On the abs-polynomial expansion of piecewise smooth functions

A. Griewank[a] and T. Streubel[b], and C Tischendorf[a]

[a]Mathematics Institute Humboldt University Berlin;[b]HU Berlin and Zuse Institute Berlin

ARTICLE HISTORY
Compiled July 22, 2020

ABSTRACT
Tom Streubel has observed that for functions in abs-normal form, generalized Taylor expansions of arbitrary order $\bar{d}-1$ can be generated by algorithmic piecewise differentiation. Abs-normal form means that the real or vector valued function is defined by an evaluation procedure that involves the absolute value function $|\cdot|$ apart from arithmetic operations and $\bar{d}$ times continuously differentiable univariate intrinsic functions. The additive terms in Streubel's expansion are abs-polynomial, i.e. involve neither divisions nor intrinsics. When and where no absolute values occur, Moore's recurrences can be used to propagate univariate Taylor polynomials through the evaluation procedure with a computational effort of $O(\bar{d}^2)$, provided all univariate intrinsics are defined as solutions of linear ODEs. This regularity assumption holds for all standard intrinsics, but for irregular elementaries one has to resort to Faa di Bruno's formula, which has exponential complexity in $\bar{d}$. As already conjectured, we show that the Moore recurrences can be adapted for regular intrincics to the abs-normal case. Finally, we observe that where the intrinsics are real analytic the expansions can be extended to infinite series that converge absolutely on spherical domains.

KEYWORDS
Nonsmooth Taylor polynomial/series; forward mode propagation; abs-normal form; abs-linear form; absolute convergence; Moore recurrences; quadratic complexity.

1. Introduction and Notation

Over the last few decades there has been an increasing interest in nonsmooth analysis and algorithmic design (see e.g. Clarke [3], Demyanov [5], Scholtes [26], Kummmer/Klatte [18], Mordukhovich[21], and Schirotzek[25] to name just a few). Fortunately, practically all nonsmooth functions arising in applications are actually piecewise smooth, and in very many cases the nonsmoothness can be cast in terms of the absolute value function, which immediately includes $\min, \max$ and the positive part function pos. In other words the vector function $f : x \in \mathcal{D} \subset \mathbb{R}^n \to y \in \mathbb{R}^m$ in question is evaluated by a sequence of assignments or *instructions*

$$v_i = v_j \circ v_k \quad \text{or} \quad v_i = \varphi_i(v_j) \quad \text{for} \quad i = 1 \ldots \ell \,.$$

CONTACT A. Griewank. Email: griewank@math.hu-berlin.de

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

$$\varphi_i \in \Phi \;\equiv\; \{\text{c, rec, sqrt}, \sin, \cos, \exp, \log, \ldots, \mathbf{abs}, \ldots\}$$

a univariate function. The symbol $c \in \mathbb{R}$ represents initialization to a constant, and to simplify the notation we interpret the division as a reciprocal $\text{rec}(u) \equiv 1/u$, followed by a multiplication. Later on we will distinguish four types of instructions: initialization and addition or subtraction, multiplication, smooth nonlinear intrinsics, and the absolute value function. Adopting the notation from [13] we partition the sequence of scalar variables v_i into the vector triple

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

such that $x \in \mathbb{R}^n$ is the vector of independent, $y \in \mathbb{R}^m$ 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, the square root and the logarithm are not globally defined. Hence, we will assume throughout this paper that all elementals are evaluated in the interior of their domain of definition, where they are $\bar{d} \leq \infty$ times continuously differentiable. In other words 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 for $x \in \mathcal{D}$. The resulting function $y = f(x)$ will then be called *abs-normal* and their set will be denoted by $\mathcal{C}^{\bar{d}}_{\text{abs}}(\mathcal{D})$. This set naturally forms a linear space over the reals and was apparently first considered in [8] in the context of algorithmic directional differentiation. Our main result is still valid if we relax the assumption of $\bar{d}$ times continuous differentiability to local Lipschitz-continuity of the $\bar{d}$-1st derivative.

The subspace of functions that are continuous and piecewise polynomial with maximal degree $\bar{d}$ on each subdomain will be denoted by

$$\mathcal{P}^{\bar{d}}_{\text{abs}}(\mathcal{D}) \;\subset\; \mathcal{C}^{\bar{d}}_{\text{abs}}(\mathcal{D}) \; .$$

It follows from the min$-$max representation in Scholtes [26] that $\mathcal{P}^1_{\text{abs}}(\mathcal{D})$ is in fact the space of all continuous piecewise linear functions, whose domain can be extended without loss of generality to all of $\mathbb{R}^n = \mathcal{D}$. For $\bar{d} > 1$ it is not clear at the moment whether the abs-polynomial space $\mathcal{P}^{\bar{d}}_{\text{abs}}(\mathcal{D})$ contains all continuous functions that are piecewise polynomial or whether it is just a subset of those. Of course evaluating a function in abs-polynomial form is much easier than a piecewise polynomial function defined by a set of polynomially constraint sub domains and a polynomial selection function on each of them. In the third section of this note we will allow the possibility $\bar{d} = \infty$ so that we are dealing with numerical series that resemble power series with respect to the implicit variable Δx.

In [9] it was shown in the style of forward automatic or algorithmic differentiation [24], [13] that the evaluation procedure defined above can be extended to yield at any reference point $\mathring{x} \in \mathcal{D}$ a piecewise linear function $\Delta^1 f(\mathring{x}; \Delta x)$ which approximates the difference $f(\mathring{x} + \Delta x) - f(\mathring{x})$ up to a term of $O(\|\Delta x\|^2)$. This local piecewise linear model can be used for optimization [14], [17], [7], [6], [15], [16],[30], [19], [4], equation solving [12], [10], [27], [23], [22], [2] and other numerical tasks [11], [1], [29].

