Quick calculus

Daniel O'Connor

These notes are intended to be a quick summary of some of the key *intuition* behind calculus. The notes are not self-contained and are meant only to supplement a calculus class, not to stand alone. Moreover, the notes are a work in progress. If you have any questions or suggestions please feel free to email me at daniel.v.oconnor@gmail.com.

Contents

1 Introduction

The purpose of these notes is not to give rigorous proofs or definitions, but just to show how easily calculus can be discovered using short, intuitive arguments.

Much of calculus comes from the equation

$$f(x + \Delta x) \approx f(x) + f'(x)\Delta x \tag{1}$$

which expresses the fact that $f'(x)$ is the instantaneous rate of change of f at x. The approximation is good when Δx is small. Equation (1) is practically the definition of $f'(x)$.

Equation (1) can be restated as

$$f(y) \approx f(x) + f'(x)(y - x). \tag{2}$$

Calculus can be viewed as the study of functions that are "locally linear", in the sense that the approximation (2) is good when y is close to x. Perhaps the phrase "f is differentiable at x" could even be replaced with "f is locally linear at x".

The key technique of integral calculus is to chop things up into tiny pieces, compute the contribution of each piece (with the help of the approximation (1)), then add up all the contributions to get the total result.

Throughout these notes, we'll assume (without saying so) that all functions are as smooth as is necessary for the arguments to make sense.

2 Single variable calculus

2.1 Fundamental theorem of calculus

Chop up the interval $[a, b]$ into tiny subintervals $[x_i, x_{i+1}]$. Then

$$\begin{aligned}
\underbrace{f(b) - f(a)}_{\text{total change}} &= \sum_i \underbrace{f(x_{i+1}) - f(x_i)}_{\text{little change}} \\
&\approx \sum_i f'(x_i)\Delta x_i \\
&\approx \int_a^b f'(x)\, dx.
\end{aligned}$$

The total change (across a big interval) is the sum of all the little changes (across tiny subintervals).

Note: It seems plausible that, by chopping up the interval $[a, b]$ into even smaller pieces, we could make the approximation better and better – in fact, it seems that we could make the approximation as close as we like. This implies that the two quantities must in fact be equal. Similar reasoning will be used throughout these notes to move from approximate equality to exact equality, and we won't bother to repeat this argument in each case.

Note: By using this strategy to compute the total change, we have found ourselves computing an "integral". This is one reason that integrals are so important. An intuitive definition of $\int_a^b g(x)\,dx$ is just this: First chop up $[a,b]$ into tiny subintervals $[x_i, x_{i+1}]$, and for each i select a point $z_i \in [x_i, x_{i+1}]$. Then

$$\sum_i g(z_i)\Delta x_i \approx \int_a^b g(x)\,dx.$$

A precise definition would state that $\int_a^b g(x)\,dx$ is in some sense a limit of such approximations.

2.2 Other fundamental theorem of calculus

Let $F(x) = \int_a^x f(s)\,ds$. Then

$$\begin{aligned} F(x+\Delta x) - F(x) &= \int_x^{x+\Delta x} f(s)\,ds \\ &\approx \int_x^{x+\Delta x} f(x)\,ds \\ &= f(x)\Delta x. \end{aligned} \tag{3}$$

By comparing (3) with (1), we discover that $F'(x) = f(x)$.

2.3 Chain rule

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

$$\begin{aligned} f(x+\Delta x) &= g(h(x+\Delta x)) \\ &\approx g(h(x) + h'(x)\Delta x) \\ &\approx \underbrace{g(h(x))}_{f(x)} + g'(h(x))h'(x)\Delta x. \end{aligned} \tag{4}$$

By comparing (4) with (1), we discover that $f'(x) = g'(h(x))h'(x)$.

2.4 Product rule

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

$$\begin{aligned} f(x+\Delta x) &= g(x+\Delta x)h(x+\Delta x) \\ &\approx (g(x) + g'(x)\Delta x)(h(x) + h'(x)\Delta x) \\ &= g(x)h(x) + g'(x)h(x)\Delta x + g(x)h'(x)\Delta x + g'(x)h'(x)\Delta x^2 \\ &\approx f(x) + (g'(x)h(x) + g(x)h'(x))\Delta x. \end{aligned} \tag{5}$$

By comparing (5) with (1), we discover that $f'(x) = g'(x)h(x) + g(x)h'(x)$.

