You Could Simplify Calculus*

by Michael Livshits

Cambridge, MA, USA 02138

Email: michaelliv@gmail.com

Web: http://www.mathfoolery.com

September 21, 2009

Abstract

I explain a direct approach to differentiation and integration. Instead of relying on the general notions of real numbers, limits and continuity, we treat functions as the primary objects of our theory, and view differentiation as division of $f(x) - f(a)$ by $x - a$ in a certain class of functions. When f is a polynomial, the division can be carried out explicitly. To see why a polynomial with a positive derivative is increasing (the monotonicity theorem), we use the estimate $|f(x) - f(a) - f'(a)(x-a)| \leqslant K(x-a)^2$. By making it into a definition we arrive at the notion of uniform Lipschitz differentiability (ULD), and see that the derivative of a ULD function is Lipschitz. Taking different moduli of continuity instead of the absolute value, we get different flavors of calculus, each rather elementary, but all together covering the total range of uniformly differentiable functions. Using the class of functions continuous at a, we recapture the classical notion of pointwise differentiability. It turns out that uniform Lipschitz differentiability is equivalent to divisibility of $f(x) - f(a)$ by $x - a$ in the class of Lipschitz functions of two variables, x and a. The same is true for any subadditive modulus of continuity. In this bottom-up, computational, one modulus of continuity at a time approach to calculus, the monotonicity theorem takes the central stage and provides the aspects of the subject that are important for practical applications. The weighty ontological issues of compactness and completeness can be treated lightly or postponed, since they are hardly used in this streamlined approach that pretty much follows the Vladimir Arnold's "principle of minimal generality, according to which every idea should first be understood in the simplest situation; only then can the method developed be applied to more complicated cases." I discuss a generalization to many variables briefly.

1 Two Stories, One Fictional, One Real

1.1 Differentiating x^4 without using limits

A teacher asks a student to calculate the derivative of x^4 at $x = a$. The student writes down the difference quotient $\frac{x^4 - a^4}{x - a}$, then, by factoring the numerator, rewrites it as $\frac{(x-a)(x+a)(x^2+a^2)}{x-a}$, then cancels $x - a$, substitutes $x = a$ into the result and gets $4a^3$, that is the right answer, of course. The teacher does not like the solution, and the following conversation takes place.

*. This document has been written using the GNU $\mathrm{T_EX_{MACS}}$ text editor (see www.texmacs.org).

T: Your answer is correct, but why didn't you use the definition of the derivative as a limit? We are studying calculus here, you know.

S: Do I really need to use limits? It looks like a waste of time, I can just simplify and plug in $x=a$ instead, it looks like it works fine.

T: But do you understand why it works?

S: Hmmm, let me see. I guess it works because the limit of $(x+a)(x^2+a^2)$ as $x \to a$ is $4a^3$, so, instead of calculating the limit we can just plug $x=a$ into $(x+a)(x^2+a^2)$.

T: How do you call such a function, that you can just plug in $x=a$ into it instead of calculating the limit of this function at a?

S: Continuous at a? Yeah, I remember.

T: Right! You know, people differentiated polynomials, roots and trig functions in the 17th century, long before they started thinking of such generalities as continuity and limits in the 19th century. Why don't you try to differentiate your way some other simple algebraic expressions, such as $\sqrt[3]{x}$ or $\frac{x^2}{3+x^3}$?

S: O.K., I will, I think I understand it a little better now.

1.2 Differentiating $\sqrt{x}$ without using limits

It happened in the fall of 1997, when I taught two calculus recitation sections at Suffolk University. The purpose of these sections was to answer the questions the students had about their homework and the subject in general. The text was Anton's Calculus, which I came to hate as the semester progressed.

It was one of the classes, and some students asked me to explain how to differentiate $\sqrt{x}$. So I wrote down the difference quotient $\frac{\sqrt{x}-\sqrt{a}}{x-a}$ on the chalkboard and said that we had to calculate the limit of this expression as x approaches a.