The salient point of piecewise linearization is that the resulting subdomains of linearity are polyhedral and that the combinatorial aspect of dealing with such a polyhedral decomposition is in principle a matter of finite linear algebra. The generalization proposed in [28] and discussed here yields semi-algebraic subdomains and functions, which are much harder to deal with, whether and when this effort pays off remains to be seen. Applications to ODE and DAE solving are studied in [28] but are not specifically addressed here.

2. Finite abs-polynomial expansion

In this section assume that $\bar{d} < \infty$, so that the intrinsics are only required to be finitely often differentiable. Given a fixed reference point $\mathring{x}$ and a variable increment $\Delta x \in \mathcal{D} - \mathring{x}$ we look for component functions

$$\Delta^d f(\mathring{x}; \Delta x) \; \in \; \mathcal{P}^d_{\text{abs}}(\mathcal{D} - \mathring{x}) \quad \text{for} \quad 0 \leq d < \bar{d} \tag{1}$$

such that

$$f(\mathring{x} + \Delta x) \; = \; \sum_{p=0}^{d-1} \Delta^p f(\mathring{x}; \Delta x) + O(\|\Delta x\|^d) \quad \text{for} \quad 0 < d \leq \bar{d}\,. \tag{2}$$

From the above relation (2) it follows immediately by subtraction that we must have

$$\Delta^d f(\mathring{x}; \Delta x) \; = \; O(\|\Delta x\|^d) \quad \text{for} \quad 0 < d \leq \bar{d}\,, \tag{3}$$

which means that the $\Delta^d f(\mathring{x}; \Delta x)$ have both the *degree* and the *order* d.

In the article [28] the formula (2) is called a generalized Taylor expansion. We have refrained from doing this for the following reasons. In the last few decades it has been recognized that power series of trigonometric functions and their inverses were described in the Kerala school of mathematics hundreds of years before the European development of calculus. Charles M. Wish, an English civil servant in the Madras Establishment of the East India Company, published in 1834 the article [31] demonstrating some amazing achievements of the Kerala school of mathematics in the 13-th and 14-th century. This went practically unnoticed by the mathematical community until about the turn of the century or should we say millenium.

Some modern authors notice critically that the Indian approach was not founded in an abstract theory of limits and ignored the inverse relation between differentiation and integration via the fundamental theorem of calculus. However, in some sense algorithmic or automatic differentiation and the piecewise smooth extension considered here follows the more algorithmic Indian tradition in that typically only the Taylor coefficients of the standard intrinsics enter into the calculation and the fundamental theorem does not need to be invoked. So one might in fact consider calling the expansions here finite and infinite Kerala series.

The remainder of this note consists essentially in the proof that (2) is true with suitable component functions satisfying (1). Of course, this will be done by induction on the the degree $0 \leq \bar{d}$ and the instruction counter $i = 1 \ldots l$.

In other words we will show that for all $i = 1-n \ldots \ell$ and $0 < d \leq \bar{d}$

$$v_i(\mathring{x} + \Delta x) - O(\|\Delta x\|^d) \;=\; \lozenge^{(d-1)} v_i(\mathring{x}; \Delta x) \;\equiv\; \sum_{p=0}^{d-1} \Delta^p v_i(\mathring{x}; \Delta x) \;, \qquad (4)$$

where also

$$\mathcal{P}^p_{\text{abs}}(\mathcal{D} - \mathring{x}) \;\ni\; \Delta^p v_i(\mathring{x}; \Delta x) \;=\; O(\|\Delta x\|^p) \;. \qquad (5)$$

Since $\mathring{x}$ and mostly also Δx are constant throughout the remainder of the paper we will omit them frequently as arguments of the component functions $\Delta^d v_i$ as well as $\lozenge^d v_i$. Notice that the increment Δx is not a direction and that $\Delta^d v_i(\mathring{x}; \Delta x)$ is generally not homogeneous with respect to even positive scaling.

Initialization and Addition/Subtraction

For all $i \leq 0$ the principal (2) relationship is satisfied with

$$\Delta^0 v_i = \mathring{x}_{n+i}, \quad \Delta^1 v_i = \Delta x_{n+i} \quad \text{and} \quad \Delta^j v_i = 0 \;\text{ for }\; 1 < j < \bar{d} \;,$$

where $\mathring{x}_{n+i}$ and Δx_{n+i} denote the i-th component of the reference vector $\mathring{x}$ and the increment Δx, respectively. Clearly all these component functions are contained in the spaces of constant or piecewise linear functions such that (5) holds. For any constant initialization $v_i = c$ we set accordingly

$$\Delta^0 v_i = c \quad \text{and} \quad \Delta^d v_i = 0 \;\text{ for }\; 0 < d < \bar{d} \;.$$

Here and throughout we use the identity $\Delta^0 v_i = v_i(\mathring{x})$ for the intermediate variables themselves. Now, when $v_i = v_j \circ v_k$ with $\circ \in \{+,-\}$ we get the linear recursions

$$\Delta^d v_i = \Delta^d v_j \circ \Delta^d v_k \;\text{ for }\; 0 \leq d < \bar{d} \;,$$

which obviously maintain the relations (5) and it can be easily checked by the triangle inequality that (4) is also satisfied.

The product rule

A little harder is the proof for the product relation $v_i = v_j \cdot v_k$, where we set

$$\Delta^d v_i(\Delta x) \;=\; \sum_{p=0}^{d} \Delta^p v_j \cdot \Delta^{d-p} v_k \;\text{ for }\; 0 \leq d < \bar{d} \;. \qquad (6)$$

Since the degree and the order a products of piecewise polynomial functions is the sum of the degrees and orders of its factors, respectively, we find that both are bounded by d for each summand on the right so that (5) is certainly maintained. The approximation property (4) is obtained as follows.

Lemma 2.1.
The definition (6) *ensures that* (4) *holds for* $v_i(x)$ *if it holds for* $v_j(x)$ *and* $v_k(x)$.

