their sum and difference. On the other hand, the expression $\min(1, 1/\mathbf{abs}(u))$ makes perfect sense in IEEE arithmetic [iee85], but rewriting it as above leads to a *NaN* at $u = 0$. While this restriction may appear quite technical, it imposes the requirement that all relevant quantities are well defined at least in some open neighborhood, which is exactly in the nature of piecewise differentiability. For an in-depth investigation of piecewise differentiable functions, see, e.g., the books by Kummer [Kum88] and Scholtes [Sch12].

The function $\Delta F(\hat{x}; \Delta x)$ is incremental in that $\Delta F(\hat{x}; 0) = 0$, but like general piecewise linear continuous functions, it is only locally and positively homogeneous so that

$$\Delta F(\hat{x}; \alpha \Delta x) = \alpha \Delta F(\hat{x}; \Delta x) \quad \text{for } 0 < \alpha \|\Delta x\| < \rho(\hat{x}) .$$

Here the bound $\rho(\hat{x})$ is positive everywhere, but generally not continuous with respect to \hat{x} .

In [Gri13] it has been shown that the Jacobians of the linear pieces of $\Delta F(\hat{x}; \alpha \Delta x)$ in the ball of radius $\rho(\hat{x})$ about \hat{x} are conically active generalized Jacobians of the underlying nonlinear function $F(x)$. We will not elaborate

on this connection here, because even in the smooth case secant approximations need not correspond to exact Jacobians. In contrast, the generalized derivative sets in the sense of Clarke [Cla83] are for piecewise differentiable functions almost everywhere just singletons, containing the classical Jacobian matrix. In floating point arithmetic the user or client algorithm will then quite likely never 'see' the nonsmoothness or gain any useful information about it.

2 Piecewise Linearization by Tangents and Secants

Suppose the vector function $F : \mathcal{D} \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ in question is evaluated by a sequence of assignments

$$v_i = v_j \circ v_k \quad \text{or} \quad v_i = \varphi_i(v_j) \quad \text{for} \quad i = 1 \dots l.$$

Here $\circ \in \{+, -, *, /\}$ is a polynomial arithmetic operation and

$$\varphi_i \in \Phi \equiv \{\text{rec, sqrt, sin, cos, exp, log, \dots, abs, \dots}\}$$

a univariate function. The user or reader may extend the library by other locally Lipschitz-continuously differentiable functions like the analysis favorites

$$\varphi(u) \equiv |u| > 0 ? u^p \cdot \sin(1/u) : 0 \quad \text{for} \quad p \geq 3.$$

To fit into the the framework they then also have to supply an evaluation procedure for both the elemental function φ and its derivative φ' , which cannot be based mechanically on the chain rule.

Following the notation from [GW08] we partition the sequence of scalar variables v_i into the vector triple

$$(x, z, y) = (v_{1-n}, \dots, v_{-1}, v_0, \dots, v_{l-m}, v_{l-m+1}, \dots, v_l) \in \mathbb{R}^{n+l}$$

such that $x \in \mathbb{R}^n$ is the vector of independent variables, $y \in \mathbb{R}^m$ the vector of dependent variables and $z \in \mathbb{R}^{l-m}$ the (internal) vector of intermediates.

Some of the elemental functions like the reciprocal $\text{rec}(u) \equiv 1/u$, the square root and the logarithm are not globally defined. As mentioned above, we will assume that the input variables x are restricted to an open domain $\mathcal{D} \subset \mathbb{R}^n$ such that all resulting intermediate values $v_i = v_i(x)$ are well defined.

Throughout we will assume that the evaluation procedure for F involves exactly $s \geq 0$ calls to **abs**(), including min and max rewritten or at least reinterpreted as discussed above. Starting from \hat{x} and an increment $\Delta x = x - \hat{x}$, we will now construct for each intermediate v_i an approximation

$$v_i(\hat{x} + \Delta x) - \hat{v}_i \approx \Delta v_i \equiv \Delta v_i(\Delta x).$$

Here the incremental function $\Delta v_i(\Delta x)$ is continuous and piecewise linear, with \hat{x} or \tilde{x} and \hat{x} considered constant in the tangent and secant case, respectively. Hence, we will often list Δx , but only rarely use \tilde{x} , \hat{x} and \hat{x} as arguments of the Δv_i in proofs.

Defining Relations for Tangent Approximation

We use the reference values $\hat{v}_i = v_i(\hat{x})$ and, assuming that all φ_i other than the absolute value function are differentiable within the domain of interest, we may use the tangent linearizations

$$\Delta v_i = \Delta v_j \pm \Delta v_k \quad \text{for } v_i = v_j \pm v_k, \quad (2)$$

$$\Delta v_i = \hat{v}_j * \Delta v_k + \Delta v_j * \hat{v}_k \quad \text{for } v_i = v_j * v_k, \quad (3)$$

$$\Delta v_i = \hat{c}_{ij} * \Delta v_j \quad \text{for } v_i = \varphi_i(v_j) \neq \mathbf{abs}(). \quad (4)$$

Here $\hat{c}_{ij} \equiv \varphi'_i(\hat{v}_j)$, which will be different for the secant linearization.

If no absolute value or other nonsmooth elemental occurs, the function $y = F(x)$ is, at the current point, differentiable and by the chain rule we have the relation

$$\Delta y = \Delta F(\hat{x}; \Delta x) \equiv F'(\hat{x})\Delta x,$$

where $F'(x) \in \mathbb{R}^{m \times n}$ is the Jacobian matrix. Thus we observe the obvious fact that smooth differentiation is equivalent to linearizing all elemental functions.

Now, let us move to the piecewise differentiable scenario, where the absolute value function does occur $s > 0$ times. We may then obtain a piecewise linearization of the vector function $F(\hat{x} + \Delta x) - F(\hat{x})$ by incrementing

$$\Delta v_i = \mathbf{abs}(\hat{v}_j + \Delta v_j) - \hat{v}_i \quad \text{when } v_i = \mathbf{abs}(v_j). \quad (5)$$

Here $\hat{v}_i = \mathbf{abs}(\hat{v}_j)$, which will be slightly different for the secant linearization. In other words, we keep the piecewise linear function $\mathbf{abs}()$ unchanged so that the resulting Δy represents, for each fixed $x \in \mathcal{D}$, the piecewise linear and continuous *increment function*

$$\Delta y = \Delta y(\Delta x) = \Delta F(\hat{x}; \Delta x) : \mathbb{R}^n \rightarrow \mathbb{R}^m.$$

Defining Relations for Secant Approximation

In the tangent approximation the reference point was always the evaluation point \hat{x} and the resulting values $\hat{v}_i = v_i(\hat{x})$. Now we will make reference to the midpoints

$$\hat{v}_i \equiv (\check{v}_i + \hat{v}_i)/2 \quad \text{of } \check{v}_i \equiv v_i(\check{x}) \quad \text{and } \hat{v}_i \equiv v_i(\hat{x}). \quad (6)$$

Consequently, we have the functional dependence $\hat{v}_i = \hat{v}_i(\check{x}, \hat{x})$, which is at least Lipschitz continuous under our assumptions. Now an intriguing observation is that the recurrences (2) and (3) for arithmetic operations can stay just the same, and the recurrence (4) for nonlinear univariates is still formally valid, except that the tangent slope $\varphi'(\hat{v}_j)$ must be replaced by the secant slope

$$c_{ij} \equiv \begin{cases} (\check{v}_i - \hat{v}_i)/(\check{v}_j - \hat{v}_j) & \text{if } \check{v}_j \neq \hat{v}_j \\ \varphi'_i(\hat{v}_j) & \text{otherwise} \end{cases}. \quad (7)$$

Theoretically, some \check{v}_i and \hat{v}_i may coincide, even if the underlying sample points \check{x} and \hat{x} are not selected identically, in which case the secant based model would reduce to the tangent based model. While exact coincidence of any pair \check{v}_i and \hat{v}_i is rather unlikely, taking the divided difference over small increments is likely to generate numerical cancellation. Therefore we will develop a division free centered form in Section 5. Finally, the nonsmooth rule (5) can stay unchanged except that we now set

$$\check{v}_i \equiv \frac{1}{2}(\check{v}_i + \hat{v}_i) = \frac{1}{2}[\mathbf{abs}(\check{v}_j) + \mathbf{abs}(\hat{v}_j)]. \quad (8)$$

Hence, it is immediately clear that the new secant approximation reduces to the old tangent approximation when $\check{x} = \hat{x}$. In general, we will denote the mapping between the input increments $\Delta x \in \mathbb{R}^n$ and the resulting values $\Delta y \in \mathbb{R}^m$ by

$$\Delta y = \Delta y(\Delta x) = \Delta F(\check{x}, \hat{x}; \Delta x) : \mathbb{R}^n \rightarrow \mathbb{R}^m .$$

Its piecewise linear structure is very much the same as that of the tangent based model, which is described in detail in [Gri13]. Here we emphasize its quality in approximating the underlying nonlinear and nonsmooth F .

