DUAL NUMBER CALCULUS

DEFINITIONS

Let $\epsilon^2 = 0$.

Also let $0 < \epsilon < r$ for all real r.

Let $I = \{r\epsilon \ : \ r \in R\}$.

We can also describe I as the set of nilpotent elements in our dual system D.

So $I = \{d \in D : d^2 = 0\}$

So I is the set of infinitesimals.

Note that $0 \in I$. In fact $R \cap I = \{0\}$.

We can also use I to get the infinitesimal part of a dual number. Let $I[a + b\epsilon] = b \in R$.

CONTINUITY

If $f(x + r\epsilon) - f(x) \in I$ for all $r \in R$, then f is **continuous** at x.

DERIVATIVES

If f is continuous, and if $f(x + \epsilon) + f(x - \epsilon) = 2f(x)$, then f is **differentiable** at x.

In other words, if f is continuous and $I[f(x + \epsilon)] = -I[f(x - \epsilon)]$, then f is also differentiable.

If f is differentiable, its **derivative** $f'(x)$ is the d in $f(x + \epsilon) = f(x) + d\epsilon$.

EXAMPLE

Let $f(x) = x^2$.

Then $f(x + r\epsilon) - f(x) = x^2 + 2xr\epsilon + r^2\epsilon^2 - x^2 = 2xr\epsilon \in I$.

So f is continuous at all $x \in R$.

Then $f(x + \epsilon) + f(x - \epsilon) = x^2 + 2x\epsilon + \epsilon^2 + x^2 - 2x\epsilon - \epsilon^2 = 2x^2 = 2f(x)$.

So f is differentiable at all $x \in R$.

Then $f(x + \epsilon) = x^2 + 2x\epsilon + \epsilon^2 = f(x) + 2x\epsilon$.

So $f'(x) = 2x$, as expected from traditional calculus.

Let D be the operator that transforms a function into its derivative.

NOTE

It's easy to check that all polynomials are differentiable.

EXAMPLE

Let $f(x) = |x|$.

Then $f(0 + d\epsilon) = |d\epsilon| \in I$.

So f is continuous. at 0.

But $f(\epsilon) + f(-\epsilon) = |\epsilon| + |-\epsilon| = 2|\epsilon| \neq 0 = 2f(x)$.

So f is not differentiable at 0, which is also true in ordinary calculus.

TRANSCENDENTAL FUNCTIONS

If we use the power series definitions for exp and sin as applied to dual inputs, we can informally confirm that exp is its own derivative, and that cos is the derivative of sin.

We could do dual analysis on these power series to make a more formal case.

We can also approach dual calculus as the attempt to make calculus more algebraic. With this in mind, we might want to avoid limits and just *define* the derivatives of useful transcendental functions. We can then use the chain rule efficiently to find the derivatives of functions composed from these given functions.

So we define, for instance, $\exp(x + d\epsilon) = \exp(x) + d\exp(x)\epsilon$.

EXAMPLE

Let $f(x) = \exp(2x)$.

Then $f(x + d\epsilon) = \exp(2(x + d\epsilon)) = \exp(2x + 2d\epsilon) = \exp(2x) + 2d\exp(2x)\epsilon$.

So $f(x + d\epsilon) - f(x) \in I$ and f is continuous.

It's easy to check that f is differentiable with $f'(x) = 2\exp(2x)$, as expected.

CHAIN RULE

Let f an g be differentiable at x.

Let $h(x) = f(g(x))$.

Then $h(x + d\epsilon) = f(g(x + c\epsilon)) = f(g(x) + b\epsilon) = f(g(x)) + a\epsilon$.

Note that a and b represent the infinitesimal part of differences (possibly 0) implied by the continuity of f and g.

Then $h(x + \epsilon) = f(g(x + \epsilon)) = f(g(x) + g'(x)\epsilon) = f(g(x)) + f'(g(x))g'(x)\epsilon$.

Also $h(x - \epsilon) = f(g(x - \epsilon)) = f(g(x) - g'(x)\epsilon) = f(g(x)) - f'(g(x))g'(x)\epsilon$.

So $I[h(x + \epsilon)] = -I[h(x - \epsilon)]$, and h is differentiable.

Of course $h'(x) = f'(g(x))g'(x)$, as expected.

VECTORS

Partial derivatives are easy with dual numbers. Let $f(x, y) = xy^2$. Then $f_x(x, y) =$

$I[f(x+\epsilon, y)] = I[(x+\epsilon)y^2] = I[y^2\epsilon] = y^2$.

This allows for “forward mode” auto-differentiation. While forward mode isn’t as efficient as backprop, it’s conceptually simpler and easier to implement. And it’s fast enough on modern computers for most problems.

We’ve used dual numbers symbolically so far, but they can be used concretely. For instance, let $f(x) = x^2$. Then $f(2+\epsilon) = (2+\epsilon)^2 = 4 + 4\epsilon + \epsilon^2 = 4 + 4\epsilon$. So we get $f(2) = 4$ and $f'(2) = 4$ through one and the same calculation.

If $f : R^n \to R$ is a loss function, we need n computations of $f(\mathbf{x})$, one for each input, to get the gradient of f at a particular x.

NOTE

Here R ambiguously stands for the set of real numbers. In concrete terms, R is the set of floating point numbers, actually a subset of the rational numbers. We calculate $exp(x)$ on a computer and get a map from floating point rationals to floating point rationals. Enlarged with floating point infinities in both directions, signed zeros, etc.

While the concrete details are more complicated and more efficient, we basically work with rational-coefficient polynomials rather than transcendental functions. In any case, functions from floats to floats stand in for the ideal transcendental functions.

Of course computer algebra systems can work with human symbols for computable reals, but this approach is an exception.

In the floating point context, we can rely on the hardware/software to approximate the ideal of $\exp(x+\epsilon) = \exp(x) + \epsilon \exp(x)$. We can even ensure that the real and infinitesimal parts are equal. But we can’t get the genuine value of $\exp(x)$ unless $x = 0$.

FUN

Julia makes dealing with pure rational (non-floating) numbers easy.

Let $f(x) = x^5 - x^4 + x^3 - x^2 + x - 1$.

Then $f(\frac{234234345}{2342342}) = \frac{6981222852652256682811353964174567427080063}{70510132136973258662751213595232}$.

Also:

$f(\frac{1371742109739369}{109739369}) = \frac{4856935312494678078738906649794560963139470024546496231042506669146000000000}{15915207065345784618237986236670245907849}$.

Rational numbers are beautifully exact, but their “granularity” makes them expensive.