Proof Abbreviation as usual $v = v_i, u = v_j, w = v_k$ and omitting the common argument Δx we find that for $0 \leq d \leq \bar{d}$

$$\begin{aligned}
v(\mathring{x} + \Delta x) - \sum_{p=0}^{d-1} \Delta^p v &= u(\mathring{x} + \Delta x) \cdot w(\mathring{x} + \Delta x) - \sum_{p=0}^{d-1} \Delta^p v \\
&= \left[\sum_{p=0}^{d-1} \Delta^p u + O(\|\Delta x\|^d)\right] \left[\sum_{p=0}^{d-1} \Delta^p w + O(\|\Delta x\|^d)\right] - \sum_{p=0}^{d-1} \Delta^p v \\
&= \left[\sum_{p=0}^{d-1} \Delta^p u\right] \left[\sum_{q=0}^{d-1} \Delta^q w\right] - \sum_{p=0}^{d-1} \Delta^p v + O(\|\Delta x\|^d) \\
&= \sum_{p+q \,\geq\, d} \Delta^p u \cdot \Delta^q w + O(\|\Delta x\|^d) = O(\|\Delta x\|^d)
\end{aligned}$$

where we have used (3) for the last order equality. **qed**

The absolute value rule

The really new aspect algorithmic piecewise differentiation is the admission of the absolute value function $v_i = \mathbf{abs}(v_j) = |v_j|$ where abbreviating $v_i = v$ and $v_j = u$ we set for $d = 0 \ldots \bar{d} - 1$

$$\Delta^d v = \left|\sum_{p=0}^{d} \Delta^p u\right| - \sum_{p=0}^{d-1} \Delta^p v \tag{7}$$

$$= \left|\Diamond^d u\right| - \left|\Diamond^d u - \Delta^d u\right| \tag{8}$$

$$= \Delta^d u \left(2\, \Diamond^d u - \Delta^d u\right) \Big/ \left(\left|\Diamond^d u\right| + \left|\Diamond^{(d-1)} u\right|\right) . \tag{9}$$

Here the partial accumulations $\Diamond^{(d)}$ are defined in (4) and the last equality follows by some elementary manipulations. The components $\Delta^d v$ inherit the right degree d from the $\Delta^p u$ for $p \leq d$ and applying the inverse triangle inequality to (8) we see hat they have the same order since $|\Delta^d v| \leq |\Delta^d u| = O(\|\Delta x\|^d)$ by induction. The reformulation (9) is more stable as it avoids cancelation errors. Finally, we obtain (4) by induction on d as

$$\begin{aligned}
v(\mathring{x} + \Delta x) - \sum_{p=0}^{d-1} \Delta^p v &= |u(\mathring{x} + \Delta x)| - \sum_{p=0}^{d-1} \Delta^p v \\
&= \left|\sum_{p=0}^{d-1} \Delta^p u\right| + O(\|\Delta x\|^d) - \sum_{p=0}^{d-2} \Delta^p v - \Delta^{d-1} v \;=\; O(\|\Delta x\|^d)
\end{aligned}$$

where we have used (7) with d replaced by $d-1$ for the last equality. For $d=1$ the relation follows from $\Delta^0 v = v(\mathring{x})$ and the directly inherited Lipschitz continuity of $v(\mathring{x}+\Delta x)$ w.r.t. Δx.

The nonlinear intrinsic rule

Thus we are left with the task to prove (4) for the case $v = v_i = \varphi(u)$ with $u = v_j$ where φ is $\bar{d}$ times continuously differentiable near the point $\mathring{u} = v_j(\mathring{u})$. Naturally we assume by induction on i that

$$u(\mathring{x}+\Delta x) - O(\|\Delta x\|^d) = \sum_{p=0}^{d-1} \Delta^p u = \mathring{u} + O(\|\Delta x\|) \quad \text{for} \quad 0 \le d < \bar{d}\,. \tag{10}$$

With $\mathring{\varphi}_q$ for $0 \le q < \bar{d}$ the Taylor coefficient of φ at $\mathring{u}$ it follows using the Lipschitz continuity of φ that for any $0 \le d < \bar{d}$

$$\begin{aligned} v(\mathring{x}+\Delta x) - O(\|\Delta x\|)^d &= \varphi(u(\mathring{x}+\Delta x)) - O(\|\Delta x\|)^d \\ &= \varphi\Big(\sum_{p=0}^{d-1} \Delta^p u\Big) = \varphi\Big(\mathring{u} + \sum_{p=1}^{d-1} \Delta^p u\Big) \\ &= \mathring{v} + \sum_{q=1}^{d-1} \mathring{\varphi}_q \Big(\sum_{p=1}^{d-1} \Delta^p u\Big)^q + O(\|\Delta x\|)^d \qquad (11) \\ &= \sum_{q=0}^{d-1} \Delta^q v + O(\|\Delta x\|)^d \qquad (12) \end{aligned}$$

The components $\Delta^q v$ for $0 \le q < d$ are obtained by collecting all products that have the same degree q in the expansion of (11). The finitely many higher order products can be incorporated into the residual term $O(\|\Delta x\|)^d$. The resulting expressions for the $\Delta^q v$ in terms of the $\Delta^p u$ and the $\mathring{\varphi}_q$ are called the Faa di Bruno formulas [20]. For a more or less explicit expression of the these formulas see [28]. They were originally obtained on the basis of the smooth univariate Taylor expansion

$$\varphi\Big(\sum_{p=0}^{d-1} t^p \Delta^p u\Big) = \sum_{q=0}^{d-1} t^p \Delta^q v + O(t^d)\,, \tag{13}$$

where Δx is considered constant. By differentiating this relation with respect to the parameter t and using the fact that φ solves a rational ODE one can obtain recurrences for the $\Delta^q v$ in terms of the $\Delta^p u$ with $p \le q$ that have only quadratic complexity in $\bar{d}$ as follows.