As soon as I uttered the word "limit" I saw many students slump in their seats, their eyes glazing over, and I had the sinking feeling that they were totally lost. I had to do something fast to help them, to pull them out of their despair, but what?

I said, look, you don't really need limits to calculate this derivative, you can do it algebraically. Let us rewrite this expression in such a way that it would make sense for $x=a$. How can we do that? Let us rewrite the denominator as $\sqrt{x}^{\,2} - \sqrt{a}^{\,2}$ and factor it as $(\sqrt{x} - \sqrt{a})(\sqrt{x} + \sqrt{a})$, so $\frac{\sqrt{x}-\sqrt{a}}{x-a} = \frac{\sqrt{x}-\sqrt{a}}{(\sqrt{x}-\sqrt{a})(\sqrt{x}+\sqrt{a})} = \frac{1}{\sqrt{x}+\sqrt{a}}$ that makes sense for $x=a$, giving us the answer $(\sqrt{x})' = 1/(2\sqrt{x})$, that's all there is to it.

I saw the students brightening up a little bit, when they realized that the problem could be solved with the tools familiar to them. And that's exactly when it dawned on me that all calculus could be done like that, differentiation being nothing but division in the class of continuous functions. It surely looked like a promising idea.

2 Calculus of Polynomials

2.1 Formal differentiation

Let us start with the simplest and most popular example, differentiating x^2. We form the difference quotient $\frac{x^2-a^2}{x-a}$ and try to make sense of it for $x=a$. The trouble is, of course, that when we just plug in $x=a$, we get $0/0$, which is undefined, because $0c=0$ for any number c. But luckily, the numerator factors as $(x-a)(x+a)$, so we can cancel $x-a$ and rewrite our expression as $x+a$ that makes sense for $x=a$, giving us $(x^2)'=2x$. To generalize to x^n, we use the factorization $x^n-a^n=(x-a)(x^{n-1}+x^{n-2}a+\ldots.+a^{n-1})$ to get $(x^n)'=nx^{n-1}$.

This trick will work for any polynomial $p(x)$, because a is a root of the polynomial $p(x)-p(a)$, and therefore it is divisible by $x-a$, so we have $p(x)-p(a)=(x-a)q(x,a)$, and we can rewrite $\frac{p(x)-p(a)}{x-a}$ as $q(x,a)$ which is a polynomial in x and a and therefore makes sense for $x=a$, giving us $p'(x)$.

Of course we don't have to divide polynomials every time we differentiate them. The first two *differentiation rules* tell us that $(f+g)'=f'+g'$ and $(kf)'=kf'$ for any constant k, in other words, differentiation is a linear operation, and therefore we can differentiate polynomials "term by term," i.e.

$$(p_0+p_1x+\ldots+p_nx^n)'=p_1+2p_2x+\ldots+np_nx^{n-1}.$$

The other two rules of differentiation, the *product* (or *Leibniz*) rule, saying that $(fg)'=f'g+fg'$ and the *chain* rule by Newton, $(f(g(x)))'=f'(g(x))g'(x)$ are a matter of algebra of polynomials.

The trick developed here can be used to differentiate all rational functions, and even algebraic functions that are defined implicitly by algebraic equations, if we use implicit differentiation.

2.2 Double roots and the basic estimate

Consider a polynomial $p(x)$. The question is: "why the *tangent* to the graph $y=p(x)$ at the point $(a,p(a))$, which is the line defined by the equation $y=p(a)+p'(a)(x-a)$ looks like a tangent, i.e. "clings" to this graph?" Let us start with a simple example, $p(x)=x^k$. Then $p'(a)=ka^{k-1}$, and $x^k-a^k-ka^{k-1}(x-a)=(x-a)(x^{k-1}+x^{k-2}a+\ldots+a^{k-1}-ka^{k-1})=(x-a)^2r(x,a)$, with r a polynomial in x and a, because the second factor vanishes for $x=a$, so it is divisible by $x-a$. A similar factoring, $p(x)-p(a)-p'(a)(x-a)=(x-a)^2r(x,a)$, holds for any polynomial p since it is a sum of monomials. It shows that $x=a$ is a double root of the equation $p(x)-p(a)-p'(a)(x-a)=0$. This fact can be taken as the definition of a tangent to a graph of a polynomial, and can be used to define the derivative for polynomials. The vertical distance $d(x,a)$ between the graph and the tangent can be written as $(x-a)^2|r(x,a)|$, with r a polynomial in x and a. When x and a are contained in some finite interval, $|r(x,a)|$ will be bounded from above by some constant K, giving us an estimate $d(x,a)\leqslant K(x-a)^2$. This *basic estimate*, that also can be written as