In contrast to the tangent model, the secant model is not a priori unique in that it depends quite strongly on the procedural representation of the vector function F and not just its values, i.e. its properties as a mathematical map. For example, one can easily check that applying the above secant modeling rules to $f(x) = \log(\exp(x))$ does not yield the same approximation $\Delta f(\check{x}, \hat{x}, \Delta x)$ as the one for $f(x) = x$. On the other hand the natural secant linearization rule for the product $v = u \cdot w$ is equivalent to that obtained by applying the Apollonius identity

$$u \cdot w = \frac{1}{4} [(u + w)^2 - (u - w)^2] .$$

Of course, the same is true for the tangent linearization and we may assume without loss of generality that we only have three kinds of elemental functions, the addition, the modulus and smooth univariate functions. That reduction greatly simplifies the theoretical analysis but might not be numerically optimal for actual implementations.

3 Approximation and Stability Properties

In contrast to the presentation in our previous papers we will now also use the nonincremental forms

$$\diamond_{\hat{x}} F(x) \equiv F(\hat{x}) + \Delta F(\hat{x}; x - \hat{x})$$

and

$$\diamond_{\check{x}} F(x) \equiv \frac{1}{2}(F(\check{x}) + F(\hat{x})) + \Delta F(\check{x}, \hat{x}; x - \hat{x}) .$$

For the square as a univariate nonlinear function $v(x) = x^2$ we find:

$$\begin{aligned} \diamond_{\hat{x}} v &= \diamond_{\hat{x}} x^2 = \frac{1}{2}[\check{x}^2 + \hat{x}^2] + \frac{\hat{x}^2 - \check{x}^2}{\hat{x} - \check{x}}(x - \frac{1}{2}[\hat{x} + \check{x}]) = \frac{1}{2}[\check{x}^2 + \hat{x}^2] + (\hat{x} + \check{x})(x - \hat{x}) \\ &= \frac{1}{2}[\check{x}^2 + \hat{x}^2] + 2\hat{x}(x - \hat{x}) = \check{v} + 2\hat{x}(x - \hat{x}). \end{aligned} \quad (9)$$

Lemma 3.1. *Plugging the secant approximation for the square into the Apollonius identity, we obtain for the general multiplication $v = u \cdot w$:*

$$\begin{aligned}\diamond_{(\tilde{u}, \tilde{w})}^{(\hat{u}, \hat{w})}(u \cdot w) &= \frac{1}{4} \left[\diamond_{(\tilde{u}, \tilde{w})}^{(\hat{u}, \hat{w})}(u + w)^2 - \diamond_{(\tilde{u}, \tilde{w})}^{(\hat{u}, \hat{w})}(u - w)^2 \right] \\ &= \frac{1}{2} (\hat{u}\hat{w} + \check{u}\check{w}) + \check{w}(u - \hat{u}) + \hat{u}(w - \check{w}).\end{aligned}$$

Proof.

$$\begin{aligned}4 \cdot \diamond_{(\tilde{u}, \tilde{w})}^{(\hat{u}, \hat{w})}(u \cdot w) &= \left[\diamond_{(\tilde{u}, \tilde{w})}^{(\hat{u}, \hat{w})}(u + w)^2 - \diamond_{(\tilde{u}, \tilde{w})}^{(\hat{u}, \hat{w})}(u - w)^2 \right] \\ &= \frac{1}{2} ([\hat{u} + \hat{w}]^2 + [\check{u} + \check{w}]^2) + 2(\hat{u} + \hat{w})(u + w - [\hat{u} + \hat{w}]) \\ &\quad - \frac{1}{2} ([\hat{u} - \hat{w}]^2 + [\check{u} - \check{w}]^2) + 2(\hat{u} - \hat{w})(u - w - [\hat{u} - \hat{w}]) \\ &= \frac{1}{2} ([\hat{u} + \hat{w}]^2 + [\check{u} + \check{w}]^2) + 2(\hat{u} + \hat{w})([u - \hat{u}] + [w - \hat{w}]) \\ &\quad - \frac{1}{2} ([\hat{u} - \hat{w}]^2 + [\check{u} - \check{w}]^2) + 2(\hat{u} - \hat{w})([u - \hat{u}] - [w - \hat{w}]) \\ &= 4 \cdot \left[\frac{1}{2} (\check{u}\check{w} + \hat{u}\hat{w}) + \check{w}[u - \hat{u}] + \hat{u}[w - \hat{w}] \right].\end{aligned}$$

□

Hereafter we will denote by $\|\cdot\| \equiv \|\cdot\|_\infty$ the infinity norm. Due to the norm equivalence in finite dimensional spaces all inequalities to be derived take the same form in other norms, provided the constants are adjusted accordingly. The infinity norm is particularly convenient, since we can then prove the following result for the general vector case $m > 1$ by considering the absolute values of the individual components $f = F_i$ for $i = 1 \dots m$. Moreover, we will make use of the Apollonius identity in the following proposition:

Proposition 3.2. *Suppose $\tilde{x}, \tilde{x}, \hat{x}, \tilde{y}, \hat{y}, \tilde{z}, \hat{z} \in \mathbb{R}^n$ are restricted to a sufficiently small closed convex neighborhood $K \subset \mathbb{R}^n$ where the evaluation procedure for $F : \mathbb{R}^n \mapsto \mathbb{R}^m$ is well defined. Then there are Lipschitz constants β_F and γ_F such that we have*

(i) *Lipschitz continuity of function, tangent and secant models:*

$$\begin{aligned}\|F(x) - F(\tilde{x})\| &\leq \beta_F \|x - \tilde{x}\| \quad \text{for } x, \tilde{x} \in K, \\ \max(\|\diamond_{\tilde{x}}^{\hat{x}} F(x) - \diamond_{\tilde{x}}^{\hat{x}} F(\tilde{x})\|, \|\diamond_{\tilde{x}} F(x) - \diamond_{\tilde{x}} F(\tilde{x})\|) &\leq \beta_F \|x - \tilde{x}\| \quad \text{for } x, \tilde{x} \in \mathbb{R}^n.\end{aligned}$$

The constant β_F can be defined by the recurrences $\beta_v = \beta_u + \beta_w$ if $v = u + w$, $\beta_v = \beta_u$ if $v = |u|$ and

$$\beta_v = \beta_u L_K(\varphi) \quad \text{if } v = \varphi(u) \quad \text{with } L_K(\varphi) \equiv \max_{x \in K} |\varphi'(u(x))|.$$

(ii) *Error between function and secant or tangent model:*

$$\begin{aligned}\|F(x) - \diamond_{\tilde{x}}^{\hat{x}} F(x)\| &\leq \frac{1}{2} \gamma_F \|x - \tilde{x}\| \|x - \hat{x}\| \\ \|F(x) - \diamond_{\tilde{x}} F(x)\| &\leq \frac{1}{2} \gamma_F \|x - \hat{x}\|^2,\end{aligned}$$

where $x \in K$. The constant γ_F can be defined using the recurrences $\gamma_v = \gamma_u + \gamma_w$ if $v = u + w$, $\gamma_v = \gamma_u$ if $v = |u|$ and

$$\gamma_v = L_K(\varphi)\gamma_u + L_K(\varphi')\beta_u^2 \quad \text{if } v = \varphi(u) \quad \text{with } L_K(\varphi') \equiv \max_{x \in K} |\varphi'(u(x))| .$$

(iii) *Lipschitz continuity of secant and tangent model:*

$$\begin{aligned} \|\diamond_{\hat{z}}^{\hat{z}} F(x) - \diamond_{\hat{y}}^{\hat{y}} F(x)\| &\leq \gamma_F \max[\|\hat{z} - \hat{y}\| \max(\|x - \hat{y}\|, \|x - \hat{z}\|), \\ &\quad \|\hat{z} - \hat{y}\| \max(\|x - \hat{y}\|, \|x - \hat{z}\|)] \\ \|\diamond_{\hat{z}} F(x) - \diamond_{\hat{y}} F(x)\| &\leq \gamma_F \|\hat{z} - \hat{y}\| \max(\|x - \hat{y}\|, \|x - \hat{z}\|) , \end{aligned}$$

where $x \in \mathbb{R}^n$.

