On Automatic Differentiation of Codes with COMPLEX Arithmetic with Respect to Real Variables

Christian Bischof Alan Carle
Gordon Pusch

CRPC-TR95549
June 1995

Center for Research on Parallel Computation
Rice University
6100 South Main Street
CRPC - MS 41
Houston, TX 77005

Also available as: *Argonne Technical Memorandum ANL/MCS-TM-188* from Argonne National Laboratory, Mathematics and Computer Science Division.

On Automatic Differentiation of Codes with COMPLEX Arithmetic with Respect to Real Variables[1]

Gordon D. Pusch[2] **Christian Bischof**[3] **Alan Carle**[4]

Argonne Technical Memorandum ANL/MCS-TM-188
CRPC Technical Report CRPC-TR95549

Abstract

We explore what it means to apply automatic differentiation with respect to a set of real variables to codes containing complex arithmetic. That is, both dependent and independent variables with respect to differentiation are real variables, but in order to exploit features of complex mathematics, part of the code is expressed by employing complex arithmetic. We investigate how one can apply automatic differentiation to complex variables if one exploits the homomorphism of the complex numbers $\mathbf{C}$ onto $\mathbf{R}^2$. It turns out that, by and large, the usual rules of differentiation apply, but subtle differences in special cases arise for `sqrt()`, `abs()`, and the power operator.

1 Introduction

In physics and engineering applications, while the underlying independent variables and operations are intrinsically *real-valued,* it is nevertheless often convenient to employ *complex-valued* representations. Applications of complex function theory in applied mathematics, physics, and engineering (and hence presumably in the respective computer codes used) may be broadly grouped into two classes:

1. pairs of real-valued functions of a set of purely real-valued parameters, whose resultants have been combined into a complex number purely for convenience; and

2. functions of pairs of real-valued variables, which for the purposes of computation may be treated as having been combined into a complex number.

[1]This work was supported by the Office of Scientific Computing, U.S. Department of Energy, under Contract W-31-109-Eng-38; by the National Aerospace Agency under Purchase Order L25935D and Cooperative Agreement No. NCCW-0027; and by the National Science Foundation through Cooperative Agreement No. CCR-8809615.

[2]Mathematics and Computer Science Division, Argonne National National Laboratory, Argonne, IL 60439, `pusch@mcs.anl.gov`.

[3]Mathematics and Computer Science Division, Argonne Laboratory, Argonne, IL 60439, `bischof@mcs.anl.gov`.

[4]Center for Research on Parallel Computation, Rice University, 6100 S. Main St., Houston, TX 77251, `carle@cs.rice.edu`.

Class (1) introduces no new rules for automatic differentiation (AD) beyond those of real-valued AD, save that all operations and resultants are declared `COMPLEX` rather than `REAL`; therefore, we discuss it no further. Class (2), however, introduces qualitatively new features. We shall discuss to what extent AD, when applied to problems for which pairs of real-valued variables are formally viewed as forming a single complex number, may be treated as a special case of differentiating maps of $\mathbf{R}^2$ onto itself, and where the peculiarities of complex analysis force exceptions.

Since we will be discussing both mathematical and computational aspects of complex numbers, it will occasionally be necessary to distinguish between a function viewed as a mathematical operator versus its computational equivalent; we shall indicate this by the typeface of the function's name—for example, exp() versus `exp()`.

2 Complex Numbers and Functions

It will be useful to recall a few points about complex numbers and functions from standard complex theory before discussing our "$\mathbf{R}^2$ viewpoint" in detail. For a more detailed exposition, see, for example, [1, 2, 3, 4].

A *complex number* is defined as a quantity of the form $x = a + ib$, where a and b are real numbers, and i is formally defined by $i^2 = -1$. The set of all complex numbers forms the *complex plane,* $\mathbf{C}$.

Since the map $(a, b) \in \mathbf{R}^2 \mapsto x \in \mathbf{C}$ is continuous, invertible, and one-to-one, it follows that $\mathbf{C}$ and $\mathbf{R}^2$ are locally equivalent spaces. However, the global topological structures of $\mathbf{C}$ and $\mathbf{R}^2$ are generally thought of as different. It is frequently convenient to consider infinity to be a point in the *extended complex plane,* $\overline{\mathbf{C}} := \{\mathbf{C} \cup \infty\}$, where infinity is defined to be the formal reciprocal of zero: $\infty := 1/0$. By considering how neighborhoods about zero and infinity transform under the inversion map $z = 1/x$, it can be shown that $\overline{\mathbf{C}}$ must therefore have the topology of a sphere[2, p. 20]; [4, p. 9, 52]. One generally ignores the distinction between $\mathbf{C}$ and $\overline{\mathbf{C}}$ in most applications of complex analysis.[5]

A *complex function of a complex variable* is defined to be a map from some domain U in $\mathbf{C}$ (or $\overline{\mathbf{C}}$) onto some range V in $\mathbf{C}$ (or $\overline{\mathbf{C}}$): $f := \{z = f(x) \mid x \in U \subset \mathbf{C} \mapsto z \in V \subset \mathbf{C}\}$. The definition of the term "function" is often somewhat abused in that the stipulation that a function be single valued is dropped; that is, a complex function is sometimes defined to be a rule that associates a unique *set* of complex values $\{z\}$ to each x in some domain of $\mathbf{C}$, rather than a single unique value. Complex functions may therefore be one-to-many as well as many-to-one. Two examples of such multiple-valued, or *multibranched,* functions are the complex square root and complex logarithm, both of which shall be discussed in detail in §3.

In keeping with our $\mathbf{R}^2$ viewpoint, with every complex function we may associate a corresponding map of $\mathbf{R}^2$ onto itself. However, the converse is not true—not every map of $\mathbf{R}^2$ onto itself may be reinterpreted as a complex-valued function, but only those maps $(c, d) = (f(a,b), g(a,b))$ that

[5] However, this is of course not possible on a finite-precision computer without introducing special arithmetic rules that properly handle points near infinity. While it is possible to define such rules, they would add additional overhead, make less efficient use of any floating-point hardware, and be awkward to implement in a language that does not support operator overloading. Furthermore, while such rules would be mathematically more correct, they would be inconsistent with both the IEEE floating-point and Fortran 77 standards because they would provide meaning to operations that are declared to represent exception conditions under those standards.

satisfy the Cauchy-Riemann (CR) condition:

$$\frac{\partial f}{\partial a} = \frac{\partial g}{\partial b}, \qquad \frac{\partial f}{\partial b} = -\frac{\partial g}{\partial a}. \tag{1}$$

A pair of real functions on $\mathbf{R}^2$ satisfying the CR condition in some neighborhood about (but not including) a point (a, b) is said to be *complex analytic* (or simply *analytic*) in that neighborhood, and the linear combination $h(x) := f(a, b) + i\,g(a, b)$ is said to be a *complex-analytic function* over that neighborhood. It is appropriate to write h as a function of x alone, because the CR condition may be interpreted as stating that $f(a, b) + i\,g(a, b)$ depends on a and b only via the linear combination $x = a + ib$.