$$|p(x)-p(a)-p'(a)(x-a)|\leqslant K(x-a)^2 \tag{1}$$

holds for any polynomial p, and explains why tangents clings the graphs. We will use it in the next subsection to understand why a polynomial with a positive derivative is increasing.

2.3 Monotonicity principle

The derivative is a mathematical metaphor for the instantaneous velocity, or the instantaneous rate of change of a function relative to its argument. So we would expect that a function with positive derivative would be increasing. Let us see why it is true for polynomials. Assume that $p'(x) \geqslant 0$ for any x such that $A \leqslant x \leqslant B$. We want to show that $p(A) \leqslant p(B)$. We can deal with a simpler case $p'(x) \geqslant C > 0$ first. Our basic estimate (1) tells us that $p(x) - p(a) \geqslant p'(a)(x - a) - K(x-a)^2$, so $p(a) \leqslant p(x)$ if $0 < x - a \leqslant C/K$. Therefore, $p(A) \leqslant p(B)$, since we can get from A to B by steps shorter than C/K. To get to the original assumption, we can consider $q(x) = p(x) + Cx$ with $C > 0$ and conclude that $p(B) - p(A) \geqslant C(A - B)$, therefore $p(A) \leqslant p(B)$ since C is arbitrary.

By applying our monotonicity principle to $f + Mx$ and $f - Mx$, we can demonstrate *the rule of bounded change:*

If $|p'| \leqslant M$, then $|p(x) - p(a)| \leqslant M|x - a|$.

When we look at definite integrals as increments of anti-derivatives, we can see how monotonicity is related to positivity of the area.

Also we need the basic estimate (1) only for small enough $|x - a|$ to show the monotonicity principle.

2.4 Formal integration

It can be introduced before the basic estimate is treated and monotonicity theorem is demonstrated, and it is very easy for polynomials. Besides, it provides a strong evidence for the Newton-Leibniz theorem. The simplest examples of course are the constants and the linear functions. A bit more work is required to calculate the areas under the other power curves, and may give the skeptics an opportunity to use such tools as algebra, the geometric series and even combinatorics (to estimate the sum $1^k + 2^k + \ldots + n^k$). Newton-Leibniz is very intuitive and can be explained early on. The integration rules are just the rules of differentiation, rewritten in terms of integrals. This formal theory can be used right away to solve some interesting problems in geometry and physics.

3 Uniform Lipschitz Calculus

How can we extend our calculus to functions more general than polynomials? As it often happens in mathematics, we just look at some useful property or a formula and make it into a definition (think about the Pythagorean Theorem). The useful property here will be the basic estimate (1) from section 2.2, so we call a function f *uniformly Lipschitz differentiable* (ULD) if the estimate

$$|f(x) - f(a) - f'(a)(x - a)| \leqslant K(x - a)^2 \tag{2}$$

holds for some constant K independent of x and a. The monotonicity principle from section 2.3 extends to ULD functions.

3.1 The automatic Lipschitz estimate for the derivative

We know that the derivatives of polynomials are also polynomials. What would be the analogous fact for ULD functions? It turns out that theis derivatives are *Lipschitz*, i.e., they satisfy the estimate $|f'(x) - f'(a)| \leqslant L|x-a|$ with L independent of x and a.