(iv) *Lipschitz continuity of the incremental part: Let $x \in K$. Abbreviating $\Delta y = \hat{y} - \check{y}$ and $\Delta z = \hat{z} - \check{z}$ we obtain in the secant case*

$$\begin{aligned} &\|\Delta_{\hat{z}}^{\hat{z}} F(\Delta x) - \Delta_{\hat{y}}^{\hat{y}} F(\Delta x)\| \\ &\leq 2\beta_F \|\hat{z} - \hat{y}\| + \frac{1}{2}\gamma_F (\|\hat{z} - \hat{y}\| + \max(\|\Delta y\|, \|\Delta z\|))^2 \\ &\quad + \gamma_F (\|\hat{z} - \hat{y}\| + \frac{1}{2}(\|\Delta y\| + \|\Delta z\|)) \|\Delta x\| , \end{aligned}$$

which reduces in the tangent case to

$$\|\Delta_{\hat{z}} F(\Delta x) - \Delta_{\hat{y}} F(\Delta x)\| \leq 2\beta_F \|\hat{z} - \hat{y}\| + \frac{1}{2}\gamma_F \|\hat{z} - \hat{y}\|^2 + \gamma_F \|\hat{z} - \hat{y}\| \|\Delta x\| .$$

Here the second bounds applying to the tangent model are always specializations of the previous ones for the secant model.

Proof. Since otherwise the bounds can be applied componentwise we may assume without loss of generality that F is a scalar function f and the norm in the range is simply the absolute value $|\cdot|$. The proof proceeds by induction on the intermediate quantities v in the computational graph of f . We will define the constants β_v and γ_v recursively on the basis of the Lipschitz constants of the elemental functions and their derivatives:

Variable initialization: The initialization of independant variables represent the minimal nodes and all assertions are trivially true with the constants $\beta_{x_i} = 1$ and $\gamma_{x_i} = 0$.

Smooth univariate operation: Let $v = \varphi(u)$, $\varphi \in \tilde{\Phi}$, be some elemental function in the computational graph of f :

$$\begin{aligned} \|\varphi(u) - \varphi(\tilde{u})\| &\leq L_K(\varphi)\|u - \tilde{u}\| \leq \beta_u L_K(\varphi)\|x - \tilde{x}\| \\ \text{and } \|\diamond_{\hat{u}}^{\hat{u}} \varphi(u) - \diamond_{\hat{u}}^{\hat{u}} \varphi(\tilde{u})\| &\leq L_K(\diamond_{\hat{u}}^{\hat{u}} \varphi)\|u - \tilde{u}\| \leq \beta_u L_K(\diamond_{\hat{u}}^{\hat{u}} \varphi)\|x - \tilde{x}\| . \end{aligned}$$

The first inequality holds due to the Lipschitz continuity of φ and $\diamond_{\hat{u}}^{\hat{u}} \varphi$. The latter inequality is the induction hypothesis and since $L_K(\diamond_{\hat{u}}^{\hat{u}} \varphi) \leq L_K(\varphi)$ holds by the mean value theorem, we may set $\beta_v \equiv \beta_u L_K(\varphi)$.

Absolute value function and sum: The absolute value function $v = \mathbf{abs}(u)$ naturally maintains the Lipschitz constant and the addition summates them.

Thus we have established (i) for all cases.

Now let us consider the approximation property (ii). For additions $v = u + w$ we may set $\gamma_v = \gamma_u + \gamma_w$ and then have by the triangle inequality

$$\begin{aligned} |v(x) - \diamond_{\hat{x}}^{\hat{x}} v(x)| &\leq |u(x) - \diamond_{\hat{x}}^{\hat{x}} u(x)| + |w(x) - \diamond_{\hat{x}}^{\hat{x}} w(x)| \\ &\leq \frac{1}{2}\gamma_u(\|x - \check{x}\|\|x - \hat{x}\|) + \frac{1}{2}\gamma_w(\|x - \check{x}\|\|x - \hat{x}\|) = \frac{1}{2}\gamma_v(\|x - \check{x}\|\|x - \hat{x}\|). \end{aligned}$$

For the absolute value function $v = \mathbf{abs}(u)$ we may also set $\gamma_v = \gamma_u$, since

$$|v(x) - \diamond_{\hat{x}}^{\hat{x}} v(x)| = ||u| - |\diamond_{\hat{x}}^{\hat{x}} u(x)|| \leq |u - \diamond_{\hat{x}}^{\hat{x}} u(x)| \leq \frac{1}{2}\gamma_u(\|x - \check{x}\|\|x - \hat{x}\|).$$

For the univariate functions $v = \varphi(u)$ we have with $\tilde{u} \equiv \diamond_{\hat{x}}^{\hat{x}} u(x)$

$$|v(x) - \diamond_{\hat{x}}^{\hat{x}} v(x)| \leq |\varphi(u(x)) - \varphi(\tilde{u})| + |\varphi(\tilde{u}) - \diamond_{\hat{u}}^{\hat{u}} \varphi(\tilde{u})|.$$

By the mean value theorem and the induction hypothesis, the first term is bounded by

$$L_K(\varphi)|u(x) - \diamond_{\hat{x}}^{\hat{x}} u(x)| \leq \frac{1}{2}L_K(\varphi)\gamma_u(\|x - \check{x}\|\|x - \hat{x}\|).$$

The second term represents the error in the Hermite interpolation of φ between \tilde{u} and \hat{u} . With $L_K(\varphi')$ a Lipschitz constant of φ' on $u(K)$ it is bounded by

$$\frac{1}{2}L_K(\varphi')|\tilde{u} - \hat{u}|\|\tilde{u} - \hat{u}\| \leq \frac{1}{2}L_K(\varphi')\beta_u^2(\|x - \check{x}\|\|x - \hat{x}\|),$$

where the last bound follows from the fact that according to (i) the approximation $\diamond_{\hat{x}}^{\hat{x}} u(x)$ has the Lipschitz constant β_u and takes on at \check{x} and \hat{x} the values \tilde{u} and \hat{u} . Hence, we have shown that (ii) holds indeed with

$$\gamma_v = L_K(\varphi)\gamma_u + L_K(\varphi')\beta_u^2. \quad (10)$$

Next we want to prove Lipschitz continuity of the model as stated in (iii). Again, we find for additions and the abs function that the assertion is almost trivial with the constants γ either being summated or just passed on. The challenge is once more the induction through the nonlinear univariates $v = \varphi(u)$. To limit the notational complexity we will connect the two point pairs at the u level by straight lines setting

$$\begin{aligned} \tilde{u} &\equiv \tilde{u}(t) \equiv u(\hat{y})(1-t) + tu(\check{z}) &\Rightarrow & \partial\tilde{u}(t)/\partial t = \Delta\tilde{u} \equiv u(\check{z}) - u(\hat{y}), \\ \hat{u} &\equiv \hat{u}(t) \equiv u(\hat{y})(1-t) + tu(\hat{z}) &\Rightarrow & \partial\hat{u}(t)/\partial t = \Delta\hat{u} \equiv u(\hat{z}) - u(\hat{y}), \\ \tilde{\tilde{u}} &\equiv \tilde{\tilde{u}}(t) \equiv \diamond_{\hat{y}}^{\hat{y}} u(x)(1-t) + t\diamond_{\hat{z}}^{\hat{z}} u(x) &\Rightarrow & \partial\tilde{\tilde{u}}/\partial t = \Delta\tilde{\tilde{u}} \equiv \diamond_{\hat{z}}^{\hat{z}} u(x) - \diamond_{\hat{y}}^{\hat{y}} u(x). \end{aligned}$$

Here x is fixed and we assume as induction hypothesis that

$$\|\Delta\tilde{\tilde{u}}\| \leq \gamma_u \max[\|\hat{z} - \hat{y}\| \max(\|x - \check{y}\|, \|x - \check{z}\|), \|\check{z} - \check{y}\| \max(\|x - \hat{y}\|, \|x - \hat{z}\|)]. \quad (11)$$

To connect the piecewise linearizations of v define

$$\tilde{v}(t) \equiv \frac{1}{2} [\varphi(\tilde{u}(t)) + \varphi(\hat{u}(t))] + \frac{\varphi(\tilde{u}(t)) - \varphi(\hat{u}(t))}{\tilde{u}(t) - \hat{u}(t)} \left[\tilde{u}(t) - \frac{1}{2} (\tilde{u}(t) + \hat{u}(t)) \right]$$

The quantity we want to find a bound for is $\Delta\tilde{v} \equiv \tilde{v}(1) - \tilde{v}(0) = \diamond_{\tilde{z}}^{\tilde{y}} v(x) - \diamond_{\hat{y}}^{\hat{y}} v(x)$. By the mean value theorem we find some $\bar{t} \in [0, 1]$ where

$$\begin{aligned} \Delta\tilde{v} &= \tilde{v}(1) - \tilde{v}(0) = \left. \frac{\partial \tilde{v}(t)}{\partial t} \right|_{t=\bar{t}} \\ &= \frac{1}{2} [\varphi'(\tilde{u})\Delta\tilde{u} + \varphi'(\hat{u})\Delta\hat{u}] + \frac{\varphi(\tilde{u}) - \varphi(\hat{u})}{\tilde{u} - \hat{u}} [\Delta\tilde{u} - \frac{1}{2} (\Delta\tilde{u} + \Delta\hat{u})] \\ &\quad + \left\{ \frac{\varphi'(\tilde{u})\Delta\tilde{u} - \varphi'(\hat{u})\Delta\hat{u}}{\tilde{u} - \hat{u}} - \frac{(\varphi(\tilde{u}) - \varphi(\hat{u}))(\Delta\tilde{u} - \Delta\hat{u})}{(\tilde{u} - \hat{u})^2} \right\} \left[\tilde{u} - \frac{1}{2} (\tilde{u} + \hat{u}) \right]. \end{aligned}$$