The relatively simple-looking condition (1) actually has deep and profound consequences: it is the necessary and sufficient condition for the complex-valued function h to be considered differentiable (in the complex sense) at the point (a, b). Furthermore, it can be shown that if the function is once differentiable at a given point, then it is also infinitely many times differentiable there.

Finally, it can be shown that unless h is everywhere constant, there must be at least one point on the extended complex plane where h has some sort of singularity—and at that singularity, the CR conditions will fail to be satisfied. In fact, it can be shown that any complex function may be completely specified by stating the locations and natures of all of its singularities. Hence, standard complex analysis places signal importance on the study of a function's singularities. (One often studies the *zeros* of a function as well, because the reciprocal of a zero is a singularity.)

A singularity may be either a *pole,* or *essential.* A pole is a singular point x_0 of f such that, for some finite integer n (called the *order* of the pole), the quantity $(x - x_0)^n f(x)$ is nonzero and nonsingular at every point in some open neighborhood about x_0, and its limit as x approaches x_0 (called the *residue* of the pole) exists and is independent of the Cauchy sequence used to approach x_0. Any singularity which is not a pole is an essential one. Furthermore, while the behavior of a function near a pole may be neatly described by the pole's order, the behavior near an essential singularity is always pathological in some way. The only type of essential singularities we shall discuss in detail in this note are branch point singularities (§3). The reciprocal of f at a pole is a zero of the same order, which is not a singularity. However, the reciprocal of an essential singularity is still a point of nonanalyticity.

Because it is foreign to our $\mathbf{R}^2$ viewpoint, we shall make no explicit use of the CR condition in this note; however, our knowledge that its consequences should still hold true will occasionally guide our development.

2.1 Complex Arithmetic Rules

Let $x = a + ib$, $y = c + id$, and $i^2 = -1$. Furthermore, let $\mathrm{Re}\,(z)$ denote the real and $\mathrm{Im}\,(z)$ the imaginary part of a complex number z. Then the results of elementary arithmetic operations have the real and imaginary parts shown in Table 1. The singularities that may result from the arithmetic operations are the union of the singularities of x and y for $x \diamond y$, $\diamond \in \{+, -, *\}$, and the union of the singularities of x with the zeros and essential singularities of y for x/y.

Table 2 shows results of some elementary functions; the properties and singularities of each will be discussed in separate subsections below. In the table, $\overline{x}$ denotes the complex conjugate of x, while $|\,x\,|$ denotes its modulus (neither $\overline{x}$ nor $|\,x\,| = \sqrt{x\overline{x}}$ are complex-analytic operations, because any appearance of $\overline{x}$ causes the CR condition to be violated).

Finally, for analysis purposes, it is useful to note that the singularities of a composite function $p(q(x))$ consist of the union of the singularities of q with the preimage of the singularities of p under q.

Table 1: Complex arithmetic operations

Operation	Re (z)	Im (z)
$z = x + y$	$a + c$	$b + d$
$z = x - y$	$a - c$	$b - d$
$z = x * y$	$ac - bd$	$ad + bc$
$z = 1/x$	$\frac{a}{a^2+b^2}$	$\frac{-b}{a^2+b^2}$
$z = x/y$	$\frac{ac+bd}{c^2+d^2}$	$\frac{bc-ad}{c^2+d^2}$

Table 2: Real and imaginary parts of some elementary functions

Operation	Re (z)	Im (z)
$z = \sqrt{x}$	$\sqrt{\frac{1}{2}(a + \lvert x \rvert)}$	$\sqrt{\frac{1}{2}(-a + \lvert x \rvert)}$
$z = e^x$	$e^a \cos(b)$	$e^a \sin(b)$
$z = \ln(x)$	$\frac{1}{2}\ln(\lvert x \rvert^2)$	$\mathrm{atan}(b/a)$
$z = \overline{x}$	a	$-b$
$z = \lvert x \rvert$	$\sqrt{a^2 + b^2}$	0

2.2 Transcendental Functions

The algebraic extension of real analytic functions to the complex plane is achieved by Taylor series expansion. For example, consider the exponential with a pure imaginary argument:

$$\begin{aligned} \exp(ib) &= \sum_{k=0}^{\infty} \frac{1}{k!}(ib)^k \\ &= \sum_{k=0}^{\infty} \frac{1}{(2k)!}(ib)^{2k} + \sum_{k=0}^{\infty} \frac{1}{(2k+1)!}(ib)^{2k+1} \\ &= \sum_{k=0}^{\infty} \frac{1}{(2k)!} i^{2k} b^{2k} + \sum_{k=0}^{\infty} \frac{1}{(2k+1)!} i^{2k+1} b^{2k+1} \end{aligned}$$

$$= \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k)!} b^{2k} + i \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)!} b^{2k+1}.$$

Comparing the final line above with the expansions for the trigonometric functions cos() and sin(), one obtains Euler's identity,

$$\exp(ib) = \cos(b) + i \sin(b). \tag{2}$$

Consider now the exponential of the complex argument $(a + ib)$: if one assumes the usual rules for exponentiation hold, then

$$\exp(a + ib) = \exp(a)\, \exp(ib). \tag{3}$$

It is straightforward but tedious to verify by Taylor expansion that the left- and right-hand sides of (3) are indeed equal term by term.

From Euler's identity and the symmetry properties of the trigonometric functions, one may derive the identities

$$\cos(b) = \frac{1}{2}(e^{ib} + e^{-ib}), \tag{4}$$

$$\sin(b) = \frac{1}{2i}(e^{ib} - e^{-ib}). \tag{5}$$

From the preceding, one can show that

$$\cosh(b) = \cos(ib), \tag{6}$$

$$i \sinh(b) = \sin(ib), \tag{7}$$

and also that

$$\cos(a + ib) = \cos(a)\cosh(b) + i \sin(a)\sinh(b), \tag{8}$$

$$\sin(a + ib) = \sin(a)\cosh(b) + i \cos(a)\sinh(b). \tag{9}$$

Here sin, cos, and exp are nonsingular everywhere in $\mathbf{C}$, but not in $\overline{\mathbf{C}}$ where they each have essential singularities at infinity.