To see it, we notice that for $x \neq a$ $|\frac{f(x)-f(a)}{x-a} - f'(a)| \leqslant K|x-a|$. By interchanging x and a we get $|\frac{f(a)-f(x)}{a-x} - f'(x)| \leqslant K|a-x|$. but $\frac{f(a)-f(x)}{a-x} = \frac{f(x)-f(a)}{x-a}$, and we see that $|f'(x) - f'(a)| \leqslant 2K|x-a|$, i.e., f' is Lipschitz.

Of course all the polynomials are Lipschitz on any finite interval, because $x-a$ is a factor in $p(x) - p(a)$, and the ULD functions are too, because their derivatives are bounded on any finite interval, and we get $|f(x) - f(a)| \leqslant M|x-a|$ from the rule of bounded change. As we will see later (for general moduli of continuity), the analogy runs even deeper, and in fact differentiation of ULD functions is related to factoring in the class of Lipschitz functions the same way as differentiation of polynomials is related to their factoring.

3.1.1 A comparison with the non-standard analysis approach

In non-standard analysis the derivatives of functions differentiable on a hyperreal interval are automatically continuous, the proof goes the same way, except we say that $\frac{f(x)-f(a)}{x-a} - f'(a)$ and $\frac{f(a)-f(x)}{a-x} - f'(x)$ are infinitely small when $x-a$ is, and conclude that in this case $f'(x) - f'(a)$ is infinitely small too. It is this fact that makes the non-standard approach to calculus simple. More generally, many pointwise estimates on a hyperreal interval are in fact uniform. In uniform differentiation theory we work with uniform estimates directly and get the results much cheaper, without any infinitesimals that are not constructive. See http://en.wikipedia.org/wiki/Hyperreal_number where I wrote a section "An intuitive approach to the ultrapower construction" and references there.

3.2 Integration, existence of a primitive and Newton-Leibniz

It is easy to integrate polynomials and rational functions since antiderivatives can be written down explicitly in terms of the elementary functions, but this situation is rather exceptional. Now, we know that the derivative of any ULD function is Lipschiz, and we ask if an antiderivative exists for any Lipschitz function, in what sense it exists, and how it can be calculated. The idea is to define the definite integral as the area under the graph and then to make sense out of the notion of the area by constructing explicit approximations (pretty much following the approach of the Greeks, later developed by Riemann, Darboux, Jordan, and Lebesque), and then prove Newton-Leibniz. The case of Lipschitz, and other uniformly continuous functions, is particularly simple, and requires hardly any sophistication. A picture (that is worth a 1000 words) is available on page 13 at http://www.mathfoolery.com/talk-2004.pdf and pages 43-44 at http://www.mathfoolery.com/lathead.pdf with a proof of Newton-Leibniz.

3.2.1 Approximation of integrals and calculus cartoon by Lin Qun

4 Other Moduli of Continuity

Sometimes calculus based on Lipschitz estimates is too restrictive, for example, the function $x^{3/2}$ has $\sqrt{x}$ for the derivative, which is not Lipschiz, since it grows too fast near $x=0$. To treat this function as differentiable, we can relax the estimate (2) defining differentiability to $|f(x)-f(a)-f'(a)(x-a)|\leqslant K|x-a|^{3/2}$. More generally, we can use the inequality

$$|f(x)-f(a)-f'(a)(x-a)|\leqslant K|x-a|m(|x-a|)$$

with some *modulus of continuity* m to define m*-differentiability*, $m(x)=\sqrt{x}$ is an example, for any positive $\gamma\leqslant 1$, x^γ is a more general example, the corresponding differentiability is called *uniform Holder*, with the exponent γ and the corresponding derivatives are *Holder*, i.e., $|f'(x)-f'(a)|\leqslant H|x-a|^\gamma$ holds. In general, we want m to be defined for $x\geqslant 0$, an increasing, continuous at 0, $m(0)=0$, and *subadditive*, i.e., $m(x+y)\leqslant m(x)+m(y)$. All the Lipschitz theory extends to the general moduli of continuity with some obvious modifications, the derivatives are m-continuous, i.e., $|f'(x)-f'(a)|\leqslant 2Km(|x-a|)$ etc.