The functions $\tilde{u}, \hat{u}, \bar{u}$ here and following are to be read as evaluated at $t = \bar{t}$. Now introduce \bar{u} as the mean value of \tilde{u} and \hat{u} at which the difference quotient of φ over the intervening interval is equal to its derivative. This yields:

$$\begin{aligned} \Delta\tilde{v} &= \frac{1}{2} [\varphi'(\tilde{u})\Delta\tilde{u} + \varphi'(\hat{u})\Delta\hat{u}] + \varphi'(\bar{u}) \left[\Delta\tilde{u} - \frac{1}{2} (\Delta\tilde{u} + \Delta\hat{u}) \right] \\ &\quad + \left\{ \frac{\varphi'(\tilde{u})\Delta\tilde{u} + \varphi'(\bar{u})(\Delta\hat{u} - \Delta\tilde{u}) - \varphi'(\hat{u})\Delta\hat{u}}{\tilde{u} - \hat{u}} \right\} \left[\tilde{u} - \frac{1}{2} (\tilde{u} + \hat{u}) \right] \\ &= \frac{1}{2} \left\{ \frac{(\varphi'(\tilde{u}) - \varphi'(\bar{u}))\Delta\tilde{u} + (\varphi'(\hat{u}) - \varphi'(\bar{u}))\Delta\hat{u}}{\tilde{u} - \hat{u}} \right\} (\tilde{u} - \hat{u}) + \varphi'(\bar{u})\Delta\tilde{u} \\ &\quad + \frac{1}{2} \left\{ \frac{(\varphi'(\tilde{u}) - \varphi'(\bar{u}))\Delta\tilde{u} - (\varphi'(\hat{u}) - \varphi'(\bar{u}))\Delta\hat{u}}{\tilde{u} - \hat{u}} \right\} [(\tilde{u} - \bar{u}) + (\bar{u} - \hat{u})] \\ &= \frac{\varphi'(\bar{u}) - \varphi'(\hat{u})}{\tilde{u} - \hat{u}} \Delta\hat{u}(\tilde{u} - \bar{u}) - \frac{\varphi'(\bar{u}) - \varphi'(\tilde{u})}{\tilde{u} - \bar{u}} \Delta\tilde{u}(\bar{u} - \hat{u}) + \varphi'(\bar{u})\Delta\tilde{u}. \end{aligned}$$

The two quotients are bounded according to

$$\begin{aligned} \left| \frac{\varphi'(\bar{u}) - \varphi'(\tilde{u})}{\tilde{u} - \hat{u}} \right| + \left| \frac{\varphi'(\bar{u}) - \varphi'(\hat{u})}{\tilde{u} - \hat{u}} \right| &\leq \left| \frac{\varphi'(\tilde{u}) - \varphi'(\bar{u})}{\tilde{u} - \bar{u}} \right| \left| \frac{\bar{u} - \tilde{u}}{\tilde{u} - \hat{u}} \right| + \left| \frac{\varphi'(\hat{u}) - \varphi'(\bar{u})}{\hat{u} - \bar{u}} \right| \left| \frac{\bar{u} - \hat{u}}{\tilde{u} - \hat{u}} \right| \\ &\leq L_K(\varphi') \left(\left| \frac{\bar{u} - \tilde{u}}{\tilde{u} - \hat{u}} \right| + \left| \frac{\bar{u} - \hat{u}}{\tilde{u} - \hat{u}} \right| \right) = L_K(\varphi'), \end{aligned}$$

where the last equality follows from \bar{u} being between \tilde{u} and \hat{u} . Hence, we find that

$$|\Delta\tilde{v}| \leq L_K(\varphi') \max(|\Delta\tilde{u}||\tilde{u} - \hat{u}| + |\Delta\hat{u}||\bar{u} - \tilde{u}|) + L_K(\varphi)|\Delta\tilde{u}|. \quad (12)$$

The factors in the middle are easily bounded by

$$|\Delta\tilde{u}| = |u(\tilde{z}) - u(\tilde{y})| \leq \beta_u \|\tilde{z} - \tilde{y}\| \quad \text{and} \quad |\Delta\hat{u}| = |u(\hat{z}) - u(\hat{y})| \leq \beta_u \|\hat{z} - \hat{y}\|.$$

That leaves us with the second factors, which are linear in t such that

$$\begin{aligned}
|\tilde{u}(t) - \check{u}(t)| &\leq \max(|\tilde{u}(0) - \check{u}(0)|, |\tilde{u}(1) - \check{u}(1)|) \\
&= \max(|\diamond_{\hat{y}}^{\hat{y}} u(x) - u(\hat{y})|, |\diamond_{\hat{z}}^{\hat{z}} u(x) - u(\hat{z})|) \\
&= \max(|\diamond_{\hat{y}}^{\hat{y}} u(x) - \diamond_{\hat{y}}^{\hat{y}} u(\hat{y})|, |\diamond_{\hat{z}}^{\hat{z}} u(x) - \diamond_{\hat{z}}^{\hat{z}} u(\hat{z})|) \\
&\leq \beta_u \max(\|x - \hat{y}\|, \|x - \hat{z}\|),
\end{aligned}$$

where the last inequality follows from (i). Analogously, we can derive

$$|\tilde{u}(t) - \hat{u}(t)| \leq \beta_u \max(\|x - \hat{y}\|, \|x - \hat{z}\|).$$

Substituting this into (12) we get

$$|\Delta \tilde{v}| \leq \gamma_v \max(\|\hat{z} - \hat{y}\| \|x - \hat{y}\|, \|\hat{z} - \hat{y}\| \|x - \hat{z}\|, \|\hat{z} - \hat{y}\| \|x - \hat{y}\|, \|\hat{z} - \hat{y}\| \|x - \hat{z}\|),$$

with $\gamma_v \equiv L_K(\varphi)\gamma_u + L_K(\varphi')\beta_u^2$, which completes the proof of (iii). Finally, we have to prove (iv), which gives a bound on the increment part only. Setting $\xi \equiv \frac{1}{2}(\hat{z} + \hat{y}) + \Delta x$ one gets, with the results already proved and a few triangle inequalities, that

$$\begin{aligned}
&\|\Delta_{\hat{z}}^{\hat{z}} F(\Delta x) - \Delta_{\hat{y}}^{\hat{y}} F(\Delta x)\| \\
&= \|\diamond_{\hat{z}}^{\hat{z}} F(\hat{z} + \Delta x) - F(\hat{z}) - (\diamond_{\hat{y}}^{\hat{y}} F(\hat{y} + \Delta x) - F(\hat{y}))\| \\
&= \|\diamond_{\hat{z}}^{\hat{z}} F(\xi) - F(\hat{z}) + \diamond_{\hat{z}}^{\hat{z}} F(\hat{z} + \Delta x) - \diamond_{\hat{z}}^{\hat{z}} F(\xi) \\
&\quad - \diamond_{\hat{y}}^{\hat{y}} F(\xi) + F(\hat{y}) - \diamond_{\hat{y}}^{\hat{y}} F(\hat{y} + \Delta x) + \diamond_{\hat{y}}^{\hat{y}} F(\xi)\| \\
&\leq \beta_F \|\hat{z} - \hat{y}\| + \beta_F \|\hat{z} + \Delta x - \frac{1}{2}(\hat{z} + \hat{y}) - \Delta x\| + \beta_F \|\hat{y} + \Delta x - \frac{1}{2}(\hat{z} + \hat{y}) - \Delta x\| \\
&\quad + \gamma_F \max[\|\hat{z} - \hat{y}\| \max(\|\xi - \hat{z}\|, \|\xi - \hat{y}\|), \|\hat{z} - \hat{y}\| \max(\|\xi - \hat{z}\|, \|\xi - \hat{y}\|)] \\
&= 2\beta_F \|\hat{z} - \hat{y}\| + \gamma_F \max[\|\hat{z} - \hat{y}\| \max(\|\xi - \hat{z}\|, \|\xi - \hat{y}\|), \|\hat{z} - \hat{y}\| \max(\|\xi - \hat{z}\|, \|\xi - \hat{y}\|)].
\end{aligned}$$