Moore recurrences

More specifically, suppose that $\varphi(u)$ satisfies for u in the interior of its domain the rational scalar ODE

$$b(u)\,\varphi'(u) - a(u)\,\varphi(u) = c(u)\ . \tag{14}$$

Here, the component functions $b(u)$, $a(u)$, and $c(u)$ are supposed to be "known" i.e., expressible in terms of the absolute value, arithmetic operations and intrinsic functions whose abs-linear expansion can already be computed. In the cases of main interest the triples $(a(u), b(u), c(u))$ are extremely simple, for example we have $(0, u, 1)$ for $\log(u)$, $(1, 1, 0)$ for $\exp(u)$, and $(r, u, 0)$ for u^r with $0 < r \in \mathbb{R}$. In other words, for any fixed increment Δx we can easily determine the abs-polynomial components $\tilde{a}_p = \Delta^p a(\Delta x)$, $\tilde{b}_p = \Delta^p b(\Delta x)$, and $\tilde{c}_p = \Delta^p c(\Delta x)$ from the scaled components $\Delta^p u(\Delta x)$. Then we may utilize the following proposition from [13]

Proposition 2.2 (Taylor Propagation for ODE Solutions).
Provided $\tilde{b}_0 \equiv b(u_0) \neq 0$ we have for $\tilde{u}_p = p\Delta^p u(\Delta x)$ and $\tilde{v}_p = p\Delta^p v(\Delta x)$

$$\tilde{v}_q \;=\; \frac{1}{\tilde{b}_0}\left[\sum_{p=1}^{q}\big(\tilde{c}_{q-p} + e_{q-p}\big)\tilde{u}_j - \sum_{p=1}^{q-1}\tilde{b}_{q-p}\tilde{v}_p\right] \quad \textit{for} \quad q = 1\ldots\bar{d}-1\ ,$$

where

$$e_q \;\equiv\; \sum_{p=0}^{q}\tilde{a}_p\,\nabla^{(q-p)}v(\Delta x) \quad \textit{for} \quad q = 0\ldots\bar{d}-1\ ,$$

where all component functions have exactly the degree and order indicated by their sub- or superscript.

The final assertion is not part of the original theorem but can be easily verified so that the resulting components $\Delta^d v_i(\Delta x) = \tilde{v}_d/d$ for $0 < d < \bar{d}$ satisfy indeed the requirement (5) and of course by (12) also the approximation property (4). Note that $\Delta^0 v_i(\Delta x) = \mathring{v}_i = \varphi(\mathring{v}_j)$ is directly available.

Summary of abs-polynomial Taylor expansion

In this main section we have shown by induction that any abs-normal function $f \in \mathcal{C}^{\bar{d}}_{\text{abs}}(\mathcal{D})$ can be expanded into a finite sum of $d < \bar{d}$ abs-polynomial components $\Delta^p f(\mathring{x}; \Delta x)$ for $p = 0\ldots d-1$ with an approximation error of $O(\|\Delta x\|^d)$. The abs-polynomial components $\Delta^p f(\mathring{x}; \Delta x)$ have both degree and order p. The right hand side of (2) will only be a useful approximation to the left hand side if the order terms are sufficiently small, i.e. for $\mathring{x} + \Delta x$ belonging to a small ball centered at the reference point $\mathring{x}$ and contained in the domain $\mathcal{D}$ where f is continuously defined.

3. The abs-polynomial series

In this final brief section we note that provided all intrinsics are real analytic we can set $\bar{d} = \infty$, so that the abs-polynomial expansions become numerical series that converge absolutely and represent $f(\mathring{x} + \Delta x)$ exactly for Δx inside a certain ball.

Proposition 3.1 (Series expansion)**.**
If all intrinsics $v_i = \varphi(v_j)$ are real analytic near their respective arguments $\mathring{v}_j$ so that $f \in \mathcal{C}^\infty_{\text{abs}}(\mathcal{D})$ there exist a positive radius $\rho \in (0, \infty]$ such that

$$f(\mathring{x} + \Delta x) \; = \; \sum_{d=0}^{\infty} \Delta^d f(\mathring{x}; \Delta x) \quad \textit{if} \quad \|\Delta x\| < \rho \, , \tag{15}$$

where now for all integer $d \geq 0$

$$\mathcal{P}^d_{\text{abs}}(\mathcal{D} - \mathring{x}) \; \ni \; \Delta^d f(\mathring{x}; \Delta x) \; = \; O(\|\Delta x\|^d) \, . \tag{16}$$

Proof To show this we proceed again by induction on the instruction index i. For the independent variables $v_i = x_{n+i}$ the expansions are finite so certainly absolutely convergent with $\rho_0 = \infty$. For linear operations and the multiplication between some v_j and v_k we get the minimal convergence radius $\rho_i = \min(\rho_j, \rho_k) > 0$. The latter follows from the fact that the Cauchy product of absolutely convergent series is also absolutely convergent to the correct limit. For the absolute value operation $v_i = |v_j|$ we find that simply $\rho_i = \rho_j$. Now let us consider finally a univariate intrinsic $v_i = \varphi(v_j)$. Provided φ is real analytic near $\mathring{v}_j$ it has a positive convergence radius $r > 0$. Then we we can restrict $\|\Delta x\| \leq \rho_i \leq \rho_j$ such that $|v_j(\mathring{x} + \Delta x) - \mathring{v}_j| = O(\|\Delta x\|) < r$ and the series for $v_j(\mathring{x} + \Delta x)$ is absolutely convergent. Consequently for $d = \infty$ the series on the right hand side of (11) is absolutely convergent and can by Fubini's principle be bracketed to yield the abs-polynomial series expansion of $v_i(\mathring{x} + \Delta x)$. Thus we can set $\rho = \min(\rho_i : \ell - m < i \leq \ell)$ to complete the proof. **qed**

Example 3.2. We first illustrate the result in one dimension for

$$v_1(x) = \sin(x) \qquad \text{and} \quad f(x) = \max(0, \sin(x)) \quad \text{at} \quad \mathring{x} = 0 \, .$$