4.1 An estimate of the difference quotient

Let m be a subadditive modulus of continuity, in particular, m is increasing, defined for $x\geqslant 0$, and $m(x)/x$ is decreasing for $x>0$, and let f be a uniformly m - differentiable function, i.e. there is a uniform in x and a estimate with some constant K:

$$|f(x)-f(a)-f'(a)(x-a)|\leqslant K|x-a|m(|x-a|) \tag{3}$$

Let the difference quotient for f be the 2-variable function

$$Q_f(x,a)=(f(x)-f(a))/(x-a)\,\text{for}\,x\neq a\,\text{and}\,Q_f(x,x)=f'(x).$$

We want to demonstrate the inequality

$$|Q_f(x,a)-Q_f(y,a)|\leqslant 2Km(|x-y|), \tag{4}$$

that means that the difference quotient is a uniformly m-continuous. It allows us to view m-differentiation as factoring $f(x)-f(a)=Q_f(x,a)(x-a)$ in a class of m-continuous functions of 2 variables. This idea (in the context of general uniformly continuous functions) is used by Mark Bridger in his innovative text Real Analysis: A Constructive Approach.

Because only the increments of the independent variable and the corresponding increments of the values of f are involved in the difference quotient, we can assume $a=0=f(0)$ and the inequality we want becomes

$$|f(x)/x-f(y)/y|\leqslant 2Km(|x-y|). \tag{5}$$

The case $x < 0 < y$ is easy because $|f(x)/x - f'(0)| \leqslant Km(|x|)$ and $|f(y)/y - f'(0)| \leqslant Km(|y|)$, so $|f(x)/x - f(y)/y)| \leqslant K(m(|x|) + m(|y|)) \leqslant 2Km(|x-y|)$ because m is increasing.

The case of x and y of the same sign, say, $0 < x < y$ is a bit more delicate. First we notice that adding any linear function to f does not change $f(x)/x - f(y)/y$, so we can assume that $f'(x) = 0$. The left-hand side of the inequality we want to prove can be rewritten as $|((y-x)f(x) - x(f(y) - f(x)))/(xy)|$. Now, $|f(y) - f(x)| \leqslant K(y-x)m(y-x)$ because $f'(x) = 0$, and also $|f(x)| \leqslant Kxm(x)$ because $f(0) = 0$. So it is enough to show that $\frac{y-x}{y}m(x) \leqslant m(y-x)$. Again the case $x \leqslant y - x$ is easy because m is increasing. We only have to use subadditivity of m when $y - x \leqslant x$. In this case $m(x)/y \leqslant m(x)/x \leqslant m(y-x)/(y-x)$ and we are done.

4.1.1 A simpler proof using integration

We can use integration to get the uniform m-continuity estimate (4) in a simpler and more conceptual manner. We have $f(x) - f(a) = \int_a^x f'(u)du$, and therefore $Q_f(x, a) = (\int_a^x f'(u)du)/(x-a)$, which is the average value of f' on the interval $[a, x]$. This average can be also written as $\int_0^1 f'(a + t(x-a))dt$. Now, the left-hand side of the inequality (4) is the absolute value of the average of $f'(a + t(x-a)) - f'(a + t(y-a))$ over the interval $[0, 1]$ in variable t. Now, by automatic m-contimuity of f', we get $|f'(a + t(x-a)) - f'(a + t(y-a))| \leqslant 2Km(|a + t(x-a) - a - t(y-a)|) = 2Km(t|x-y|) \leqslant 2Km(|x-y|)$, the last inequality holds because m is increasing, and we are done. Notice that we did not use subadditivity of m, only its monotonicity.

4.2 Epsilon-delta and moduli of continuity

We used different moduli of continuity to describe uniform continuity and differentiability. The question is: "how much of the classical theory of continuous and smooth functions do we miss, if any?" The answer to this question is "nothing." Let us consider uniform continuity of a function defined on a finite interval, uniform differentiability is analogous.