Now, since for example

$$\|\frac{1}{2}(\hat{y} + \hat{z}) - \hat{z}\| = \|\frac{1}{2}(\hat{y} - \hat{z}) - \hat{z} + \frac{1}{2}(\hat{z} + \hat{z})\| \leq \frac{1}{2}(\|\hat{y} - \hat{z}\| + \|\Delta z\|),$$

both inner maxima can be bounded by the same expression, namely

$$\|\Delta x\| + \frac{1}{2}(\|\hat{z} - \hat{y}\| + \max(\|\Delta z\|, \|\Delta y\|)),$$

so that we obtain the upper bound

$$\begin{aligned}
&\|\Delta_{\hat{z}}^{\hat{z}} F(\Delta x) - \Delta_{\hat{y}}^{\hat{y}} F(\Delta x)\| \\
&\leq 2\beta_F \|\hat{z} - \hat{y}\| + \gamma_F \max(\|\hat{z} - \hat{y}\|, \|\hat{z} - \hat{y}\|) \left[\|\Delta x\| + \frac{1}{2}(\|\hat{z} - \hat{y}\| + \max(\|\Delta z\|, \|\Delta y\|)) \right].
\end{aligned}$$

Finally, we can also bound

$$\|\hat{y} - \hat{z}\| = \|\hat{y} - \hat{y} - \hat{z} + \hat{z} + \hat{y} - \hat{z}\| \leq \|\hat{y} - \hat{z}\| + \frac{1}{2}(\|\Delta y\| + \|\Delta x\|),$$

which yields the assertion after some elementary modifications. From the secant result we can easily get the bound for the tangent model by setting $\hat{z} = \hat{z} = \hat{z}$ and $\hat{y} = \hat{y} = \hat{y}$. \square

As one can see by setting $\tilde{y} = x = \tilde{z}$ the assertion (ii) almost follows from (iii), except that a factor of 2 is lost in the constants. The proposition also states that the values of F at \tilde{x} and \hat{x} are reproduced exactly by our approximation as one would expect from a secant approximation. This property clearly nails down the piecewise linearization rules (5) and (4) with (7) for all univariate functions. Also, there is no doubt that addition and subtraction should be linearized according to (2) and that multiplications $v_i = c v_j$ by constants c should yield the differentiated version $\Delta v_i = c \Delta v_j$, which is a special case of (3). For general multiplications $v_i = v_j * v_k$ the two values $\tilde{v}_i = \tilde{v}_j * \tilde{v}_k$ and $\hat{v}_i = \hat{v}_j * \hat{v}_k$ could also be interpolated by linear functions other than the one defined by (3). However, we currently see no possible gain in that flexibility, and maintaining the usual product rule form seems rather attractive.

Stable Surjectivity

Recall that a continuous function $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is called proper if the preimage of every compact set is compact. One can easily see that any piecewise linear function is proper if and only if it maps no affine ray $\{a + \lambda b : \lambda \geq 0\}$, ($a, b \in \mathbb{R}^n$), to a point. This is trivially the case if $n = m$ and the Jacobians of all selection functions of F are invertible. In [Rad17] it was proved that if a piecewise linear function $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is proper, there exists a $d \in \mathbb{Z}$ such that for all regular values y of F , i.e. all values y such that the Jacobian at all preimages of y exists and is invertible [which is trivially the case if $F^{-1}(y) = \emptyset$], it holds

$$d = \sum_{x \in F^{-1}(y)} \text{sign}[\det(D_x F)] =: \text{deg}(F) .$$

Hence, any regular value y of a proper piecewise linear function $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ has at least one preimage if $\text{deg}(F) \neq 0$. This implies surjectivity of F by the closedness of piecewise linear functions (cf. [Sch12]) and the well known fact that regular values lie dense in the range. We call $\text{deg}(F)$ the degree of F .

Lemma 3.3. ([Rad17, Cor. 5.2]) *Assume a piecewise linear function $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is composed of k affine functions F_i with invertible linear parts A_i and define*

$$\rho_F := \min \left\{ \frac{1}{\|A_1^{-1}\|}, \dots, \frac{1}{\|A_k^{-1}\|} \right\} .$$

Moreover, let $0 < \varepsilon < \rho_F$ and $l \geq 0$. Then all piecewise linear functions $G : \mathbb{R}^n \rightarrow \mathbb{R}^n$ with

$$\|F(x) - G(x)\| \leq \varepsilon \|x\| + l \quad \forall x \in \mathbb{R}^n$$

are proper and their degree equals that of F .

The latter statement enables us to prove another stability result:

Proposition 3.4. *Let $F \in \text{span}(\Phi)$ where $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ and assume a piecewise linearization $\diamond_{\tilde{x}} F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is composed of k affine functions F_i with invertible*

linear parts A_i . Moreover, define ρ_F as in Lemma 3.3. Then for all tangent and secant mode piecewise linearizations with development points in the ball $B(\hat{x}, \rho_F/\gamma_F)$, where γ_F is defined as in Proposition 3.2, the mapping degree is well defined and equals that of $\diamond_{\hat{x}}F$.

Proof. Just plug Proposition 3.2. (iii) into Lemma 3.3. \square

The fact that the mapping degree of a coherently oriented piecewise linear function cannot be zero yields a rather pointed statement:

Corollary 3.5. *In the situation of Proposition 3.4, if $\diamond_{\hat{x}}F$ is coherently oriented, then all tangent and secant mode piecewise linearizations with development points in the ball $B(\hat{x}, \rho_F/\gamma_F)$ are surjective.*

4 Generalized Newton Methods by Piecewise Linearization

We will proceed by proposing and analyzing a possible application of the tangent and secant approximations developed in the previous sections. For this we present generalized Newton's methods for composite piecewise smooth functions $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ based on the piecewise linear approximations, both for the tangent and the secant mode. The merit of these methods is the fact that they impose no strong differentiability requirements but require only piecewise differentiability at the root in question.

Definition 1 (Newton operator). Let $F \in \text{span}(\Phi)$ and x^* be an isolated root of F in an open neighborhood \mathcal{D} . The Newton step for F is definable on \mathcal{D} in tangent or secant mode if the piecewise linear equation $\diamond_{\hat{x}}F(x) = 0$ resp. $\diamond_{\hat{x}}^{\hat{x}}F(x) = 0$ has at least one root for all $\hat{x} \in \mathcal{D}$ resp. $\hat{x}, \hat{x} \in \mathcal{D}$.

Then the Newton operator is defined in tangent mode as

$$N(\hat{x}) = \arg \min \{ \|x - \hat{x}\| : \diamond_{\hat{x}}F(x) = 0 \}$$

and in secant mode as

$$N(\hat{x}, \tilde{x}) = \arg \min \{ \|x - \frac{1}{2}(\hat{x} + \tilde{x})\| : \diamond_{\hat{x}}^{\hat{x}}F(x) = 0 \} .$$

Definition 1 is the minimal assumption under which we can conclude, using the approximation and Lipschitz continuity results of Section 3, that the iteration

$$x_{k+1} = N(x_k) \quad (\text{tangent}) \quad x_{k+1} = N(x_k, x_{k-1}) \quad (\text{secant}) \quad (13)$$

converges locally to an isolated root x^* . Assuming that the piecewise linearization $\diamond_{x^*}F$ is bijective on a ball $\mathcal{D} = B(x^*, \bar{R})$ we will proceed to show that close to x^*

- the Newton step can be defined

- the Newton step stays close to x^*
- the tangent mode Newton method converges quadratically and the secant mode Newton method converges with the golden mean as order.

We know that the assumption of bijectivity on a ball can be relaxed through the use of degree theory, but this requires a significant technical effort for which we refer to subsequent works. Note that the minimal assumption for the local convergence of semismooth Newton [Hin10] is similar, namely, that there always exists a generalized Jacobian $J(x)$ that has a uniformly bounded inverse over all $x \in B(x^*, \bar{\rho})$.

For most of the following results up to the last it is sufficient to consider the secant mode piecewise linear approximation of F , as results for the tangent mode can be obtained by setting $\tilde{x} = \hat{x} = \hat{x}$.

The general bound for the difference of piecewise linearizations with different basis points \tilde{z}, \hat{z} and $\tilde{y} = \hat{y} = \hat{y}$ can be refined to a closer bound in the case that one is to express that bound only in distances to \hat{y} .