It is frequently convenient to use the polar coordinate representation of a complex number:

$$x = \rho e^{i\theta} = |\, x \,| \exp(i \arg(x)), \tag{10}$$

where $\rho = |\, x \,| := [a^2 + b^2]^{1/2}$ is called the "modulus" of x, and

$$\theta = \arg(x) := \begin{cases} \operatorname{atan}(y/x), & x > 0, \quad y > 0 \quad \text{(quadrant I)}; \\ \operatorname{atan}(y/x) + \pi, & x < 0, \quad y > 0 \quad \text{(quadrant II)}; \\ \operatorname{atan}(y/x) - \pi, & x < 0, \quad y \leq 0 \quad \text{(quadrant III)}; \\ \operatorname{atan}(y/x), & x > 0, \quad y \leq 0 \quad \text{(quadrant IV)}; \\ \frac{\pi}{2}\operatorname{sign}(y), & x = 0, \quad y \neq 0 \quad (y \text{ axis}); \\ \text{NaN}, & x = 0, \quad y = 0 \quad \text{(origin)}. \end{cases}$$

is called the "argument" or "phase" of x. "NaN" denotes "Not a Number" because in this case the phase is indeterminate—the left- and right-hand sides of (10) are equal no matter what value is chosen for $\arg(x)$.

If one assumes that the usual identity $\ln(\, x\, y\,) = \ln(x) + \ln(y)$ still holds for complex numbers, then from the polar representation we obtain

$$\mathrm{Ln}(x) = \ln(\,|\,x\,|\,) + i\,\arg(x), \tag{11}$$

where the capitalized notation "Ln" indicates that this defines the "principal branch" of the complex logarithm (principal branches of functions will be discussed in more detail in section 3). So, unlike its real cousin, the complex natural logarithm is defined for all $x \in \overline{\mathbf{C}} \backslash \{0, \infty\}$. Its singularities at zero and infinity are *branch points,* a concept which we will discuss in more detail in §3.

In addition to the principal branch of the logarithm, there are an infinite number of additional branches [3, p. 77], [4, p. 71], differing from the principal branch by an additive factor of $2\pi i\, n$, where n is any integer:

$$\ln(x) = \ln(\,|\,x\,|\,) + i\,\arg(x) + 2\pi i\, n, \qquad n = 0, \pm 1, \pm 2, \pm 3, \ldots. \tag{12}$$

This is because, from Euler's identity, $e^{2\pi i\, n} \equiv 1$, so that $e^{\mathrm{Ln}(x)+2\pi i\, n} = e^{\mathrm{Ln}(x)} = x$. Therefore, in the sense that it satisfies the identity $e^{\ln(x)} = x$, $\mathrm{Ln}(x) + 2\pi i\, n$ has as much claim to being "a logarithm of x" as $\mathrm{Ln}(x)$. In other words, the logarithm is an *infinitely many-valued* function.

2.3 SQRT()

Two simple approaches exist for obtaining the complex extension of sqrt(). One is to consider $z = \pm\sqrt{x}$ to be the solution of $z^2 = x \Longrightarrow (c^2 - d^2 + 2icd) = (a + ib)$, which one may formally solve for c and d in terms of a and b. Another simpler approach is to apply the identity $x^{1/2} \equiv \sqrt{x}$ to the polar representation of x. Let $a = \rho\cos(\theta)$, $b = \rho\sin(\theta)$, where $\rho := \,|\,x\,|$, $\theta := \arg(x)$. Then

$$\begin{aligned}
x^{1/2} &= [\rho\exp(i\theta)]^{1/2} = \rho^{1/2}\,[\exp(i\theta)]^{1/2} \\
&= \rho^{1/2}\exp\left[\tfrac{1}{2}i\theta\right] \\
&= \rho^{1/2}\left\{\cos\left[\tfrac{1}{2}\theta\right] + i\sin\left[\tfrac{1}{2}\theta\right]\right\} \\
&= \rho^{1/2}\left\{\sqrt{\frac{[1+\cos(\theta)]}{2}} + i\sqrt{\frac{[1-\cos(\theta)]}{2}}\right\} \\
&= \left\{\sqrt{\tfrac{1}{2}(\rho + a)} + i\sqrt{\tfrac{1}{2}(\rho - a)}\right\}. \qquad (13)
\end{aligned}$$

However expression (13) is not the only solution of $z^2 = x$; its negative solves this equation, also. Hence, sqrt() is a double-valued function; each object point in the x-plane has two image points in the z-plane, $z = \pm\sqrt{x}$, with the $+$ and $-$ signs distinguishing the two branches of sqrt() (we shall discuss branches and branch-point singularities in more detail in §3). Therefore, in terms of its real and imaginary components the complex square-root is

$$\begin{aligned}
\mathrm{Re}\,(z) &= \pm\sqrt{\frac{1}{2}\left[a + \sqrt{a^2+b^2}\right]}, \qquad (14) \\
\mathrm{Im}\,(z) &= \pm\sqrt{\frac{1}{2}\left[-a + \sqrt{a^2+b^2}\right]}, \qquad (15)
\end{aligned}$$

where the $\pm$ signs are to be taken as the same in both equations. It is conventional to choose the $+$ branch as the principal branch.

The only singularities of sqrt() are the origin and infinity, both of which are branch points. That infinity should be a singularity is obvious; however the singularity at the origin is more subtle—while the value of sqrt() is well behaved at this point (it vanishes), the CR condition is violated there.

3 Branch Cuts and Riemann Sheets

Many-valued complex functions are rendered single-valued by the imposition of an artificial boundary called a branch cut [3, p. 117]. The principal branch of the complex logarithm is conventionally obtained by selecting $n = 0$ in (12), and choosing a cut along the negative real axis of the x-plane: $\mathrm{Re}\,(x) < 0$, $\mathrm{Im}\,(x) = 0$. The branch cut is introduced to coerce ln() to be single valued; its image under ln() topologically separates the image plane into inequivalent "sheets," each of which contains only one image of the object point. Each image point represents *a* logarithm of the object point, but generally one refers to the principal-branch value as *the* logarithm of the object point.

It is important to recognize that the location of the branch cut is purely a matter of convention and that nothing unusual happens to the function in question there. It can be shown that it does not matter where one draws the branch cut, so long as it is a non-self-intersecting curve connecting the branch points at the origin and infinity, which are the only true singularities of ln(). The branch cut's function is somewhat analogous to that of the International Date Line in that it represents an arbitrary directed boundary establishing a convention as to where one should consider oneself to have transitioned onto a different sheet. On circumnavigating any closed path in the complex plane, one considers oneself to have moved forward or backward by a number of sheets equal to the net number of branch-cut crossings.

3.1 Ln()

Return now to the principal branch Ln() of the complex logarithm. $z = \mathrm{Ln}(x)$ maps the cut x-plane into the strip $0 \leq \mathrm{Im}\,(z) < 2\pi$. It appears to be discontinuous, jumping from $\ln(\rho) + \pi i$ to $\ln(\rho) - \pi i$ as x crosses the branch cut. Hence, one might fear that the logarithm is nondifferentiable at the cut. However, in reality this is not so, for the principal branch is but a part of the complex logarithm function, and the sheet that it represents patches continuously and smoothly onto the $n = 1$ and $n = -1$ sheets.