2.5 Integration by parts

We can integrate both sides of the product rule to obtain the integration by parts rule

$$\int_a^b \frac{dg}{dx} h \, dx = -\int_a^b g \frac{dh}{dx} \, dx + gh|_a^b.$$

Linear algebra intuition (optional) From a linear algebra point of view, integration by parts says that the adjoint of $\frac{d}{dx}$ is $-\frac{d}{dx}$ (in a setting where the boundary term vanishes). In other words, $\frac{d}{dx}$ is anti-self-adjoint, hence normal. We thus have reason to hope (based on the spectral theorem) that there is (in some sense) an orthonormal basis of eigenvectors for $\frac{d}{dx}$. Fourier series can be discovered in this way.

2.6 Taylor series approximation

$$\begin{aligned} f(x) &= f(x_0) + \int_{x_0}^{x} f'(s) \, ds \\ &\approx f(x_0) + \int_{x_0}^{x} \underbrace{f'(x_0) + f''(x_0)(s - x_0)}_{\text{first-order approximation to } f'(s)} \quad ds \\ &= f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2} f''(x_0)(x - x_0)^2. \end{aligned}$$

Using a higher-order approximation to $f'(s)$, such as

$$f'(s) \approx f'(x_0) + f''(x_0)(s - x_0) + \frac{1}{2} f'''(x_0)(s - x_0)^2,$$

yields higher-order Taylor series approximations for $f(x)$.

2.7 L'hospital's rule

Assume that $f(x_0) = g(x_0) = 0$, and $g'(x_0) \neq 0$. Then

$$\begin{aligned} \lim_{x \to x_0} \frac{f(x)}{g(x)} &= \lim_{x \to x_0} \frac{f(x_0) + f'(x_0)(x - x_0)}{g(x_0) + g'(x_0)(x - x_0)} \\ &= \frac{f'(x_0)}{g'(x_0)}. \end{aligned}$$

3 Multivariable calculus

Equation (1) works perfectly in the multivariable case where $f : \mathbb{R}^n \to \mathbb{R}^m$.

$$f(\underbrace{x}_{n \times 1} + \underbrace{\Delta x}_{n \times 1}) \approx \underbrace{f(x)}_{m \times 1} + \underbrace{f'(x)}_{m \times n} \underbrace{\Delta x}_{n \times 1}.$$

Note that $f'(x)$ is now an $m \times n$ matrix.

If we prefer to think in terms of linear transformations rather than matrices, we can write

$$f(x + \Delta x) \approx f(x) + Df(x)\Delta x$$

where $Df(x)$ is a linear transformation that takes Δx as input. This is what it means for f to be "locally linear" in the multivariable case.

(Extending the notion of "locally linear" to the multivariable case is one motivation for studying linear transformations in the first place. The need to describe linear transformations concisely leads us to introduce matrices.)

If $f : \mathbb{R}^n \to \mathbb{R}$, then $f'(x)$ is a $1 \times n$ matrix, and $f'(x)\Delta x = \langle \nabla f(x), \Delta x\rangle$ where $\nabla f(x) = f'(x)^T$. In this case, (1) can be written as

$$f(x + \Delta x) \approx f(x) + \langle \underbrace{\nabla f(x)}_{n\times 1}, \underbrace{\Delta x}_{n\times 1}\rangle. \tag{6}$$

3.1 Directional derivative

Let $f : \mathbb{R}^n \to \mathbb{R}$ and let $u \in \mathbb{R}^n$. Then

$$\begin{aligned} D_u f(x) &= \lim_{t\to 0} \frac{f(x + tu) - f(x)}{t} \qquad \text{(by definition)} \\ &= \lim_{t\to 0} \frac{f(x) + \langle \nabla f(x), tu\rangle - f(x)}{t} \\ &= \langle \nabla f(x), u\rangle. \end{aligned}$$

When $u = e_i$ (the ith standard basis vector), we discover that the ith component of $\nabla f(x)$ is the ith partial derivative of f at x:

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{bmatrix}. \tag{7}$$

What direction u should we choose to make the directional derivative $D_u f(x)$ as large as possible? To make the inner product $\langle \nabla f(x), u\rangle$ as large as possible, we should choose u to be in the same direction as $\nabla f(x)$. Hence $\nabla f(x)$ points in the direction of steepest ascent for f at x.

3.2 Jacobian

Let $f : \mathbb{R}^n \to \mathbb{R}^m$. The matrix $f'(x)$ is called the "Jacobian" of f at x.