Lemma 4.1. *Let the second order constant γ_F be valid on some convex set U and $x, \hat{y}, \tilde{z}, \hat{z} \in U$. Then*

$$\|\diamond_{\tilde{z}}^{\hat{z}} F(x) - \diamond_{\hat{y}} F(x)\| \leq \gamma_F \left[\max(\|\hat{z} - \hat{y}\|, \|\tilde{z} - \hat{y}\|) \cdot \|x - \hat{y}\| + \frac{1}{2} \|\hat{z} - \hat{y}\| \|\tilde{z} - \hat{y}\| \right].$$

Proof. Select some $N \in \mathbb{N}$ and consider the subdivision of the segments $[\tilde{z}, \hat{y}]$ and $[\hat{z}, \hat{y}]$ by $\tilde{u}_k = \hat{y} + \frac{k}{N}(\tilde{z} - \hat{y})$ and $\hat{u}_k = \hat{y} + \frac{k}{N}(\hat{z} - \hat{y})$. Then by Prop. 3.2.(iii)

$$\begin{aligned} \|\diamond_{\tilde{u}_{k+1}}^{\hat{u}_{k+1}} F(x) - \diamond_{\hat{u}_k} F(x)\| &\leq \gamma_F \max \begin{bmatrix} \|\hat{u}_{k+1} - \hat{u}_k\| \max(\|x - \tilde{u}_{k+1}\|, \|x - \tilde{u}_k\|) \\ \|\tilde{u}_{k+1} - \tilde{u}_k\| \max(\|x - \hat{u}_{k+1}\|, \|x - \hat{u}_k\|) \end{bmatrix} \\ &\leq \gamma_F \max \begin{bmatrix} \frac{1}{N} \|\hat{z} - \hat{y}\| (\|x - \hat{y}\| + \frac{k+1}{N} \|\tilde{z} - \hat{y}\|) \\ \frac{1}{N} \|\tilde{z} - \hat{y}\| (\|x - \hat{y}\| + \frac{k+1}{N} \|\hat{z} - \hat{y}\|) \end{bmatrix} \\ &= \gamma_F \left[\frac{1}{N} \max(\|\hat{z} - \hat{y}\|, \|\tilde{z} - \hat{y}\|) \|x - \hat{y}\| + \frac{k+1}{N^2} \|\hat{z} - \hat{y}\| \|\tilde{z} - \hat{y}\| \right]. \end{aligned}$$

Summation and limit $N \rightarrow \infty$ results in the claim. \square

We assume hereafter that $\diamond_{x^*} F$ is bijective on a ball $\mathcal{D} = B(x^*, \bar{R})$, where x^* is an isolated root of F , which implies that it is also coherently oriented and metrically regular on the latter. This means that there exists a constant $c > 0$ such that for all $x \in \mathcal{D}$, $y \in \diamond_{x^*} F(\mathcal{D})$

$$\left\| x - (\diamond_{x^*} F|_{\mathcal{D}})^{-1}(y) \right\| \leq c \|\diamond_{x^*} F(x) - y\|.$$

Lemma 4.2 (existence of roots of the PL approximation). *If $R \leq \min(\bar{R}, \frac{1}{c\gamma_F})$ and $\rho = \frac{R}{3}$ then for any $\tilde{x}, \hat{x} \in B(x^*, \rho)$ the piecewise linearization $\diamond_{\tilde{x}}^{\hat{x}} F$ has a root in $B(x^*, R)$.*

Proof. We intend to apply the Brouwer theorem to the fixed point operator

$$T(x) = x + x^* - (\diamond_{x^*} F|_{\mathcal{D}})^{-1} (\diamond_{\hat{x}} F(x)). \quad (14)$$

Any fixed point is then a root of $\diamond_{\hat{x}} F$. As T is obviously continuous, we only need to show that T maps $B(x^*, R)$ into itself. Using metric regularity and Lemma 4.1 we find

$$\begin{aligned} \|T(x) - x^*\| &= \left\| x - (\diamond_{x^*} F|_{\mathcal{D}})^{-1} (\diamond_{\hat{x}} F(x)) \right\| \\ &\leq c \left\| \diamond_{x^*} F(x) - \diamond_{\hat{x}} F(x) \right\| \\ &\leq c\gamma_F \left[\max(\|\tilde{x} - x^*\|, \|\hat{x} - x^*\|) \|x - x^*\| + \frac{1}{2} \|\tilde{x} - x^*\| \|\hat{x} - x^*\| \right] \\ &\leq \frac{\rho}{R} \left(R + \frac{1}{2}\rho \right) \leq \frac{7}{18}R. \end{aligned}$$

□

Lemma 4.3 (containment of the roots). *Under the same assumptions any root of $\diamond_{\hat{x}} F$ in $B(x^*, R)$ is actually contained in $B(x^*, \rho)$.*

Proof. Let x be a root of $\diamond_{\hat{x}} F$. Then again using metric regularity and Lemma 4.1 the distance to x^* has the bound

$$\begin{aligned} \|x - x^*\| &\leq c \|\diamond_{x^*} F(x)\| = c \|\diamond_{x^*} F(x) - \diamond_{\hat{x}} F(x)\| \\ &\leq c\gamma_F \left[\max(\|\tilde{x} - x^*\|, \|\hat{x} - x^*\|) \|x - x^*\| + \frac{1}{2} \|\tilde{x} - x^*\| \|\hat{x} - x^*\| \right] \\ &\leq \frac{1}{3} \|x - x^*\| + \frac{\|\tilde{x} - x^*\| \|\hat{x} - x^*\|}{6\rho} \end{aligned}$$

so that

$$\|x - x^*\| \leq \frac{\|\tilde{x} - x^*\| \|\hat{x} - x^*\|}{4\rho} \leq \frac{1}{4} \min(\|\tilde{x} - x^*\|, \|\hat{x} - x^*\|) < \frac{1}{4}\rho. \quad (15)$$

□

Corollary 4.4 (Newton iteration). *Let $\diamond_{x^*} F$ be bijective on a ball $\mathcal{D} = B(x^*, \bar{R})$, where x^* is an isolated root of $F \in \text{span}(\Phi)$. Then on $B(x^*, \rho)$, where ρ is defined as in Lemma 4.2, the Newton step is defined in both the tangent and secant mode and maps back into $B(x^*, \rho)$. The thus definable Newton iteration converges at least linearly.*

Proof. As the next Newton iterate is among the roots of minimal distance to the basis point(s) of the linearization, one needs to ensure that any root that $\diamond_{\hat{x}} F$ may have outside the ball $B(x^*, R)$ has a larger distance to $\{\tilde{x}, \hat{x}\}$ than the root that is known to exist inside $B(x^*, \frac{1}{4}\rho)$. The distance from the basis points inside $B(x^*, \rho)$ to the outside of $B(x^*, R)$ is at least $R - \rho = \frac{2}{3}R$. The distance from the root of $\diamond_{\hat{x}} F$ inside $B(x^*, \frac{1}{4}R)$ to the basis points is at most $\frac{1}{4}\rho + \rho = \frac{5}{12}R$ and thus the smaller distance.

By equation (15) of the last lemma we also see that the distance of the root of $\diamond_{\hat{x}} F$ to x^* is at most $\frac{1}{4}$ the distance of the basis points to x^* , which implies linear convergence. □

Corollary 4.5 (convergence rates). *Let $R = \min(\bar{R}, \frac{1}{c\gamma_F})$ and $\rho = \frac{R}{3}$. Then for all initial points $x_0, x_1 \in B(x^*, \rho)$ the Newton iteration in tangent mode*

$$x_{k+1} = N(x_k)$$

converges quadratically resp. in secant mode

$$x_{k+1} = N(x_k, x_{k-1})$$

converges with order $\frac{1+\sqrt{5}}{2}$ towards x^ .*

Proof. With the choice of the initial points and by Lemma 4.3, the full iteration sequence stays inside $B(x^*, \rho)$.

Tangent mode: By replacing x, \check{x}, \hat{x} with x_{j+1}, x_j, x_j equation (15) in Lemma 4.3 we get

$$\|x_{j+1} - x^*\| \leq \frac{1}{4\rho} \|x_j - x^*\|^2$$

which implies the quadratic convergence of the sequence (x_j) towards x^* ,

$$\|x_j - x^*\| \leq 4\rho \left(\frac{\|x_0 - x^*\|}{4\rho} \right)^{2^j}.$$

Secant mode: Replacing x, \check{x}, \hat{x} with x_{j+2}, x_{j+1}, x_j in equation (15) in Lemma 4.3 one finds

$$\|x_{j+2} - x^*\| \leq \frac{1}{4\rho} \|x_{j+1} - x^*\| \|x_j - x^*\|.$$

As in the scalar secant method, this implies convergence with rate $\phi = \frac{1+\sqrt{5}}{2}$ or more precisely

$$\|x_j - x^*\| \leq 4\rho \left(\frac{\|x_0 - x^*\|}{4\rho} \right)^{F_{j-1}} \left(\frac{\|x_1 - x^*\|}{4\rho} \right)^{F_j} \leq 4\rho \left(\frac{\max(\|x_0 - x^*\|, \|x_1 - x^*\|)}{4\rho} \right)^{\phi^j}.$$

where (F_j) is the Fibonacci sequence with $F_0 = 0, F_1 = 1$ and $F_{j+1} \leq \phi^j$. \square

The inner iterations of (13) require the solution of piecewise linear equations. However, these may possess several solutions Δx_j , of which we must find one of minimal norm. So far, the solvers that we know and surveyed in [Gri13, GBR15, SGR14, Rad16] require at least coherent orientation to guarantee convergence. Hence, the actual implementation of successive piecewise linearization needs further study.