3.2 sqrt()

The apparent singularity of sqrt() is more subtle. Suppose one arbitrarily selected the $+$ branch to represent sqrt() as in the real case. Then while both $\mathrm{Re}\,(z)$ and $\mathrm{Im}\,(z)$ are continuous everywhere in the complex plane, at first glance one might think that the derivative of $\mathrm{Re}\,(z)$ would fail to be defined where the argument of the root in (14) vanishes. This will occur if $a = -\rho$, which is true when $a \leq 0$ and $b = 0$. By a similar argument regarding (15), one might also expect that the derivative of $\mathrm{Im}\,(z)$ would fail to be defined if $a \geq 0$, $b = 0$. If true, this would put one in the awkward position of having the derivative of at least one part of sqrt() fail to exist at every point on the real axis.

Fortunately, this singularity in the derivatives turns out to be illusory. The flaw in the preceding argument is that both the $+$ and $-$ branches of sqrt() are required in order to completely describe this function. sqrt() is *intrinsically double valued,* identifying two points of the image complex plane with each point of the object plane.

sqrt() is coerced into being single valued by again introducing an artificial branch cut on the object plane, whose image under sqrt() topologically separates the image plane into two inequivalent sheets, each of which contains only one of the two images of the object point. For sqrt(), the branch cut is conventionally taken to be the *negative real axis.* The $+$ branch of sqrt() maps the entire cut x-plane onto the right half of the z-plane, plus the upper portion of the imaginary axis: $\{\mathrm{Re}\,(z) > 0\} \cup \{\mathrm{Re}\,(z) = 0, \mathrm{Im}\,(z) > 0\}$. However, the $-$ branch of sqrt() also maps that cut x-plane onto the left half of the z-plane, plus the lower portion of the imaginary axis: $\{\mathrm{Re}\,(z) < 0\} \cup \{\mathrm{Re}\,(z) = 0, \mathrm{Im}\,(z) < 0\}$. In other words, for $z = \mathrm{sqrt}(x)$, the preimage of the z-plane is a *double covering* of the x-plane.

Note once more that the location of the branch cut is purely a matter of convention and nothing unusual happens to the derivatives there. Wherever one chooses to draw the branch cut, a careful analysis will show that the $+$ and $-$ branches match together smoothly everywhere along the branch cut except at the branch points at the origin and infinity, which are the only true singularities of sqrt(). Another way of seeing that $\mathrm{sqrt}() \equiv x^{1/2}$ is nonsingular on the real axis is to note that the "power function" x^y to be discussed in §3.3 has no singularity there for $y = 1/2$.

3.3 The Power Function, x^y

A complex number raised to a complex power is defined by the identity $x^y = \exp(\, y \ln(x)\,)$. The function exp() is regular everywhere except infinity; therefore x^y will be singular at the singularities of $y \ln(x)$, which in turn will be singular at the singularities of y or the zeros and singularities of x (because ln() has branch points at both zero and infinity). With no loss of generality, we may assume that any singularities in x or y will have already generated exceptions during their computation. Therefore, it follows that for x^y, the only singularity of interest is the branch point at $x = 0$. Like the logarithm, the power function will in general have infinitely many branches,

$$x^y = \Big\{ \exp\left[\, y\,(\mathrm{Ln}(x) + 2\pi i\, n\,)\right], \;\; n \in \{0, \pm 1, \pm 2, \ldots\} \Big\}. \tag{16}$$

However, if y happens to be real and rational ($y = p/q$ with p and q integers) then x^y has only q inequivalent branches [4, p. 73], because $\exp\left[\,(p/q)(\,2\pi i\, n\,)\right] = \exp\left[\,(p/q)(\,2\pi i\, m\,)\right]$ if $n = m\,(\mathrm{mod}\; q)$. A corollary is that for y a nonzero integer, x^y has only a single branch (since $q = 1$ in this case). The case $y = 0$ is more subtle and will be deferred to the next section.

The cut for x^y is conventionally chosen to be the negative real axis. (The image of the cut plane is a sector subtending $2\pi p/q$ radians if y is real and rational.)

The definitions in subsequent sections will implicitly assume that the principal branch is always chosen. As with sqrt(), there will be an apparent discontinuity in the derivative of the principal branch of x^y. However, it is merely an illusion introduced by the arbitrary imposition of a branch cut.

3.3.1 Asymptotic Behavior of x^y

The singularity relevant to exception conditions for x^y occurs at its branch point $x = 0$. To determine how to properly handle this exception, one should examine the asymptotic behavior of x^y

under various singular limits.

Let $x = a + ib = \rho e^{i\theta}$, and $y = c + id$. The usual rules for exponentiation still apply to complex numbers, so one may write

$$\begin{aligned} x^y &= \left(\rho e^{i\theta}\right)^{(c+id)} = \left(\rho e^{i\theta}\right)^c \left(\rho e^{i\theta}\right)^{id} \\ &= \left(\rho^c e^{ic\theta}\right)\left(\rho^{id} e^{-d\theta}\right) = \rho^c \left(e^{ic\theta} e^{id\ln\rho}\right) e^{-d\theta} \\ &= \rho^c e^{-d\theta} e^{i(c\theta + d\ln\rho)}. \end{aligned} \tag{17}$$

Therefore,

$$|\, x^y \,| = \rho^c e^{-d\theta}, \tag{18}$$

$$\arg(x^y) = c\theta + d\ln\rho. \tag{19}$$

Let us now examine the limiting cases $x \to 0$ and $y \to 0$.

Limit when $x \longrightarrow 0$, $y \neq 0$

It is sufficient to consider $\lim_{\rho\to 0^+} x^y$ with fixed θ and fixed $y = c + id$. One obtains the limiting behaviors shown in Table 3 for the image point as ρ approaches zero from above (where "[C]CW" means "[Counter]Clock-Wise").

Table 3: Limiting behavior of x^y as $\rho = |\, x \,| \to 0^+$

$z = \lim\limits_{\rho\to 0^+} x^y$	$d < 0$	$d = 0$	$d > 0$
$c < 0$	spirals outward CCW to ∞	approaches ∞ along $\texttt{arg}(z) = c\theta$	spirals outward CW to ∞
$c = 0$	circles endlessly CCW at $\texttt{abs}(z) = e^{-d\theta}$	equals $1 + i0$ for all values of $\rho > 0$ and θ	circles endlessly CW at $\texttt{abs}(z) = e^{-d\theta}$
$c > 0$	spirals inward CCW to 0	approaches 0 along $\texttt{arg}(z) = c\theta$	spirals inward CW to 0

Limit when $y \longrightarrow 0$, $x \neq 0$

Similarly, let us consider x^y for $y \to 0$. Let us also assume fixed $x \neq 0$; we shall defer the behavior near the branch point $x = 0$ until the next section.