The classical way to describe uniform continuity of a function f is to say that for any $\varepsilon > 0$ there is $\delta > 0$ such that $|f(x) - f(a)| < \varepsilon$ when $|x - a| < \delta$.

We want to show that there is a modulus of continuity m, such that the inequality $|f(x) - f(a)| \leqslant m(|x-a|)$ holds. Let us consider the following function: $g(h) = \sup\{|f(x) - f(a)| : |x - a| \leqslant h\}$. We know that g will be positive, increasing, and $g(h) \longrightarrow 0$ as $h \longrightarrow 0$, so g will become continuous at 0 if we put $g(0) = 0$. Now, consider the set $\{(y, h)\colon y \leqslant g(h),\ 0 \leqslant y,\ 0 \leqslant h \leqslant A\}$ of points under the graph of g in the first quadrant of the (h, y) plane. Take the convex hull of this set. The upper edge of this convex hull will be the graph of a concave (and therefore subadditive) modulus of continuity for f.

It is needless to say that in some questions (such as topological classification of dynamical systems) keeping track of the particular moduli of continuity may be a nuisance, and not fruitful. Then we can throw all the uniformly continuous or uniformly differentiable functions into one big pile and enjoy the generality.

4.2.1 Continuity of increasing functions: an intuitive approach

4.2.2 A minimalistic theory of the reals

5 Series and Sequences of Functions

5.1 Power series: term-by-term differentiation

5.2 Term-by-term integration of sequences of functions

6 Some Pedagogical implications

6.1 Calculus by problem solving: a still unrealized dream

http://www.mathfoolery.com/Problem_sets/hw.html

6.2 Suggested curricula

7 Many Variables

7.1 Differentiability

Similar to the case of one variable, we define differentiability by the inequality

$$|f(x+h)-f(x)-f'(x)h| \leqslant K|h|m(|h|).$$

Here $|.|$ denotes some norm, for example, the Euclidean norm, $f'(x)$ is a linear map depending on x, K is a constant and m is a modulus of continuity.

7.2 Automatic continuity of the derivative

We want to show that the uniform derivative is uniformly continuous with the modulus of continuity m from the definition, i.e., the inequality

$$|f'(x+h)-f'(x)| \leqslant Lm(|h|)$$

holds for some constant L that will depend on K in the definition. Here $|.|$ is the norm of the linear operators, $|A| = \sup\{|Ak|, |k| = 1\}$.

The idea of the simplest proof I could come up with is the following. There are two ways to get from x to $x + h + k$. We can go directly, or we can go from x to $x + h$ first and then from $x + h$ to $x + h + k$. The corresponding increments of the function f should be the same. Now consider the approximation of these increments by the differentials.

$$|f(x+h) - f(x) - f'(x)h| \leqslant K|h|m(|h|)$$

$$|f(x+h+k) - f(x+h) - f'(x+h)k| \leqslant K|k|m(|k|)$$

$$|-f(x+h+k) + f(x) + f'(x)(h+k)| \leqslant K|h+k|m(|h+k|)$$

By "adding" all of these inequalities and using the triangle inequality, $|a + b| \leqslant |a| + |b|$, and linearity, $f'(x)(h+k) = f'(x)h + f'(x)k$, we conclude that

$$|f'(x)k - f'(x+h)k| \leqslant K(|h|m(|h|) + |k|m(|k|) + |h+k|m(|h+k|)).$$

But $|h + k| \leqslant |h| + |k|$ and $m(|h + k|) \leqslant m(|h| + |k|) \leqslant m(|h|) + m(|k|)$ (triangle, m is increasing and subadditive). Finally, by taking $|k| = |h|$, we get

$$|(f'(x+h) - f'(x))k| = |f'(x)k - f'(x+h)k| \leqslant 6Km(|h|)|k|$$

that means that $|f'(x+h) - f'(x)| \leqslant 6Km(|h|)$, so we can take $L = 6K$. Done.