5 Singularity Free Implementation

In contrast to the tangent mode of piecewise linearization, the secant mode involves two points of evaluation $\check{v}_i, \hat{v}_i \in \mathbb{R}^n$, which means that its computational cost of the primal values are roughly twice that of the tangent mode. These define a line segment $[\check{v}_i, \hat{v}_i] \equiv \{\lambda\check{v}_i + (1-\lambda)\hat{v}_i \mid \lambda \in [0, 1]\}$ for any intermediate operation. The formal definition of the secant slope given in equation (7) may cause numerically unstable divisions when the denominator gets small during the transition from secant to tangent mode, e.g. when the secant mode Newton iteration scheme converges.

However, in this section we will provide *singularity free* closed form expressions. Therefore an exception handling at $\check{v}_i = \hat{v}_i$ will no longer be necessary. To that end we move from the line segment representation to a midpoint-radius based representation. Now let $v = \varphi(u)$, where $\varphi \in \{\sin, \exp, \dots\}$ is some elementary operation and

$$(\check{v}_i, \hat{v}_i) = (\varphi(\check{v}_j), \varphi(\hat{v}_j)) \mapsto (\hat{v}_i, \delta v_i), \quad \text{where } \hat{v}_i = \frac{\check{v}_i + \hat{v}_i}{2} \text{ and } \delta v_i = -\frac{\check{v}_i - \hat{v}_i}{2}.$$

We adopted the concept of representing intervals via midpoint and radius from interval arithmetic calculus (described in detail e.g. by Siegfried Rump in [Rum99], or by Götz Alefeld and Jürgen Herzberger in [AH12]). We remark though, that in the present setting the radius $\delta v_i \in \mathbb{R}$ is allowed to become negative as well.

Now one can rewrite the secant slope of differentiable functions to the new representation

$$c_{ij} \equiv \frac{\check{v}_i - \hat{v}_i}{\check{v}_j - \hat{v}_j} = \frac{\delta v_i}{\delta v_j}.$$

Using some algebraic manipulations one can find individual formulas for the aforementioned propagation rules of the secant mode:

binary operation	\hat{v}_i	δv_i	c_{ij}	c_{ik}
$v_i = v_j + v_k$	$\hat{v}_j + \hat{v}_k$	$\delta v_j + \delta v_k$	1	1
$v_i = v_j - v_k$	$\hat{v}_j - \hat{v}_k$	$\delta v_j - \delta v_k$	1	-1
$v_i = v_j \cdot v_k$	$\hat{v}_j \hat{v}_k + \delta v_j \delta v_k$	$\delta v_j \hat{v}_k + \hat{v}_j \delta v_k$	\hat{v}_k	\hat{v}_j
$v_i = \frac{v_j}{v_k}$	$\frac{\hat{v}_j \hat{v}_k - \delta v_j \delta v_k}{\hat{v}_k^2 - \delta v_k^2}$	$\frac{\delta v_j \hat{v}_k - \hat{v}_j \delta v_k}{\hat{v}_k^2 - \delta v_k^2}$	$\frac{1}{\hat{v}_k}$	$-\frac{\hat{v}_j}{\hat{v}_k^2 - \delta v_k^2}$

Alternatively, we could represent $v_i = v_j/v_k$ as an application of a multiplication on the reciprocal $1/v_k$. Furthermore, we can represent the multiplication by the

Appolonius identity as above. Moreover, for unary operations we get:

unary operation	\hat{v}_i	δv_i	c_{ij}
$v_i = \sin(v_j)$	$\sin(\hat{v}_j) \cos(\delta v_j)$	$\cos(\hat{v}_j) \sin(\delta v_j)$	$\cos(\hat{v}_j) \operatorname{sinc}(\delta v_j)$
$v_i = \cos(v_j)$	$\cos(\hat{v}_j) \cos(\delta v_j)$	$-\sin(\hat{v}_j) \sin(\delta v_j)$	$-\sin(\hat{v}_j) \operatorname{sinc}(\delta v_j)$
$v_i = \exp(v_j)$	$\exp(\hat{v}_j) \cosh(\delta v_j)$	$\exp(\hat{v}_j) \sinh(\delta v_j)$	$\exp(\hat{v}_j) \operatorname{sinhc}(\delta v_j)$
$v_i = \log(v_j)$	$\frac{1}{2} \log(\hat{v}_j^2 + \delta v_j^2)$	$\operatorname{artanh}\left(\frac{\delta v_j}{\hat{v}_j}\right)$	$\frac{1}{\hat{v}_j} \operatorname{artanhc}\left(\frac{\delta v_j}{\hat{v}_j}\right)$

Note that by, e.g., [Wei16] $\operatorname{sinc}(x)$ and $\operatorname{sinhc}(x)$ (hyperbolic $\operatorname{sinc}(x)$) have regular Taylor expansions

$$\sin(x) = x \cdot \operatorname{sinc}(x) = x \cdot \sum_{i=0}^{\infty} \frac{(-x^2)^n}{(2n+1)!}, \quad \operatorname{sinc}(x) \equiv x \cdot \operatorname{sinhc}(x) \equiv x \cdot \sum_{i=0}^{\infty} \frac{x^{2n}}{(2n+1)!}.$$

We want to define $\operatorname{artanhc}(x)$ similar to $\operatorname{sinhc}(x)$ via its Taylor expansion:

$$\tanh^{-1}(x) = \operatorname{artanh}(x) = x \cdot \operatorname{artanhc}(x) = x \cdot \sum_{i=0}^{\infty} \frac{x^{2n}}{2n+1}$$

and it can be implemented in a similar fashion as $\operatorname{sinhc}(x)$ from the boost c++ libraries (see [lib]). For Root functions ($v_i = \sqrt[n]{v_j}$), general Powers ($v_i = v_j^c$, or in a binary fashion $v_i = v_j^{v_k}$) and monomials ($v_i = v_j^n$) one can use the identity

$$v_i = v_j^c = \exp(c \cdot \log(v_j)) \quad \text{or} \quad v_i = v_j^{v_k} = \exp(v_k \cdot \log(v_j))$$

and apply the rules above. Of course, the base $v_j > 0$ has to be positive, but there is a less restrictive alternative for monomials:

monomials	\hat{v}_i	δv_i	c_{ij}
$v_i = v_j^n$ (n natural number)	$\sum_{\substack{k \leq n \\ \text{even}}} \binom{n}{k} \hat{v}_j^{n-k} \delta v_j^k$	$\sum_{\substack{k \leq n \\ \text{odd}}} \binom{n}{k} \hat{v}_j^{n-k} \delta v_j^k$	$\sum_{\substack{k \leq n \\ \text{odd}}} \binom{n}{k} \hat{v}_j^{n-k} \delta v_j^{k-1}$
$v_i = v_j^2$ (special case: square)	$\hat{v}_j^2 + \delta v_j^2$	$2\hat{v}_j \delta v_j$	$2\hat{v}_j$

Remark 1 (General approximation for unary operations). Of course one can find a lot more singularity free formulas for the secant slopes of other operations. Using a Taylor expansion approach one can provide general approximation for-

mulas for the triplet $\hat{v}_i, \delta v_i$ and c_{ij} by:

$$\hat{v}_i = \frac{1}{2} \left(\sum_{k \geq 0} \frac{\varphi^{(k)}(\hat{v}_j)}{k!} \delta v_j^k + \sum_{k \geq 0} \frac{\varphi^{(k)}(\hat{v}_j)}{k!} (-\delta v_j)^k \right) = \sum_{\substack{k \geq 0 \\ k \text{ even}}} \frac{\varphi^{(k)}(\hat{v}_j)}{k!} \delta v_j^k,$$

$$\delta v_j = \frac{\varphi(\hat{v}_j + \delta v_j) - \varphi(\hat{v}_j - \delta v_j)}{2} = \sum_{\substack{k > 0 \\ k \text{ odd}}} \frac{\varphi^{(k)}(\hat{v}_j)}{k!} \delta v_j^k, \quad c_{ij} = \sum_{\substack{k > 0 \\ k \text{ odd}}} \frac{\varphi^{(k)}(\hat{v}_j)}{k!} \delta v_j^{k-1}.$$

6 Numerical Example

Consider the function

$$F(x) = \begin{bmatrix} \cos(\varphi(\angle x) - \angle x) & -\sin(\varphi(\angle x) - \angle x) \\ \sin(\varphi(\angle x) - \angle x) & \cos(\varphi(\angle x) - \angle x) \end{bmatrix} \cdot x - c,$$

where $c = [1.001, 10.01]^\top$, and $\angle x \in [0, 2\pi[$ is the angle of $x = (x_1, x_2)^\top$ in polar coordinate representation. Moreover, φ , which is defined by

$$\varphi(\psi) \equiv \psi + \frac{8}{5\pi} \psi^2 - \frac{8}{5\pi^2} \psi^3 + \frac{2}{5\pi^3} \psi^4$$

maps $[0, 2\pi[$ strictly monotonically onto itself. Hence, $F(x)$ is bijective. Furthermore, the function is differentiable everywhere except at the origin. There it is, just as the Euclidean norm, locally Lipschitz continuous and not even piecewise differentiable in the sense of [Sch12].