To take this limit, we introduce a real parameter λ such that $y = \lambda y_0 = \lambda(c_0 + id_0)$, where y_0 is any fixed complex number. Then the limit $|\, y \,| \to 0$ with $\arg(y)$ fixed is equivalent to $\lim_{\lambda\to 0^+}$.

Proceeding much as before, one obtains

$$x^y = \rho^{\lambda c_0} e^{-\lambda d_0 \theta} e^{i\lambda(c_0\theta + d_0 \ln \rho)}.$$

Therefore,

$$\lim_{\lambda\to 0^+} |\, x^{\lambda y_0} \,| = \lim_{\lambda\to 0^+} \left(\rho^{c_0} e^{-d_0\theta}\right)^{\lambda} = 1, \qquad \forall\, \rho > 0, \tag{20}$$

$$\lim_{\lambda\to 0^+} \arg(x^{\lambda y_0}) = \lim_{\lambda\to 0^+} \lambda(c_0\theta + d_0 \ln \rho) = 0, \qquad \forall\, \rho > 0. \tag{21}$$

Note that both limits above are well behaved and independent of x; taken together, they imply that $x^0 = 1 + i0$ for all $x \neq 0$ (consistent with the real result). Therefore it follows that $x \neq 0$, $y = 0$ is not an exception condition—although it may still be advantageous to treat it as a special case, to avoid unnecessarily computing a `log()` and and `exp()`.

Double Limit: Should 0^0 Be NaN?

Since we have shown that $\lim_{|\,x\,|\to 0} x^y = 1 + i0$ for $y = 0$, and also that $\lim_{|\,y\,|\to 0} x^y = 1 + i0$ for $x \neq 0$, no new rules appear to be needed for the case $x \to 0$, $y \to 0$ simultaneously. However, one should not be quite so sanguine about this, as we shall now show.

When evaluating limits of indeterminate forms, one must always exercise caution, particularly when taking multiple limits (since the result may depend on the order in which the limits are taken). We shall see that when we more carefully consider the limit of x^y when $x \to 0$ *and* $y \to 0$, it is not clear that any meaning can be assigned to this form, because

$$\lim_{\lambda\to 0^+} \lim_{|\,x\,|\to 0^+} |\, x^{\lambda y_0} \,| = \lim_{\lambda\to 0^+} \lim_{\rho\to 0^+} \left(\rho^{c_0} e^{d_0\theta}\right)^{\lambda} = \lim_{\lambda\to 0^+} (0)^{\lambda} = 0, \tag{22}$$

while

$$\lim_{|\,x\,|\to 0^+} \lim_{\lambda\to 0^+} |\, x^{\lambda y_0} \,| = \lim_{\rho\to 0^+} \lim_{\lambda\to 0^+} \left(\rho^{c_0} e^{d_0\theta}\right)^{\lambda} = \lim_{\rho\to 0^+} (1) = 1. \tag{23}$$

Since the limits do not commute, even a generalized limit does not exist, suggesting that one should define

$$x^y\big|_{x=0,\, y=0} = \mathrm{NaN}. \tag{24}$$

In summary, the combined results on the limiting behavior of x^y reduce to the following rules for exceptional cases:

$$\boxed{x^y = \begin{cases} 0,^{\dagger} & x = 0,\ \mathrm{Re}\,(y) > 0 \\ \mathrm{NaN}, & x = 0,\ \mathrm{Re}\,(y) = 0,\ \forall\, \mathrm{Im}\,(y) \\ \infty,^{\dagger} & x = 0,\ \mathrm{Re}\,(y) < 0 \\ 1,^{\ddagger} & x \neq 0,\ y = 0 \\ x^y & \text{otherwise.} \end{cases}} \tag{25}$$

It should be noted that while the above are *mathematically* correct in the various limiting cases, they do not necessarily agree with the defined behavior of the F77 intrinsics or IEEE floating-point

standards. In fact, on two workstation platforms using IEEE arithmetic, a test-program yielded NaN for $0 ** y$ independent of the value of y, which disagrees with both cases marked by a † above. The appropriate behavior should, of course, be defined to reproduce the behavior of the original program. It should also be noted that the case marked by ‡ above is not an exception, but merely a special case of the general form x^y.

4 `COMPLEX` Computations in Fortran 77

Table 4 shows the Fortran 77-supported intrinsic generic functions and operations involving type `COMPLEX`. The symbols `I`, `R`, `D`, `C` stand for `INTEGER`, `REAL`, `DOUBLE PRECISION`, and `complex`, respectively. `X` stands for any one of `I`, `R`, `D`; `Z` stands for any one of `I`, `R`, `D`, `C`.

Table 4: Fortran 77 complex functions and operations

Operation	Generic Name	Range and Domain
Type Conversion	`int`	$\mathbf{C} \to \mathbf{I}$
	`real`	$\mathbf{C} \to \mathbf{R}$
	`aimag`	$\mathbf{C} \to \mathbf{R}$
	`dble`	$\mathbf{C} \to \mathbf{D}$
	`cmplx`	$\mathbf{Z} \to \mathbf{C}$
		$\mathbf{X}^2 \to \mathbf{C}$
Arithmetic	{ `+,-,*,/,**` }	$\mathbf{C}\times\mathbf{I} \to \mathbf{C}$
		$\mathbf{I}\times\mathbf{C} \to \mathbf{C}$
		$\mathbf{C}\times\mathbf{R} \to \mathbf{C}$
		$\mathbf{R}\times\mathbf{C} \to \mathbf{C}$
		$\mathbf{C}\times\mathbf{C} \to \mathbf{C}$
Modulus	`abs`	$\mathbf{C} \to \mathbf{R}$
Conjugate	`conjg`	$\mathbf{C} \to \mathbf{C}$
Square Root	`sqrt`	$\mathbf{C} \to \mathbf{R}$
Exponential	`exp`	$\mathbf{C} \to \mathbf{R}$
Natural Log	`log`	$\mathbf{C} \to \mathbf{R}$
Sine	`sin`	$\mathbf{C} \to \mathbf{R}$
Cosine	`cos`	$\mathbf{C} \to \mathbf{R}$

`int()` truncates the real part of a complex argument, coercing it to `INTEGER`, and discards the imaginary part. `real()` ans `aimag()` extract the real and imaginary parts and coerce the result to `REAL`. `DBLE()` extracts the real part, coercing it to `DOUBLE PRECISION`, and discards the imaginary part. `CMPLX` can accept either one or two arguments: if there is a single argument of type `X`, it is converted to `REAL` and assigned to the real part of the `COMPLEX` result while zero is assigned to the imaginary part (a single argument of type `C` is passed through unaltered); whereas if there are two arguments of type `X`, they are each converted to `REAL` and assigned to the real and imaginary parts, respectively (both arguments must be of the same type).