The the corresponding abs-linear expansions are given by

$$\begin{aligned} v_1 \;&\equiv\; \sin(\Delta x) = \underbrace{0}_{\equiv \mathring{v}_0} + \underbrace{\Delta x}_{\equiv \Delta^1 v_1(\Delta x)} + \underbrace{0}_{\equiv \Delta^2 v_1(\Delta x)} - \underbrace{\tfrac{1}{6}(\Delta x)^3}_{\equiv \Delta^3 v_1(\Delta)} + \ldots + \underbrace{\tfrac{(-1)^d}{(2d+1)!}(\Delta x)^{2d+1}}_{\equiv \Delta^{2d+1} v_1(\Delta x)} + \ldots \\ v_2 \;&\equiv\; |v_1| = \underbrace{0}_{\equiv \mathring{v}_2} + \underbrace{|\Delta x|}_{\equiv \Delta^1 v_2(\Delta x)} + \underbrace{0}_{\equiv \Delta^2 v_2(\Delta x)} + \underbrace{\left|\Delta x - \tfrac{1}{6}(\Delta x)^3\right| - |\Delta x|}_{\equiv \Delta^3 v_2(\Delta x)} + \underbrace{0}_{\equiv \Delta^4 v_2(\Delta x)} + \ldots \\ f \;&\equiv\; \tfrac{1}{2}(v_1 + v_2) = 0 + \underbrace{\tfrac{1}{2}(\Delta x + |\Delta x|)}_{\equiv \Delta^1 f(\Delta x)} + 0 + \underbrace{\tfrac{1}{2}\left(\left|\Delta x - \tfrac{1}{6}(\Delta x)^3\right| - |\Delta x| - \tfrac{1}{6}(\Delta x)^3\right)}_{\equiv \Delta^3 f(\Delta x)} + \ldots \end{aligned}$$

where we have used throughout the positive part representation $\max(0, v) = \frac{1}{2}(v + |v|)$.

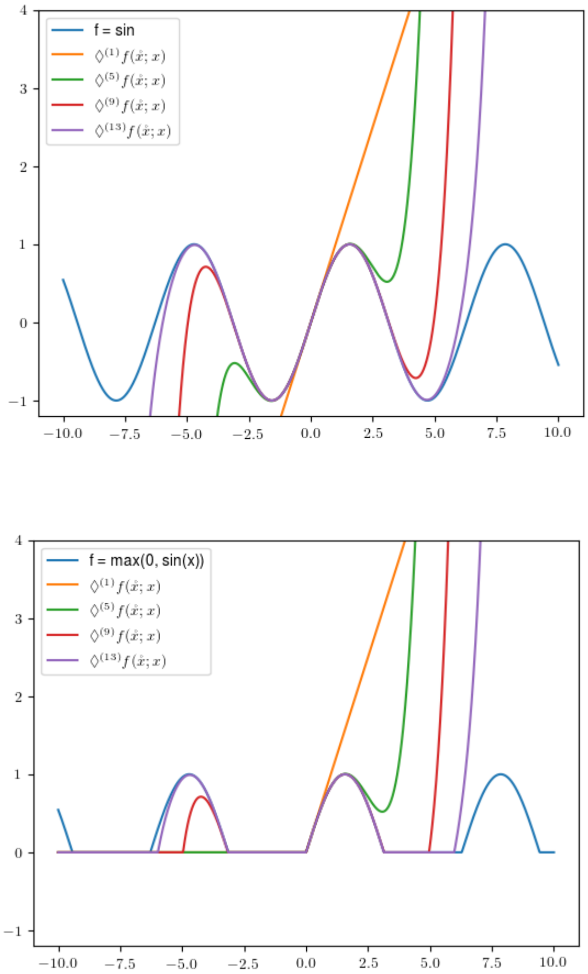


Figure 1. Abs-polynomial approximations of Example 3.2 with $\Delta x \equiv (x, y)$.

As one can see in Fig. 3.2 we get for the first intermediate variable $v_1(x) = \sin(x)$ just the Taylor polynomials, which oscillate rather widely as one moves away from the reference point. A similar effect occurs for the positive part function applied to the sine. It is worth noting that the third abs-polynomial component $\Delta^3 f$, though in agreement with (16) of degree and order 3, is not just a combination of cubic monomials with or without absolute values. While this holds in general, we have in this special case $\Diamond^d \max(0, \sin(x)) = \max(0, \Diamond^d \sin(x))$ so that the $d-$th abs-polynomial expansion is simply obtained by replacing $\sin(x)$ with its $d-$th Taylor expansion and then applying the positive part function. That is no longer the case when there are several nonlinear elemental functions that superimpose each other like in the last example 3 below.

Example 3.3. Secondly, we illustrate the result on the bivariate nonsmooth function

$$y = f(x_0, x_1) = |\exp(x_0) - |x_1|| \quad \text{at} \quad \mathring{x} = (0, 1)$$

We list first the intermediate variables v_1, v_2, v_3 and last the dependent $f \equiv v_4$.

$$
\begin{aligned}
v_1 &\equiv \exp(x_0) = \underbrace{1}_{\equiv \mathring{v}_0} + \underbrace{\Delta x_0}_{\equiv \Delta^1 v_1(\Delta x_0)} + \underbrace{\tfrac{1}{2}(\Delta x_0)^2}_{\equiv \Delta^2 v_1(\Delta x_0)} + \underbrace{\tfrac{1}{6}(\Delta x_0)^3}_{\equiv \Delta^3 v_1(\Delta x_0)} + \ldots + \underbrace{\tfrac{1}{d!}(\Delta x_0)^d}_{\equiv \Delta^d v_1(\Delta x_0)} \ldots \\
v_2 &\equiv |x_1| = \underbrace{1}_{\equiv \mathring{v}_1} + \underbrace{(|1+\Delta x_1| - 1)}_{\equiv \Delta^1 v_2(\Delta x_1)} + 0 \ldots 0 \ldots \\
v_3 &\equiv v_1 - v_2 = \underbrace{\Delta x_0 - |1+\Delta x_1| + 1}_{\equiv \Delta^1 v_3(\Delta x)} + \underbrace{\tfrac{1}{2}(\Delta x_0)^2}_{\equiv \Delta^2 v_3(\Delta x)} + \underbrace{\tfrac{1}{6}(\Delta x_0)^3}_{\equiv \Delta^3 v_3(\Delta x)} + \ldots + \underbrace{\tfrac{1}{d!}(\Delta x_0)^d}_{\equiv \Delta^d v_3(\Delta x)} \\
f &\equiv |v_3| = \underbrace{|\Delta x_0 - |1+\Delta x_1| + 1|}_{\equiv \Delta^1 f(\Delta x)} + \underbrace{\left|\Delta x_0 - |1+\Delta x_1| + 1 + \tfrac{1}{2}(\Delta x_0)^2\right| - \Delta^1 f(\Delta x)}_{\equiv \Delta^2 f(\Delta x)} \\
&\quad + \underbrace{\left|\Delta x_0 - |1+\Delta x_1| + 1 + \tfrac{1}{2}(\Delta x_0)^2 + \tfrac{1}{6}(\Delta x_0)^3\right| - \Delta^2 f(\Delta x)}_{\equiv \Delta^3 f(\Delta x)} + \ldots +
\end{aligned}
$$