We investigated the behavior of the tangent and secant mode Newton on F both with and without noise. That is, we investigated F and \tilde{F} , where

$$\tilde{F}(x) = F(x) + \frac{\sin(5000 \cdot [x_1 + x_2])}{10^4}.$$

The secant mode was started with the initial values $\tilde{x} = [-3.7, -2.05]^\top$ and $\hat{x} = [7.0, 8.0]^\top$. The tangent mode was started with the mean value of the latter points.

We recall the well known formula for approximating the convergence rate numerically:

$$\gamma \approx \frac{\log \left| \frac{x_{n+1} - x_n}{x_n - x_{n-1}} \right|}{\log \left| \frac{x_n - x_{n-1}}{x_{n-1} - x_{n-2}} \right|}.$$

The first table shows both modes' residuals in the iterations without noise.

iteration	residual with tangent mode	res. with secant mode
0	13.3919956235	5.81435555868
1	5.65630249881	19.6157765738
2	3.39957297287	4.99712831052
3	0.00920821188601	1.2635478817
4	$2.65403135735e - 06$	0.0882985228824
5	$2.13162820728e - 13$	0.00192152171278
6	$8.881784197e - 15$	$5.19844852542e - 06$
7		$3.0607116841e - 10$
8		$8.881784197e - 16$
γ	$2.07814399547 \approx 2$	$1.64753467681 \approx \frac{1+\sqrt{5}}{2}$

The next table shows the residuals of the iterations with noise.

iteration	residual with tangent mode	res. with secant mode
0	13.3920216719	5.81445152008
1	60.308012713	19.6157077846
2	81.8554424857	4.99709320593
3	8.54532016753	1.26352999658
4	5.69986744799	0.0883961689187
5	1.92933721639	0.00178884970844
6	0.425650504358	0.00013247524387
7	0.127098157087	$3.54530493811e - 05$
8	0.0253173333077	$2.27749388494e - 06$
9	0.00248354296218	$3.65673925216e - 08$
10	0.000950425306504	$3.71373256312e - 11$
11	$6.728080753e - 05$	$4.95408564432e - 15$
12	$1.87922375892e - 06$	
13	$1.57356337999e - 09$	
14	$2.89904818901e - 15$	
γ	$2.01181918489 \approx 2$	$1.66153582947 \approx \frac{1+\sqrt{5}}{2}$

Both methods attain their theoretical convergence rates in either iteration. However, we observe that the secant mode fares better with the problem with noise as it cuts through the latter, while the tangent Newton is thrown off for several iteration steps.

7 Final Remarks

The framework developed in the present paper, as well as in [Gri13, GBR15, SGRB14, Rad16, GHRS16], aims at presenting viable alternatives to current approaches to piecewise differentiability as it may occur, e.g., in nonsmooth nonlinear systems or ODEs with nonsmooth right hand side. The piecewise

linearizations, which were first introduced in [Gri13] can be obtained in an automated fashion by an adaptation of AD tools such as ADOL-C [WG12].

The generalized Newton iterations introduced in Section 5 are intended as an alternative to semismooth Newton [Hin10]. The quadratic convergence rate of the tangent version is in line with that of semismooth Newton. While for one step of semismooth Newton an appropriate element of the generalized derivative has to be calculated, the piecewise linear Newton's methods solve a piecewise linear system in each step. Both problems are NP-hard in general, but may be solvable in practice with essentially the same effort as an ordinary linear system. Hence, it is likely highly situation dependent, which approach yields better performance.

It is our hope that, in combination with the formulas for numerically stable implementation of the secant linearization, the generalized Newton's methods can be developed into robust and stable workhorse algorithms, which might even outperform the quadratic methods on selected problems. For example on problems with oscillating noise we observed that the secant method required fewer Newton steps to converge, as it cuts through the oscillations.

Acknowledgements

The work for the article has been partially conducted within the Research Campus MODAL funded by the German Federal Ministry of Education and Research (BMBF) (fund number 05M14ZAM).

References

- [AH12] ALEFELD, G. ; HERZBERGER, J.: *Introduction to Interval Computation*. Elsevier Science, 2012 (Computer Science and Applied Mathematics). <https://books.google.de/books?id=rUsX5x00qUcC>. – ISBN 9780080916361
- [Cla83] CLARKE, F.H.: *Optimization and Nonsmooth Analysis*. Wiley-Interscience, 1983 (Canadian Mathematical Society Series of Monographs and Advanced Texts). – ISBN 047187504X
- [GBRS15] GRIEWANK, A. ; BERNT, J.U. ; RADONS, M. ; STREUBEL, T.: Solving piecewise linear systems in abs-normal form. In: *Linear Algebra and its Applications* 471 (2015), Nr. 0, 500 - 530. <http://www.sciencedirect.com/science/article/pii/S0024379514008209>
- [GHR16] GRIEWANK, A. ; HASENFELDER, R. ; RADONS, M. ; STREUBEL, T.: Integrating Lipschitzian Dynamical Systems using Piecewise Algorithmic Differentiation. In: *Submitted 2016* (2016)
- [Gri13] GRIEWANK, A.: On stable piecewise linearization and generalized algorithmic differentiation. In: *Optimization Methods and Software* 28

- (2013), Nr. 6, 1139-1178. <http://dx.doi.org/10.1080/10556788.2013.796683>. – DOI 10.1080/10556788.2013.796683
- [GW08] GRIEWANK, A. ; WALTHER, A.: *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation*. Society for Industrial and Applied Mathematics (SIAM), 2008 (Other Titles in Applied Mathematics). <http://epubs.siam.org/doi/book/10.1137/1.9780898717761>. – ISBN 978-0-89871-776-1
- [Hin10] HINTERMULLER, M.: *Semismooth Newton methods and applications*. Oberwolfach Seminar on Mathematics of PDE-Constrained Optimization, Mathematisches Forschungsinstitut, Oberwolfach, Germany, 2010
- [iee85] IEEE Standard for Binary Floating-Point Arithmetic. In: *ANSI/IEEE Std 754-1985* (1985). <http://dx.doi.org/10.1109/IEEESTD.1985.82928>
- [Kum88] KUMMER, B.: Newton's method for non-differentiable functions. In: *Advances in Math. Optimization* (1988)
- [lib] LIBRARIES boost c.: [boost/math/special_functions/sinhc.hpp](http://www.boost.org/doc/libs/1_62_0/boost/math/special_functions/sinhc.hpp). – retrieved from http://www.boost.org/doc/libs/1_62_0/boost/math/special_functions/sinhc.hpp, Last visited on 12/01/2016
- [Rad16] RADONS, M.: Direct solution of piecewise linear systems. In: *Theoretical Computer Science* 626 (2016), S. 97–109
- [Rad17] RADONS, M.: Degree theory for piecewise affine functions. In: *Submitted 2017* (2017)
- [Rum99] RUMP, Siegfried M.: Fast and Parallel Interval Arithmetic. In: *BIT Numerical Mathematics* 39 (1999), Nr. 3, 534–554. <http://dx.doi.org/10.1023/A:1022374804152>. – DOI 10.1023/A:1022374804152. – ISSN 1572-9125
- [Sch12] SCHOLTES, S.: *Introduction to Piecewise Differentiable Equations*. Springer New York, 2012 (SpringerBriefs in optimization). <http://link.springer.com/book/10.1007/978-1-4614-4340-7>. – ISBN 978-1-4614-4340-7
- [SGRB14] STREUBEL, T. ; GRIEWANK, A. ; RADONS, M. ; BERNT, J.U.: Representation and Analysis of Piecewise Linear Functions in Abs-normal form. In: *Proc. of the IFIP TC 7* (2014), S. 323–332
- [Wei16] WEISSTEIN, Eric W.: *Sinhc Function*. From MathWorld—A Wolfram Web Resource. [\url{http://mathworld.wolfram.com/SinhcFunction.html}](http://mathworld.wolfram.com/SinhcFunction.html). Version:1999-2016. – Last visited on 12/01/2016

- [WG12] WALTHER, A. ; GRIEWANK, A.: Getting started with ADOL-C.
In: NAUMANN, U. (Hrsg.) ; SCHENK, O. (Hrsg.): *Combinatorial
Scientific Computing*. Chapman-Hall CRC Computational Science,
2012, Kapitel 7, S. 181–202