Note that `abs()` is explicitly defined to return a `REAL` result.

Note also that if one operand of a binary arithmetic operator is `COMPLEX`, the standard explicitly prohibits the other operand from being `DOUBLE PRECISION`.

The standard explicitly specifies that "the result of a function of type `COMPLEX` is the principal value." In particular, it states that:

- "The result of `CSQRT` is the principle value with the real part greater than or equal to zero. When the real part of the result is zero, the imaginary part is greater than or equal to zero."
- "The value of the argument of `CLOG` must not be `(0.,0.)`. The range of the imaginary part of the result of `CLOG` is: $-\pi <$ imaginary part $\leq \pi$. The imaginary part is π only when the real part of the argument is less than zero and the imaginary part of the argument is zero."

It also defines `x ** y` to be $exp(\, y \ln(x)\,)$, which implies the power function returns the principal value.

The above facts will be needed to guide our definitions of gradient rules, and also our handling of exception conditions.

5 Complex Differentiation Rules

Again, let $x = a+ib$ and $y = c+id$. The mapping $x \mapsto (a,b) = (\mathrm{Re}\,(x)\,, \mathrm{Im}\,(x))$ is a homomorphism from $\mathbf{C}$ to $\mathbf{R}^2$. Hence, one possibility for complex differentiation would be to rewrite all complex operations explicitly in terms of their real and imaginary parts, converting complex arithmetic into the corresponding $\mathbf{R}^2$ arithmetic. While such an approach would lead to a tool that probably would be useful in other circumstances, it would prevent the utilization of complex arithmetic hardware, would greatly increase the length of the code, and would impair code readability.

The question then is whether the results of differentiation in $\mathbf{R}^2$ can be expressed in terms of complex arithmetic. That is, for a complex variable x, consider its real part $\mathrm{Re}\,(x)$ with its associated derivatives $\nabla\mathrm{Re}\,(x)$, and its imaginary part $\mathrm{Im}\,(x)$ with derivatives $\nabla\mathrm{Im}\,(x)$. An elementary complex operation involving x, and possibly another variable y, defines new values for the real and imaginary parts of the result by the rules set forth in §2.1. We wish to know whether the derivative computations induced by these computations can be easily expressed in complex arithmetic if one adopts the convention that

$$\boxed{\nabla x := (\nabla\mathrm{Re}\,(x)\,, \nabla\mathrm{Im}\,(x)).} \tag{26}$$

5.1 Gradient Rules for the Reciprocal, $1/x$

Let us consider, as an example, $z = 1/x$. We know that $\mathrm{Re}\,(z) = \frac{a}{a^2+b^2}$ and $\mathrm{Im}\,(z) = \frac{-b}{a^2+b^2}$. If we define

$$t = a^2 + b^2, \tag{27}$$

the differentiation rules of real calculus imply that

$$\begin{aligned}
\frac{\partial \mathrm{Re}\,(z)}{\partial a} &= \frac{1}{t} - \frac{2a^2}{t^2}, && (28)\\
\frac{\partial \mathrm{Re}\,(z)}{\partial b} &= -\frac{2ab}{t^2}, && (29)\\
\frac{\partial \mathrm{Im}\,(z)}{\partial a} &= \frac{2ab}{t^2}, && (30)\\
\frac{\partial \mathrm{Im}\,(z)}{\partial b} &= -\frac{1}{t} + \frac{2b^2}{t^2}, && (31)
\end{aligned}$$

and we obtain

$$\begin{aligned}
\nabla \mathrm{Re}\,(z) &= \left(\frac{1}{t} - \frac{2a^2}{t^2}\right) \nabla \mathrm{Re}\,(x) - \frac{2ab}{t^2} \nabla \mathrm{Im}\,(x), && (32)\\
\nabla \mathrm{Im}\,(z) &= \frac{2ab}{t^2} \nabla \mathrm{Re}\,(x) + \left(\frac{2b^2}{t^2} - \frac{1}{t}\right) \nabla \mathrm{Im}\,(x). && (33)
\end{aligned}$$

On the other hand, let us consider

$$w = -\frac{1}{x^2} \nabla x$$

in complex arithmetic. With t defined as in (27), we easily obtain

$$\begin{aligned}
\mathrm{Re}\left(-\frac{1}{x^2}\right) &= \frac{b^2 - a^2}{t^2}, && (34)\\
\mathrm{Im}\left(-\frac{1}{x^2}\right) &= \frac{2ab}{t^2}, && (35)
\end{aligned}$$

and employing the rules of complex multiplication, we obtain

$$\begin{aligned}
\mathrm{Re}\left(-\frac{1}{x^2}\nabla x\right) &= \frac{b^2 - a^2}{t^2} \mathrm{Re}\,(\nabla x) - \frac{2ab}{t^2} \mathrm{Im}\,(\nabla x), && (36)\\
\mathrm{Im}\left(-\frac{1}{x^2}\nabla x\right) &= \frac{2ab}{t^2} \mathrm{Re}\,(\nabla x) + \frac{b^2 - a^2}{t^2} \mathrm{Im}\,(\nabla x). && (37)
\end{aligned}$$

It is easily seen that (32) and (36) as well as (33) and (37) are identical. Hence, the differentiation rules for the real reciprocal also apply to the complex reciprocal.

It is not difficult (albeit tedious) to show in an analogous fashion that the usual differentiation rules apply to complex addition, subtraction, multiplication, division, square root, and exponential.

5.2 Gradient Rules for `ln()`

From the identity $e^{\ln(x)} = x$ one may obtain $e^{\ln(x)}\, d\ln(x)/dx = 1$, which leads to the usual rule for the derivative of the natural logarithm, $d\ln(x)/dx = 1/x$. It is worth noting that this result is clearly finite, single valued, and well behaved everywhere except at the origin, despite the multivalued nature of ln(). Also, it is clear that the derivative shows no evidence of the apparent discontinuity in the principal branch caused by the branch cut on the negative real axis. This discontinuity is purely the result of selecting out only the sheet corresponding to the principal branch for consideration, despite the fact that it is smoothly connected to its neighboring branches. The final gradient rule is

$$\boxed{\nabla\left(\ln(x)\right) = \frac{1}{x}\nabla x, \qquad x \neq 0.} \tag{38}$$

Note that in addition to the function x being nonvanishing at the point of interest, the entire preceding argument implicitly assumes that it is also regular there; that is, it does not have a pole, branch point, or other form of singularity at that point.