Note that $\mathring{v}_2 = 0 = \mathring{f}$ so that these two terms could be omitted. All series are globally convergent so that Proposition 3.2 applies in fact with $\rho = \infty$. As one can see the abs-polynomial components $\Delta^d f(\mathring{x}; \Delta x)$ have the right degree and order. Moreover, they are telescoping so that we obtain exactly the finite expansions

$$
f(\mathring{x} + \Delta x) = \underbrace{\left| \Delta x_0 - |1 + \Delta x_1| + 1 + \sum_{p=0}^{d-1} \frac{1}{p!}(\Delta x_0)^p \right|}_{\equiv \lozenge^{(d-1)} f(\mathring{x}; \Delta x)} + O(\|\Delta x\|^d)
$$

It should be noted that unless an absolute value $v(\mathring{x} + \Delta x) = |u(\mathring{x} + \Delta x)|$ is zero and thus also its argument $u(\mathring{x} + \Delta x)$ vanishes, the partial sums of latter's absolutely convergent series will have eventually all the same sign s so that $\delta^d v(\mathring{x}; \Delta x) = s \cdot \delta^d u(\mathring{x}; \Delta x)$ for all large d. In that way an extra calculation effort and potential numerical cancelation can be avoided. Graphic representations of the successive approximations are given in Fig. 1. As one can see they converge more or less monotonically right form the start reflecting the essential features in the domain considered. On the following second example one has to take many more terms.

Example 3.4.

$$
y \; = \; f(x_1, x_2) \; = \; \sin(|x_1 - \exp(|x_2 - x_1|)|) \quad \text{at} \quad \mathring{x} = (1, 1)
$$

As one can see in Fig. 3 the combination of the exponential and the sine generate rather large abs-polynomial terms, which have to be cut off to yield informative plots. Only the approximations of order above 40 reproduce the original features reasonably faithfully as one can see in Fig. 3. It is remarkable that in all examples the expansion can *see* across several kinks to approximate function values on the other side. This is of course impossible for a classical Taylor expansion, which has only information about the smooth piece containing the development point. For the abs-polynomial expansion the development point can be located on none or several kinks.

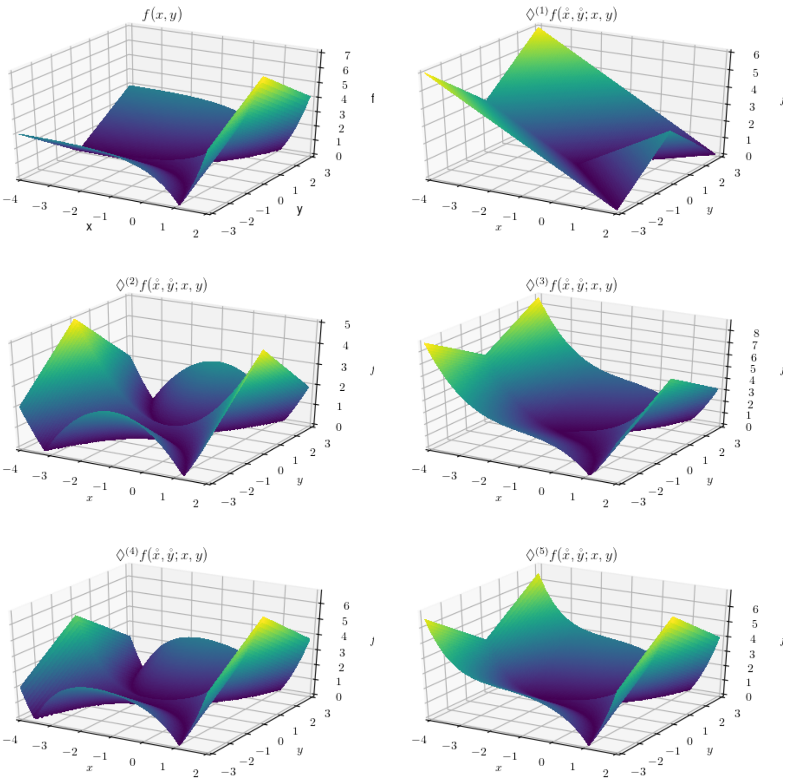
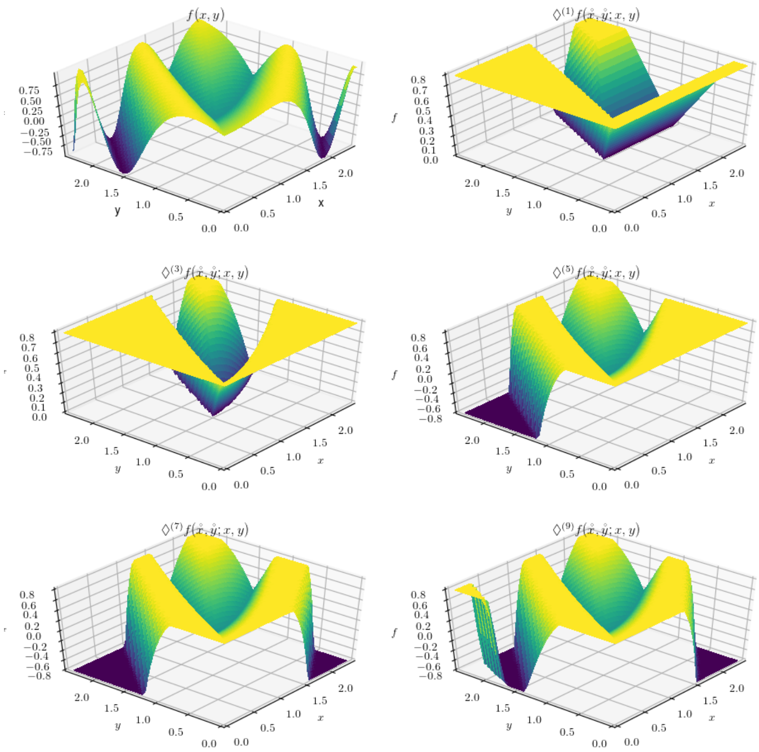