Let v_i be the ith row of $f'(x)$. Looking at equation (1) component by component, we see that

$$f_i(x + \Delta x) \approx f_i(x) + v_i \Delta x,$$

where f_i is the ith component function of f. This reveals that $v_i = f_i'(x) = \nabla f_i(x)^T$. Using (7), we obtain the formula

$$f'(x) = \begin{bmatrix} \frac{\partial f_1(x)}{\partial x_1} & \cdots & \frac{\partial f_1(x)}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m(x)}{\partial x_1} & \cdots & \frac{\partial f_m(x)}{\partial x_n} \end{bmatrix}. \tag{8}$$

3.3 Note on matrix transpose

If u and v are column vectors in $\mathbb{R}^n$, then $\langle u, v\rangle = u^T v$. We will sometimes use the fact that if $M \in \mathbb{R}^{m\times n}$, then

$$\begin{aligned} \langle Mx, y\rangle &= (Mx)^T y \\ &= x^T M^T y \\ &= \langle x, M^T y\rangle. \end{aligned}$$

This is the key property of the transpose matrix.

3.4 Hessian

Let $f : \mathbb{R}^n \to \mathbb{R}$, and let

$$g(x) = \nabla f(x).$$

So $g : \mathbb{R}^n \to \mathbb{R}^n$.

The matrix $g'(x)$ is called the "Hessian" of f at x, and is sometimes denoted $Hf(x)$. Equations (7) and (8) together yield the formula

$$Hf(x) = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_n \partial x_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix}.$$

Note that

$$\nabla f(x + \Delta x) \approx \nabla f(x) + Hf(x)\Delta x.$$

We might notice by experimentation that mixed partials are equal, which implies that $Hf(x)$ is symmetric. On the other hand, we can argue directly that $Hf(x)$ is symmetric, as follows. First note that

$$\begin{aligned} f(x + \Delta u + \Delta v) &\approx f(x + \Delta u) + \langle \nabla f(x + \Delta u), \Delta v\rangle \\ &\approx f(x) + \langle \nabla f(x), \Delta u\rangle + \langle \nabla f(x), \Delta v\rangle + \langle Hf(x)\Delta u, \Delta v\rangle. \end{aligned}$$

Alternatively,

$$\begin{aligned} f(x + \Delta u + \Delta v) &\approx f(x + \Delta v) + \langle \nabla f(x + \Delta v), \Delta u\rangle \\ &\approx f(x) + \langle \nabla f(x), \Delta v\rangle + \langle \nabla f(x), \Delta u\rangle + \langle Hf(x)\Delta v, \Delta u\rangle. \end{aligned}$$

Comparing these two approximations shows that

$$\langle Hf(x)\Delta u, \Delta v\rangle \approx \langle \Delta u, Hf(x)\Delta v\rangle$$

when Δu and Δv are small, which shows that $Hf(x)$ is symmetric.

The symmetry of $Hf(x)$ implies that mixed partials are equal.

(A similar argument could directly show equality of mixed partials without mentioning the Hessian.)

3.5 A multivariable product rule

Suppose $g, h : \mathbb{R}^n \to \mathbb{R}^m$ and

$$f(x) = \langle g(x), h(x)\rangle.$$

Then

$$\begin{aligned} f(x+\Delta x) &\approx \langle g(x) + g'(x)\Delta x, h(x) + h'(x)\Delta x\rangle \\ &\approx \langle g(x), h(x)\rangle + \langle g'(x)\Delta x, h(x)\rangle + \langle g(x), h'(x)\Delta x\rangle \\ &= f(x) + \langle \underbrace{g'(x)^T h(x) + h'(x)^T g(x)}_{\nabla f(x)}, \Delta x\rangle. \end{aligned}$$

3.6 Multivariable chain rule

The chain rule derivation above works perfectly in the case where $h : \mathbb{R}^n \to \mathbb{R}^p$ and $g : \mathbb{R}^p \to \mathbb{R}^m$. However, it's also enlightening to directly intuit the chain rule formula in the special case where

$$f(x) = g(h_1(x), \ldots, h_p(x)),$$

and $g : \mathbb{R}^p \to \mathbb{R}$ and $h_i : \mathbb{R} \to \mathbb{R}$. (So $f : \mathbb{R} \to \mathbb{R}$.) Let's assume $p = 2$ for simplicity.