5.3 Gradient Rules for SQRT()

To obtain the derivative rules for sqrt(), it is best to start over from first principles. Since the analytic derivative of the complex square root formally looks the same as its real counterpart, we may proceed as follows:

$$\begin{aligned}
\pm\frac{dx^{1/2}}{dx} &= \pm\frac{1}{2}\,x^{-1/2} = \pm\frac{1}{2}\left(\rho e^{i\theta}\right)^{-1/2} && (39)\\
&= \pm\frac{1}{2}\rho^{-1/2}\exp\left[-i\frac{\theta}{2}\right] && (40)\\
&= \pm\frac{1}{2\rho^{1/2}}\left\{\cos\left[\frac{\theta}{2}\right] - i\sin\left[\frac{\theta}{2}\right]\right\} && (41)\\
&= \pm\frac{1}{2\rho^{1/2}}\left\{\sqrt{\frac{[1+\cos(\theta)]}{2}} - i\sqrt{\frac{[1-\cos(\theta)]}{2}}\right\} && (42)\\
&= \pm\frac{1}{2}\left\{\sqrt{\frac{\rho+a}{2\,\rho^2}} - i\sqrt{\frac{\rho-a}{2\,\rho^2}}\right\}; && (43)
\end{aligned}$$

therefore,

$$\begin{aligned}
\mathrm{Re}\left(\pm dx^{1/2}/dx\right) &= \pm\frac{1}{2}\sqrt{\frac{\rho+a}{2\,\rho^2}} && (44)\\
\mathrm{Im}\left(\pm dx^{1/2}/dx\right) &= \mp\frac{1}{2}\sqrt{\frac{\rho-a}{2\,\rho^2}}. && (45)
\end{aligned}$$

The chain-rule formula required for automatic differentiation is

$$\begin{aligned}
\pm\nabla\sqrt{x} &= \pm\frac{dx^{1/2}}{dx}\nabla x && (46)\\
&= \left[\mathrm{Re}\left(dx^{1/2}/dx\right) + i\,\mathrm{Im}\left(dx^{1/2}/dx\right)\right]\left(\nabla a + i\nabla b\right) && (47)
\end{aligned}$$

$$
\begin{aligned}
&= \left[\mathrm{Re}\left(dx^{1/2}/dx\right)\nabla a - \mathrm{Im}\left(dx^{1/2}/dx\right)\nabla b\right] \\
&\quad + i\left[\mathrm{Im}\left(dx^{1/2}/dx\right)\nabla a + \mathrm{Re}\left(dx^{1/2}/dx\right)\nabla b\right]; \qquad (48)
\end{aligned}
$$

therefore,

$$
\mathrm{Re}\left(\nabla x^{1/2}\right) = \pm\frac{1}{2}\left\{\sqrt{\frac{\rho+a}{2\,\rho^2}}\nabla a + \sqrt{\frac{\rho-a}{2\,\rho^2}}\nabla b\right\}, \qquad (49)
$$

$$
\mathrm{Im}\left(\nabla x^{1/2}\right) = \pm\frac{1}{2}\left\{\sqrt{\frac{1}{2}\frac{\rho+a}{\rho^2}}\nabla b - \sqrt{\frac{1}{2}\frac{\rho-a}{\rho^2}}\nabla a\right\}. \qquad (50)
$$

Note that, while the above formulae are everywhere finite and continuous (except at the origin), they again have the same branch-cut-induced *apparent* singularity in their derivatives as before.

5.4 Gradient Rules for the Power Function, x^y

The usual rules for the power operation x^y follow from the logarithm and exponential, via the identity $x^y = \exp(\,y\,\mathrm{Ln}(x)\,)$. By direct differentiation one gets

$$
\frac{\partial x^y}{\partial x} = y\,e^{y\,\mathrm{Ln}(x)}\frac{1}{x} = y\,x^y\frac{1}{x} = y\,x^{(y-1)}, \qquad (51)
$$

$$
\frac{\partial x^y}{\partial y} = \mathrm{Ln}(x)\,e^{y\,\mathrm{Ln}(x)} = \mathrm{Ln}(x)\,x^y. \qquad (52)
$$

Therefore, the appropriate gradient rule is

$$
\boxed{\nabla x^y = y\,x^{(y-1)}\nabla x + \mathrm{Ln}(x)\,x^y\nabla y.} \qquad (53)
$$

However, (53) must be applied with caution, for in the final step in each of (51) and (52) we assumed that we could use the definition of the principal value of the logarithm Ln(), that neither x nor y have singularities at the point of interest, and that $|\,x\,| \neq 0$. If any of the preceding assumptions is invalid, then the final step in each of (51) and (52) is arithmetically invalid, and one cannot proceed to (53). We shall discuss this problem further in §6.3.

6 Exceptional Cases

While the rules of real calculus apply in the cases where the elementary functions in question are well defined, subtle differences exist for exceptional cases. Again, $x = a + ib$ and $y = c + id$.

6.1 Exceptions for ABS()

From the definition of the absolute value of a complex number we readily deduce that $z = \mathrm{abs}(x) = \sqrt{a^2+b^2}$ implies

$$
\boxed{\nabla z = \frac{a\,\nabla a}{\sqrt{a^2+b^2}} + \frac{b\,\nabla b}{\sqrt{a^2+b^2}}} \qquad (54)
$$

and, in particular, $\mathrm{Im}(\nabla z) = 0$. Depending on the curve along which one approaches the origin, one obtains a different directional derivative (consider, for example, the cases $b = 3a$ and $b = 8a$), so that not even a generalized limit (e.g., ∞) exists for $x = 0$. This is because $|\, x\,| = \sqrt{x\overline{x}}$, while everywhere continuous, is nowhere *complex analytic.* It does not satisfy the CR condition, because it depends on the complex conjugate variable $\overline{x}$ as well as x; therefore, its directional derivatives cannot be interpreted as a complex scalar. To attempt to assign a meaning to $\nabla \mathrm{abs}(x)$, one should probably proceed by asserting that `abs()` should be interpreted as a map $\mathbf{C} \rightarrow \mathbf{R}$. *Therefore, the resultant is no longer an element of the complex number field and should not be interpreted as such.* If this approach is taken, then the above gradient should be interpreted as a "conventional" directional derivative, that is, a map

$$\mathbf{C} \xrightarrow{\text{abs}} \mathbf{R} \xrightarrow{\nabla} \mathbf{R}^2.$$

The resultant of this map is therefore an ordinary two-dimensional vector and *not* a complex number anymore. We believe this interpretation is probably the one most often of interest; however, it must only appear as the terminal complex operation, and the result assigned to a `REAL` variable. If `abs()` were to occur as an intermediate step whose result could be acted on by further complex-valued operations, then it is not clear how the derivatives should be propagated or even whether any meaning may be attributed to them, since the complex gradient of a non-complex-analytic function is mathematically meaningless; a strong case can be made that in this circumstance the gradients should not be propagated at all, but rather the expression should be flagged as an error. Unfortunately, nothing in the F77 standard forbids one from constructing such expressions.