Figure 2. First five abs-polynomial approximations of Example 3.3 with $\Delta x \equiv (x,y)$.



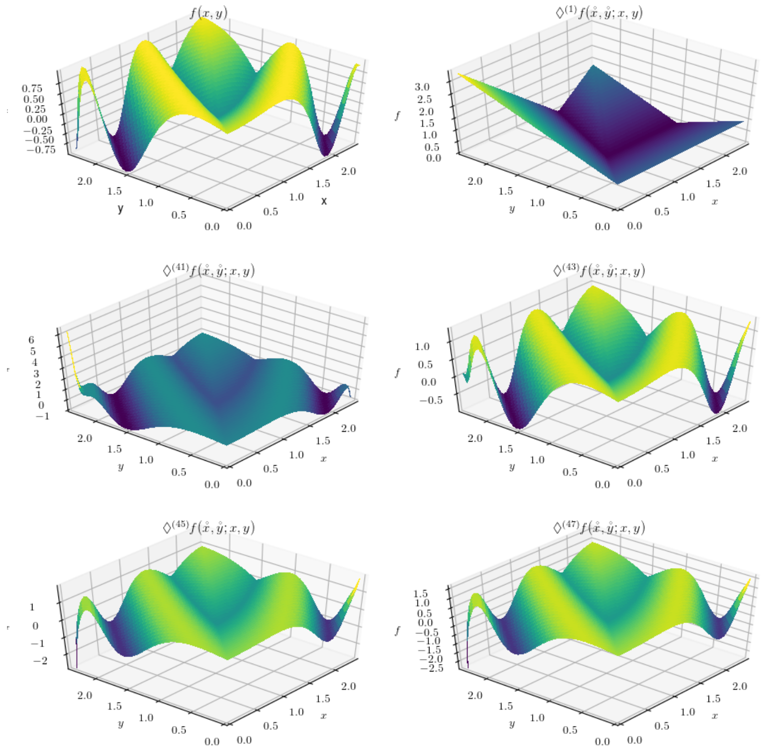


Figure 3. Abs-polynomial approximations of Example 3 with $\Delta x \equiv (x, y)$.

Of course our examples are somewhat special and since the exponential and sine are entire functions the abs-polynomial series converge globally with $\rho = \infty$. One may wonder whether and when it is really useful to compute the abs-polynomial components. In general, one certainly would not want to determine their algebraic form, but only evaluate them at a given increment Δx using the recurrences described in Section 2. Essentially, this is also true in the smooth case where we simply obtain a multi-dimensional Taylor expansion. Then the component functions $\Delta^d f$ are homogeneous higher order polynomials, which are numerically only slightly easier to deal with than general real analytic nonlinear functions. It is not clear whether and when the fact that they are scaled derivatives makes a real difference to their usability.

4. Summary and Conclusion

We have extended the result of [28] to infinite series and shown that for ODE based intrinsics the Moore recurrences are still applicable. Overall we have the somewhat surprising observation that not only smooth but also many or even most piecewise smooth functions can be expanded as absolutely convergent sums of component functions with increasing degree and order at a given reference point in their domain of continuity. Applications of these abs-polynomial expansions remain to be investigated.

Acknowledgement

The authors thank the anonymous referee for a very careful reading of the manuscript.

References

[1] P. Boeck, B. Gompil, A. Griewank, R. Hasenfelder, and N. Strogies. Experiments with generalized midpoint and trapezoidal rules on two nonsmooth ODE's. *Mongolian Mathematical Journal*, 17:39–49, 2013.

[2] T. Bosse, S.H.K. Narayanan, and L. Hascoet. Piecewise linear AD via source transformation. 2016. published as preprint via Argonne National Laboratory.

[3] F. H. Clarke. *Optimization and nonsmooth analysis.* Wiley, New York, 1983.

[4] Tanja Clees, Igor Nikitin, and Lialia Nikitina. Making network solvers globally convergent. In Mohammad S. Obaidat, Tuncer Ören, and Yuri Merkuryev, editors, *Simulation and Modeling Methodologies, Technologies and Applications*, pages 140–153, Cham, 2018. Springer International Publishing. doi:10.1007/978-3-319-69832-8_9.

[5] V. F. Demyanov and A. M. Rubinov. An introduction to quasidifferential calculus. In *Quasidifferentiability and related topics*, volume 43 of *Nonconvex Optim. Appl.*, pages 1–31. Kluwer Acad. Publ., Dordrecht, 2000.

[6] S. Fiege, A. Walther, and A. Griewank. An algorithm for nonsmooth optimization by successive piecewise linearization. *Mathematical Programming*, Apr 2018. doi:10.1007/s10107-018-1273-5.