First note that

$$\begin{aligned} g(u_1 + \Delta u_1, u_2 + \Delta u_2) - g(u_1, u_2) &= g(u_1 + \Delta u_1, u_2 + \Delta u_2) - g(u_1, u_2 + \Delta u_2) \\ &\quad + g(u_1, u_2 + \Delta u_2) - g(u_1, u_2) \\ &\approx D_1 g(u_1, u_2 + \Delta u_2)\Delta u_1 + D_2 g(u_1, u_2)\Delta u_2 \\ &\approx D_1 g(u_1, u_2)\Delta u_1 + D_2 g(u_1, u_2)\Delta u_2. \end{aligned}$$

This is just another way to say that $g(u+\Delta u) \approx g(u) + \langle \nabla g(u), \Delta u\rangle$. We already knew this, but this derivation explains why we expect g to be differentiable when g has continuous partial derivatives.

Now we compute $f'(x)$:

$$\begin{aligned} f(x+\Delta x) &= g(h_1(x+\Delta x), h_2(x+\Delta x)) \\ &\approx g(h_1(x) + h_1'(x)\Delta x, h_2(x) + h_2'(x)\Delta x) \\ &\approx g(h_1(x), h_2(x)) + D_1 g(h_1(x), h_2(x))h_1'(x)\Delta x + D_2 g(h_1(x), h_2(x))h_2'(x)\Delta x \\ &= f(x) + \underbrace{(D_1 g(h_1(x), h_2(x))h_1'(x) + D_2 g(h_1(x), h_2(x))h_2'(x))}_{f'(x)}\Delta x. \end{aligned}$$

Of course, this is just another way of saying that $f'(x) = g'(h(x))h'(x)$, which we already knew.

3.7 Multivariable Taylor series

Taylor series approximations to $f : \mathbb{R}^n \to \mathbb{R}$ can be derived by introducing $g(t) = f(x_0 + t(x - x_0))$, and computing the single variable Taylor series approximations to g. For example,

$$g(1) \approx g(0) + g'(0) + \frac{1}{2}g''(0). \tag{9}$$

From the chain rule,

$$\begin{aligned} g'(t) &= f'(x_0 + t(x - x_0))(x - x_0) \\ &= \langle \nabla f(x_0 + t(x - x_0)), x - x_0 \rangle. \end{aligned}$$

The product rule and the chain rule together allow us to compute $g''(t)$:

$$g''(t) = \langle x - x_0, Hf(x_0 + t(x - x_0))(x - x_0) \rangle.$$

Equation (9) becomes

$$f(x) \approx f(x_0) + \langle \nabla f(x_0), x - x_0 \rangle + \frac{1}{2}(x - x_0)^T Hf(x_0)(x - x_0) \tag{10}$$

which is the second order Taylor series approximation to f at x. Higher order Taylor series approximations can be derived also, but this requires a skillful use of notation.

3.8 Classifying critical points

When $\nabla f(x_0) = 0$, equation (10) gives us useful information about how f behaves near x_0. In particular, if $Hf(x_0)$ is positive definite, then (by definition)

$$(x - x_0)^T Hf(x_0)(x - x_0) > 0$$

for all $x \neq x_0$, which shows that f has a local minimum at x_0.

Similarly, if $Hf(x_0)$ is negative definite, then f has a local maximum at x_0. If $Hf(x_0)$ is indefinite, then f has a saddle point at x_0.

3.9 Lagrange multipliers

Suppose $x^\star$ is a local minimizer for the problem

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & g(x) = 0. \end{aligned}$$

Here $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}$. Then, for all sufficiently small Δx, we have:

$$\text{if } g(x^\star + \Delta x) = 0, \text{ then } f(x^\star + \Delta x) \geq f(x^\star).$$

(Otherwise $x^\star$ is not a local minimizer.) Making the approximations

$$\begin{aligned} & g(x^\star + \Delta x) \approx g(x^\star) + \langle \nabla g(x^\star), \Delta x \rangle \\ \text{and} \quad & f(x^\star + \Delta x) \approx f(x^\star) + \langle \nabla f(x^\star), \Delta x \rangle, \end{aligned}$$

we conclude that

$$\text{if } \langle \nabla g(x^\star), \Delta x \rangle = 0, \text{ then } \langle \nabla f(x^\star), \Delta x \rangle \geq 0$$

for sufficiently small Δx. It follows that

$$\text{if } \langle \nabla g(x^\star), \Delta x \rangle = 0, \text{ then } \langle \nabla f(x^\star), \Delta x \rangle = 0$$

for sufficiently small Δx.