It is worth noting at this point that similar problems will be encountered with the functions $\mathrm{Re}()$, $\mathrm{Im}()$, $\arg()$, and the complex-conjugate operation, none of which are by themselves complex-analytic, even though the combinations "$\mathrm{Re}() + i\,\mathrm{Im}()$" and "$\ln(\mathrm{abs}()) + i \arg()$" are analytic. The equivalent FORTRAN 77 intrinsic functions are $\mathrm{Re}()$ = `real()`, $\mathrm{Im}()$ = `imag()`, and the complex-conjugate is `conjg()`; The F77 standard supplies "`abs()`" for complex numbers, but does not supply "`arg()`;" `arg()` may be constructed via the composite-function `imag( log() )`.

6.2 Exceptions from `sqrt()`

If the + (principal) branch is chosen to define `sqrt()`, then the gradients should be propagated by using the partials

$$\boxed{\frac{\partial\sqrt{x}}{\partial x} := \begin{cases} \dfrac{1}{2\sqrt{x}}, & b \neq 0 \ \text{ or } \ a > 0,\, b = 0 \\ \dfrac{+i}{2\sqrt{|\,a\,|}}, & a < 0,\, b = 0 \\ \mathrm{NaN}, & a = 0,\, b = 0. \end{cases}} \tag{55}$$

6.3 Exceptions from the Power Function x^y

Since the power function depends on two variables, computing gradients with respect to either x or y may produce exception conditions. We argue that, as required by complex analyticity theory, the derivatives of the power function will be well defined at any point where the function itself is defined and, conversely, will fail to be defined at any point where the function is undefined.

6.3.1 $\partial x^y/\partial x$

The derivative of x^y with respect to x is naively $y\,x^{(y-1)}$. Therefore, by a similar analysis to Section 3.3, one might conclude that the branch point of this derivative will occur at $x = 0,\ y = 1$, rather than $x = 0,\ y = 0$. If true, this conclusion would violate one of the most fundamental results of complex analyticity theory: that for any neighborhood in which a function is analytic, its derivative is *also* analytic—and therefore must be well defined at any point where the function itself is defined (the converse is also true). Hence, we must go back to the gradient rules (51) and (52) and see what happens when the steps that lead to them are arithmetically invalid.

Restating one of the intermediate steps in (51), we have

$$\frac{\partial x^y}{\partial x} = y\,x^y\,\frac{1}{x}. \tag{56}$$

It is reasonable to assume that x, y, and x^y have already be computed; Hence, the only new factor in the above expression that could generate and exception is $1/x$, which will fail if $|\,x\,| = 0$. However, from the first three clauses of (25), one sees that x^y has already generated an exception if $|\,x\,| = 0$; therefore, no new exception conditions will be generated by $\partial x^y/\partial x$.

In summary, the exceptional cases for $\partial x^y/\partial x$ are

$$\boxed{\frac{\partial x^y}{\partial x} = \begin{cases} 0,\ ^{\dagger} & x = 0,\ \mathrm{Re}(y) > 0 \\ \mathrm{NaN}, & x = 0,\ \mathrm{Re}(y) = 0,\ \forall\,\mathrm{Im}(y) \\ \infty,\ ^{\dagger} & x = 0,\ \mathrm{Re}(y) < 0 \\ 0,\ ^{\ddagger} & x \neq 0,\ y = 0 \\ \dfrac{y\,x^y}{x},\ ^{\S} & \text{otherwise.} \end{cases}} \tag{57}$$

Again, the cases marked † may disagree with the IEEE standards definition of NaN, while ‡ denotes a case that is not an exception but merely a special case of the general form $y\,x^y$. We propose that the specific computational form marked § be used, rather than the algebraically equivalent form $y\,x^{y-1}$, since for both forward and reverse mode the value of x^y should already be available—thus eliminating a `log()` and `exp()` in favor of a division.

6.3.2 $\partial x^y/\partial y$

The naive derivative of x^y with respect to y yields $\partial x^y/\partial y = \ln(x)\,x^y$. In the limit of vanishing x, one obtains

$$\lim_{\rho\to 0^+} |\,\mathrm{Ln}(x)\,x^y\,| = \lim_{\rho\to 0^+} \sqrt{(\ln\rho)^2+\theta^2}\,\rho^c e^{-d\theta} \tag{58}$$

$$= \lim_{\rho\to 0^+} |\ln\rho\,|\sqrt{1+(\theta/\ln\rho)^2}\,\rho^c e^{-d\theta} \tag{59}$$

$$= \begin{cases} \infty, & c \le 0, \\ 0, & c > 0, \end{cases} \tag{60}$$

because a logarithmic singularity is weak enough to be overcome by any positive power-law.

In the limit of vanishing y, one obtains

$$\lim_{\lambda \to 0^+} \left.\frac{\partial x^y}{\partial y}\right|_{y=\lambda y_0} = \lim_{\lambda \to 0^+} \ln(x)\, x^y|_{y=\lambda y_0} = \ln(x), \quad \forall\, x \neq 0, \tag{61}$$

which adds no new conditionals. Therefore, in summary we have

$$\boxed{\frac{\partial(x^y)}{\partial y} = \begin{cases} 0,^{\dagger} & x = 0,\ \ \mathrm{Re}\,(y) > 0,\ \forall\,\mathrm{Im}\,(y) \\ \infty,^{\dagger} & x = 0,\ \ \mathrm{Re}\,(y) \leq 0,\ \forall\,\mathrm{Im}\,(y) \\ \ln(x),^{\ddagger} & x \neq 0,\ \ y = 0 \\ \ln(x)\, x^y & \text{otherwise} \end{cases}} \tag{62}$$

Cases marked † may disagree with the F77 and IEEE standards, while ‡ denotes a case that is not an exception but merely a special case of the general form $\ln(x)\, x^y$.

Acknowledgments

We thank Andreas Griewank and Andrew Mauer for useful discussions, James Lyness for clarifying the nature of the branch point singularity of x^y at $x = 0$, and G. E. Lee-Whiting for his critical reading of this Techmemo.

References

[1] I. Bronstein and K. Semendjajew. *Taschenbuch der Mathematik.* Verlag Harri Deutsch, Frankfurt, Germany, 1979.

[2] Ruel V. Churchill, James W. Brown, and Roger F. Verhey. *Complex Variables and Applications.* McGraw-Hill, 3rd edition, 1974.

[3] Gino Moretti. *Functions of a Complex Variable.* Prentice-Hall, 1964.

[4] Rolf Nevanlinna and V. Paatero. *Introduction to Complex Analysis.* Addison-Wesley, 1969.