7.3 Differentiation as factoring

How the idea from section 4, that uniform differentiation can be viewed as factoring of functions of 2 variables, can be extended to the case of many variables? The difficulty is that we can no longer divide by $x - a$ to define the difference quotient $Q_f(x, a)$ for $x \neq a$. To make things worse, even if there is a factoring like $f(x) - f(a) = P(x, a)(x - a)$, the first factor is not uniquely defined, since its value can be changed by any covector orthogonal to $x - a$ without changing the product, so the whole idea looks a bit shaky. Novertheless, it survives.

First, to see that factoring implies differentiability, assume that P is m-continuous in x and a. Then

$$|f(x) - f(a) - P(a,a)(x-a)| = |(P(x,a) - P(a,a))(x-a)| \leqslant Lm(|x-a|)|x-a|$$

and f is uniformly m-differentiable, with $f'(x) = P(x, x)$.

Now we can define the difference quotient as the average value of f' over the segment $[a, x]$. The argument that $Q_f(x, a)$ is m-continuous from subsubsection 4.1.1 still applies. Notice that the difference quotient defined this way is a symmetric function of x and a, i.e., $Q_f(a, x) = Q_f(x, a)$, as in the one variable case.

What is so attractive about the idea of differentiation as factoring? It brings calculus into the realm of algebra and helps to keep away infinitesimals, limits and other ghosts of departed quantities. It also makes differentiation rules very natural. Take, for example, the chain rule.

7.3.1 Chain rule

We want to calculate the derivative of of composition $f \circ g$ of two functions, assuming they are both m-differentiable. We have factoring $f(g(x)) - f(g(a)) = Q_f(g(x), g(a))(g(x) - g(a)) = Q_f(g(x), g(a))Q_g(x, a)(x - a)$. By taking $a = x$ we get $(f \circ g)'(x) = f'(g(x))g'(x)$ it works for many variables as well as for one. We only have to check that $Q_f(g(x), g(a))Q_g(x, a)$ is m-continuous.

We can do it in several easy steps, that are of interest in their own right.

First we notice that m-continuous functions are bounded in any bounded region. This implies that a product of two m-continuous functions is m-continuous. Now, since g' is bounded, g is Lipschitz in any bounded region by the law of bounded change. Finally, a composition of an m-continuous function with a Lipschitz function is m-continuous.

The reasoning simplifies somewhat if we use general uniformly continuous functions since we stop keeping track of the specific moduli of continuity.

7.4 Equality of the mixed derivatives

Probably the simplest way to understand why $f_{xy} = f_{yx}$ is to use the Green's formula. Here is how. Let us consider a rectangle $ABCD$ on the (x, y) – plane where f is defined. $A = (a, c)$, $B = (b, c)$, $C = (b, d)$ and $D = (a, d)$. There are two ways to get from A to C. We can go from A to B and then from B to C, or we can go from A to D and then from D to C. The total change in f should be the same for both ways. Let us write down this change in terms of the line integrals of the partial derivatives.

$$f(A) - f(C) = f(B) - f(A) + f(C) - f(B) = \int_a^b f_x(x, c)dx + \int_c^d f_y(b, y)dy$$

and on the other hand,

$$f(A) - f(C) = f(D) - f(A) + f(C) - f(D) = \int_c^d f_y(a, y)dy + \int_a^b f_x(x, d)dx$$

so we have

$$\int_a^b (f_x(x, d) - f_x(x, c))dx - \int_c^d (f_y(b, y) - f_y(a, y))dy = 0,$$

but $f_x(x, d) - f_x(x, c) = \int_c^d f_{xy}(x, y)dy$ and $f_y(b, y) - f_y(a, y) = \int_a^b f_{yx}(x, y)dx$ and, replacing the iterated integrals with the double integrals, we conclude that $\int_{ABCD} (f_{xy} - f_{yx})dxdy = 0$ for any rectangle $ABCD$. It is only possible if $f_{xy} - f_{yx} = 0$, so $f_{xy} = f_{yx}$ and we are done.