In other words, $\nabla f(x^\star)$ is orthogonal to everything orthogonal to $\nabla g(x^\star)$. This implies that $\nabla f(x^\star)$ is parallel to $\nabla g(x^\star)$:

$$\nabla f(x^\star) = \lambda \nabla g(x^\star)$$

for some $\lambda \in \mathbb{R}$.

A similar argument works when $g : \mathbb{R}^n \to \mathbb{R}^m$, but in that case we need to use the four subspace theorem from linear algebra.

3.10 Definition of integral

Suppose $f : R \to \mathbb{R}$, where $R = [a, b] \times [c, d] \subset \mathbb{R}^2$. Chop up $[a, b]$ into tiny subintervals $[x_i, x_{i+1}]$, and chop up $[c, d]$ into tiny subintervals $[y_j, y_{j+1}]$. The rectangle R is correspondingly chopped up into tiny subrectangles $R_{ij} = [x_i, x_{i+1}] \times [y_j, y_{j+1}]$. For each (i, j), pick a point $z_{ij} \in R_{ij}$. Then

$$\sum_{i,j} f(z_{ij}) \Delta x_i \Delta y_j \approx \int_R f \, dx \, dy.$$

This integral is also denoted $\int_R f(x, y) \, dx \, dy$. A precise definition would state that $\int_R f \, dx \, dy$ is in some sense a limit of approximations like this.

Now suppose that $f : \Omega \to \mathbb{R}$, where $\Omega \subset \mathbb{R}^2$ is not a rectangle, but is contained in a rectangle $R = [a, b] \times [c, d]$. We can extend f to a function $\tilde{f}$ defined on the entire rectangle R by declaring that $\tilde{f}$ is equal to 0 at all points of R that don't belong to Ω. We can then define $\int_\Omega f \, dx \, dy = \int_R \tilde{f} \, dx \, dy$.

A similar definition allows us to integrate over subsets of $\mathbb{R}^n$ when $n > 2$.

By the way, notice that

$$\begin{aligned}
\int_R f(x,y)\,dx\,dy &\approx \sum_{i,j} f(x_i,y_j)\Delta x_i \Delta y_j \\
&= \sum_i \left(\sum_j f(x_i,y_j)\Delta y_j \right) \Delta x_i \\
&\approx \sum_i \int_c^d f(x_i,y)\,dy\,\Delta x_i \\
&\approx \int_a^b \int_c^d f(x,y)\,dy\,dx.
\end{aligned}$$

Of course, we could equally well argue that

$$\int_R f(x,y)\,dx\,dy = \int_c^d \int_a^b f(x,y)\,dx\,dy.$$

This is Fubini's theorem. Similar arguments give us Fubini's theorem in higher dimensions.

3.11 Change of variables formula

Let X and Y be open subsets of $\mathbb{R}^n$, and assume that $T : X \to Y$ is 1-1 and onto. Let $f : Y \to \mathbb{R}$.

Chop up Y into tiny subsets Y_i. Because of the 1-1 correspondence between X and Y, X is correspondingly chopped up into tiny subsets X_i such that $T(X_i) = Y_i$.

For each i, pick a point $y_i \in Y_i$. Let x_i be the corresponding point in X_i. So $T(x_i) = y_i$.

If x is close to x_i, then

$$T(x) \approx T(x_i) + T'(x_i)(x - x_i).$$

The function

$$T_i(x) = T(x_i) + T'(x_i)(x - x_i)$$

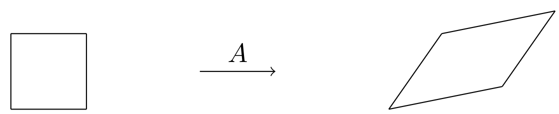
is called the "local linear approximation" to T at x_i, and $T(x) \approx T_i(x)$ when x is near x_i.

Let $\hat{Y}_i = T_i(X_i)$. $\hat{Y}_i$ is an approximation of Y_i. A key fact is that

$$m(\hat{Y}_i) = |\det T'(x_i)| m(X_i).$$

Here $m(S)$ denotes the "measure" of a subset S of $\mathbb{R}^n$, as discussed in section 3.10.