[7] S. Fiege, A. Walther, K. Kulshreshtha, and A. Griewank. Algorithmic differentiation for piecewise smooth functions: a case study for robust optimization. *Optimization Methods and Software*, 33(4-6):1073–1088, 2018. doi:10.1080/10556788.2017.1333613.

[8] A. Griewank. Automatic Directional Differentiation of Nonsmooth Composite Functions. In Roland Durier, editor, *Recent developments in Optimization / Seventh French-German Conference on Optimization, Dijon 1994*, pages 155–169. Springer Verlag, 1995.

[9] A. Griewank. On stable piecewise linearization and generalized algorithmic differentiation. *Optimization Methods and Software*, 28(6):1139–1178, 2013. doi:10.1080/10556788.2013.796683.

[10] A. Griewank, J.U. Bernt, M. Radons, and T. Streubel. Solving piecewise linear systems in abs-normal form. *Linear Algebra and its Applications*, 471:500 – 530, 2015. doi:https://doi.org/10.1016/j.laa.2014.12.017.

[11] A. Griewank, R. Hasenfelder, M. Radons, L. Lehmann, and T. Streubel. Integrating Lipschitzian dynamical systems using piecewise algorithmic differentiation. *Optimization Methods and Software*, 33(4-6):1089–1107, 2018. doi:10.1080/10556788.2017.1378653.

[12] A. Griewank, T. Streubel, L. Lehmann, M. Radons, and R. Hasenfelder. Piecewise linear secant approximation via algorithmic piecewise differentiation. *Optimization Methods and Software*, 33(4-6):1108–1126, 2018. doi:10.1080/10556788.2017.1387256.

[13] A. Griewank and A. Walther. *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation.* Other Titles in Applied Mathematics. Society for Industrial and Applied Mathematics (SIAM), 2008. doi:10.1137/1.9780898717761.

[14] A. Griewank and A. Walther. First- and second-order optimality conditions for piecewise smooth objective functions. *Optimization Methods and Software*, 31(5):904–930, 2016. doi:10.1080/10556788.2016.1189549.

[15] A. Griewank and A. Walther. Finite convergence of an active signature method to local minima of piecewise linear functions. *Optimization Methods and Software*, 34(5):1035–1055, 2019.

[16] A. Griewank and A. Walther. Relaxing kink qualifications and proving convergence rates in piecewise smooth optimization. *SIAM Journal on Optimization*, 29:262–289, 2019.

[17] A. Griewank, A. Walther, S. Fiege, and T. Bosse. On Lipschitz optimization based on gray-box piecewise linearization. *Mathematical Programming*, 158(1):383–415, Jul 2016. doi:10.1007/s10107-015-0934-x.

[18] D. Klatte and B. Kummer. *Nonsmooth equations in optimization*, volume 60 of *Nonconvex Optimization and its Applications.* Kluwer Academic Publishers, Dordrecht, 2002. Regularity, calculus, methods and applications.

[19] Koichi Kubota. Enumeration of subdifferentials of piecewise linear functions with abs-

normal form. *Optimization Methods and Software*, 33(4-6):1156–1172, 2018. doi:10.1080/10556788.2018.1458848.

[20] R.B. Leipnik and C. E. M. Pearce. The multivariate Fa di Bruno formula and multivariate Taylor expansions with explicit integral remainder term. *The ANZIAM Journal*, 48(3):327341, 2007. doi:10.1017/S1446181100003527.

[21] B. S. Mordukhovich. *Variational analysis and generalized differentiation. I*, volume 330 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2006. Basic theory.

[22] M. Radons. Direct solution of piecewise linear systems. *Theoretical Computer Science*, 626:97 – 109, 2016. doi:https://doi.org/10.1016/j.tcs.2016.02.009.

[23] M. Radons. A note on surjectivity of piecewise affine mappings. *Optimization Letters*, May 2018. doi:10.1007/s11590-018-1271-9.

[24] Louis B. Rall and George F. Corliss. An introduction to automatic differentiation. pages 1–17. SIAM, Philadelphia, PA, 1996.

[25] Winfried Schirotzek. *Nonsmooth analysis*. Universitext. Springer, Berlin, 2007. URL: http://dx.doi.org/10.1007/978-3-540-71333-3, doi:10.1007/978-3-540-71333-3.

[26] S. Scholtes. *Introduction to Piecewise Differentiable Equations*. SpringerBriefs in optimization. Springer New York, 2012. doi:10.1007/978-1-4614-4340-7.

[27] T. Streubel, A. Griewank, M. Radons, and J.U. Bernt. Representation and analysis of piecewise linear functions in abs-normal form. In Christian Pötzsche, Clemens Heuberger, Barbara Kaltenbacher, and Franz Rendl, editors, *System Modeling and Optimization*, pages 327–336, Berlin, Heidelberg, 2014. Springer Berlin Heidelberg. doi:10.1007/978-3-662-45504-3_32.

[28] T. Streubel, C. Tischendorf, and A. Griewank. Piecewise polynomial taylor expansions - the generalization of faà di bruno's formula. Technical Report 18-24, ZIB, Takustr. 7, 14195 Berlin, 2018.

[29] Tom Streubel, Christian Strohm, Philipp Trunschke, and Caren Tischendorf. Generic construction and efficient evaluation of flow network DAEs and their derivatives in the context of gas networks. In N. Kliewer, J.-F. Ehmke, and R. Borndörfer, editors, *Operations Research Proceedings 2017*, pages 627–632, Cham, 2018. Springer International Publishing. doi:10.1007/978-3-319-89920-6_83.

[30] A. Walther and A. Griewank. Characterizing and testing subdifferential regularity in piecewise smooth optimization. *SIAM Journal on Optimization*, 29:1473–1501, 2019.

[31] Charles M. Whish. XXXIII. On the Hind Quadrature of the Circle, and the infinite Series of the proportion of the circumference to the diameter exhibited in the four S'stras, the Tantra Sangraham, Yucti Bhsh, Carana Padhati, and Sadratnamla, July 1834. URL: https://doi.org/10.1017/s0950473700001221, doi:10.1017/s0950473700001221.