When $n = 2$, you can derive this fact easily by drawing a picture. If R is a tiny square, and A is a 2×2 matrix, then AR is a parallelogram. With high school geometry you can compute the area of this parallelogram, and discover that the answer is $|\det A|m(R)$. If the determinant has previously been discovered by deriving formulas for the solution of 2×2 or 3×3 linear systems (discovering Cramer's rule), then it's surprising and beautiful that the determinant pops up here too. A picture proof is also straightforward (but tedious) in the case where $n = 3$. Based on this evidence, we would not hesitate to guess that the formula holds for any n. This can be proved using linear algebra – for example the SVD provides a nice way to look at it.



We're now ready to derive the change of variables formula for integration:

$$\begin{aligned}
\int_Y f(y)\,dy &\approx \sum_i f(y_i)m(Y_i) \\
&\approx \sum_i f(y_i)m(\hat{Y}_i) \\
&= \sum_i f(T(x_i))|\det T'(x_i)|m(X_i) \\
&\approx \int_X f(T(x))|\det T'(x)|\,dx.
\end{aligned}$$

3.12 Definition of line integral

Let C be a smooth directed curve in $\mathbb{R}^n$, and let f be a vector field on C (so $f : C \to \mathbb{R}^n$).

Chop up C into tiny curves C_i, each of which is approximated by a line segment spanned by a vector Δx_i. (The direction of Δx_i is chosen to be consistent with the direction of C.)

For each i, pick a point $z_i \in C_i$. Then

$$\sum_i \langle f(z_i), \Delta x_i \rangle \approx \int_C f \cdot dx.$$

A precise definition would state that $\int_C f \cdot dx$ is in some sense a limit of approximations like this.

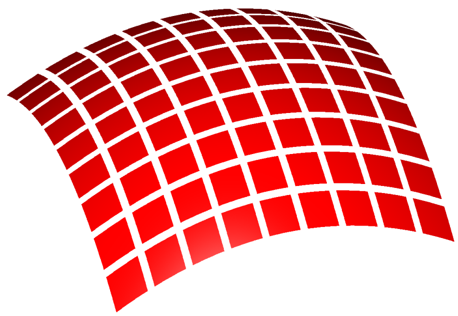
A similar definition allows us to integrate a scalar-valued function over C. In this case we don't require C to have a direction.

3.13 Definition of surface integral

Let $S \subset \mathbb{R}^3$ be a smooth oriented surface, and let f be a vector field on S (so $f : S \to \mathbb{R}^3$).

(For S to be "oriented" means, roughly speaking, that one side of S has been designated the "outside" and the other side has been designated the "inside". Some surfaces, such as a Mobius strip, can't be oriented.)

Chop up S into tiny pieces S_i, each of which is approximated by a parallelogram spanned by vectors u_i and v_i (chosen so that $u_i \times v_i$ points "outward").



For each i, pick a point $z_i \in S_i$. Then

$$\sum_i \langle f(z_i), u_i \times v_i \rangle \approx \int_S f \cdot dA.$$

A precise definition would state that $\int_S f \cdot dA$ is in some sense a limit of approximations like this.

A similar definition allows us to integrate scalar-valued functions over S. In this case we don't require S to have an orientation.

Differential forms viewpoint. At each point $x \in S$, let $\omega(x)$ be the alternating bilinear function that maps (u, v) to $\langle f(x), u \times v \rangle$. The function ω is a "differential 2-form" on S. We can integrate a differential 2-form over S using a similar definition:

$$\sum_i \omega(z_i)(u_i, v_i) \approx \int_S \omega.$$

This differential forms viewpoint suggests how to generalize the idea of integration to higher dimensional manifolds.

Question: Is it obvious that S can be chopped up into tiny pieces, each of which is approximately a parallelogram?

Hint: Consider the case where S has a parametrization $g : R \to S$, where $R = [a, b] \times [c, d]$. Chop up R into tiny rectangles, and imagine how S is correspondingly chopped up. Use the fact that g is "locally linear". A linear transformation maps a rectangle to a parallelogram.

By working this out in detail, we could express an integral over S in terms of an integral over R. This gives us a way to evaluate surface integrals explicitly.

3.14 Fundamental theorem of calculus for line integrals

Suppose that C is a curve connecting points $a, b \in \mathbb{R}^n$, and let $f : \mathbb{R}^n \to \mathbb{R}$. Chop up C into tiny curves C_i that start at x_i and end at x_{i+1}. Then

$$\begin{aligned}\underbrace{f(b) - f(a)}_{\text{total change}} &= \sum_i \underbrace{f(x_{i+1}) - f(x_i)}_{\text{little change}} \\ &\approx \sum_i \langle \nabla f(x_i), \Delta x_i \rangle \\ &\approx \int_C \nabla f(x) \cdot dx.\end{aligned}$$

(The total change is the sum of all the little changes.)

If C is a closed curve, so $a = b$, then $\int_C \nabla f(x) \cdot dx = 0$. On the other hand, suppose that g is a vector field on $\mathbb{R}^n$ and that the integral of g over any closed curve is equal to 0. Can we conclude that $g = \nabla f$ for some function $f : \mathbb{R}^n \to \mathbb{R}$? Yes. Select a point $x_0 \in \mathbb{R}^n$ arbitrarily, and define $f(x) = \int_{x_0}^x g(s) \cdot ds$, where the line integral defining f is taken over *any* curve connecting x_0 to x. (It doesn't matter which curve you pick, because the integral of g over any closed curve is 0, which implies that any two curves from x_0 to x must yield the same result.) Then

$$\begin{aligned}f(x + \Delta x) - f(x) &= \int_x^{x+\Delta x} g(s) \cdot ds \\ &\approx \int_x^{x+\Delta x} g(x) \cdot ds \\ &= \langle g(x), \Delta x \rangle. \end{aligned} \tag{11}$$

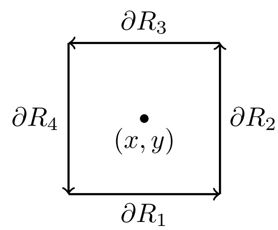
Comparing (11) with (6), we see that $\nabla f(x) = g(x)$.

In physics, a vector field g whose integral over any closed curve is 0 is called "conservative". The function f is called a "potential function" for g.

3.15 Green's theorem

Let Ω be an open subset of $\mathbb{R}^2$, with a piecewise smooth boundary $\partial\Omega$. Let f be a vector field on $\mathbb{R}^2$, with component functions f_1 and f_2. We want to compute $\int_{\partial\Omega} f \cdot dr$, where $\partial\Omega$ is oriented counterclockwise. Our strategy is to chop up Ω into tiny squares and triangles Ω_i, and compute $\int_{\partial\Omega_i} f \cdot dr$ for each i. Each boundary $\partial\Omega_i$ is given a counterclockwise orientation. When we add up all those individual line integrals, wonderful cancellation occurs and we are left with $\int_{\partial\Omega} f \cdot dx$.

Let R be a tiny square of width Δx and height Δy, centered at the point (x, y). Then ∂R consists of 4 pieces:



Therefore

$$\begin{aligned}
\int_{\partial R} f \cdot dr &= \int_{\partial R_1} f \cdot dr + \int_{\partial R_2} f \cdot dr + \int_{\partial R_3} f \cdot dr + \int_{\partial R_4} f \cdot dr \\
&\approx \quad f_1\left(x, y - \Delta y/2\right) \Delta x + f_2\left(x + \Delta x/2, y\right) \Delta y \\
&\quad - f_1\left(x, y + \Delta y/2\right) \Delta x - f_2\left(x - \Delta x/2, y\right) \Delta y \\
&\approx -\frac{\partial f_1(x,y)}{\partial y} \Delta y \Delta x + \frac{\partial f_2(x,y)}{\partial x} \Delta x \Delta y \\
&= \left(\frac{\partial f_2(x,y)}{\partial x} - \frac{\partial f_1(x,y)}{\partial y} \right) \Delta x \Delta y.
\end{aligned}$$

A similar calculation works for triangles. Adding up all the contributions from the squares and triangles Ω_i, we find that

$$\int_{\partial \Omega} f \cdot dr = \int_{\Omega} \left(\frac{\partial f_2(x,y)}{\partial x} - \frac{\partial f_1(x,y)}{\partial y} \right) dx\, dy.$$

3.16 Divergence theorem

A very similar argument can be used to derive the divergence theorem:

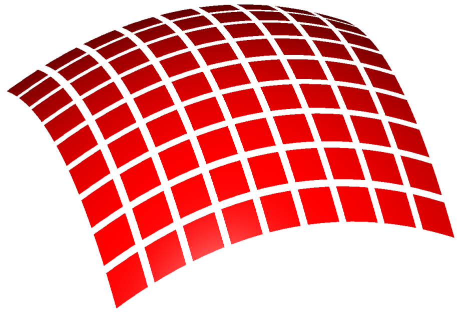
$$\int_{\partial \Omega} f \cdot dA = \int_{\Omega} \operatorname{div} f \, dx.$$

Here Ω is an open subset of $\mathbb{R}^n$ with a piecewise smooth boundary, and f is a vector field on $\mathbb{R}^n$. $\partial\Omega$ is given the outward orientation. The strategy is to chop up Ω into tiny n-cubes Ω_i, and compute the integral of f over $\partial\Omega_i$ for each i.

3.17 Stokes' theorem (classical version)

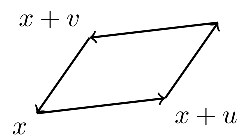
Let S be an oriented surface in $\mathbb{R}^3$ with a piecewise smooth boundary. Let F be a vector field on $\mathbb{R}^3$. We want to extend the derivation of Green's theorem to this situation, to discover a theorem like Green's theorem that relates $\int_{\partial S} F \cdot dr$ to an integral over S.

Our strategy is to chop up S into a bunch of tiny pieces S_i, each of which is approximately a parallelogram.



Then we'll compute $\int_{\partial S_i} F \cdot dr$ for each i. When we add up all these tiny contributions, wonderful cancellation occurs, and we're left with $\int_{\partial S} F \cdot dr$.

The key step is to calculate $\int_{\partial P} F \cdot dr$, where $P \subset \mathbb{R}^3$ is a tiny oriented parallelogram. Assume that the corners of P are x, $x+u$, $x+v$, and $x+u+v$.



Then

$$\begin{aligned}\int_{\partial P} F \cdot dr \approx \quad & \langle F(x), u\rangle + \langle F(x+u), v\rangle \\ & - \langle F(x+v), u\rangle - \langle F(x), v\rangle \\ \approx & -\langle F'(x)v, u\rangle + \langle F'(x)u, v\rangle.\end{aligned}$$

In the last step, we used the approximations

$$F(x+u) \approx F(x) + F'(x)u, \quad F(x+v) \approx F(x) + F'(x)v.$$

At this point, the rest of the calculation is completely straightforward. We are nearly done already. All we need to do now is write out everything in terms of their components. Picking up where we left off:

$$\begin{aligned}
& -\langle F'(x)v, u\rangle + \langle F'(x)u, v\rangle \\
= \quad & \langle u, (F'(x)^T - F'(x))v\rangle \\
= \quad & u_1 \left(\left(\frac{\partial F_2}{\partial x_1} - \frac{\partial F_1}{\partial x_2} \right) v_2 + \left(\frac{\partial F_3}{\partial x_1} - \frac{\partial F_1}{\partial x_3} \right) v_3 \right) \\
& + u_2 \left(\left(\frac{\partial F_1}{\partial x_2} - \frac{\partial F_2}{\partial x_1} \right) v_1 + \left(\frac{\partial F_3}{\partial x_2} - \frac{\partial F_2}{\partial x_3} \right) v_3 \right) \\
& + u_3 \left(\left(\frac{\partial F_1}{\partial x_3} - \frac{\partial F_3}{\partial x_1} \right) v_1 + \left(\frac{\partial F_2}{\partial x_3} - \frac{\partial F_3}{\partial x_2} \right) v_2 \right) \\
= \quad & \left(\frac{\partial F_3}{\partial x_2} - \frac{\partial F_2}{\partial x_3} \right) (u_2 v_3 - u_3 v_2) \\
& + \left(\frac{\partial F_1}{\partial x_3} - \frac{\partial F_3}{\partial x_1} \right) (u_3 v_1 - u_1 v_3) \\
& + \left(\frac{\partial F_2}{\partial x_1} - \frac{\partial F_1}{\partial x_2} \right) (u_1 v_2 - u_2 v_1) \\
= \; & \langle \nabla \times F, u \times v \rangle
\end{aligned}$$

All the partial derivatives, as well as $\nabla \times F$, are evaluated at x. It's beautiful that the answer can be expressed so simply, in terms of ∇ and the cross product.

We're now ready to derive the classical Stokes' theorem. Chop up S into tiny pieces S_i, each of which is approximated by a parallelogram spanned by vectors u_i and v_i (chosen so that $u_i \times v_i$ points "outward"). Then

$$\begin{aligned}
\int_{\partial S} F \cdot dr &= \sum_i \int_{\partial S_i} F \cdot dr \\
&\approx \sum_i \langle \nabla \times F, u_i \times v_i \rangle \\
&\approx \int_S \nabla \times F \cdot dA.
\end{aligned}$$