Taylor's Theorem for smooth functions

Let p be a point of a real affine space $\mathbb{M}$ of finite dimension, whose tangent space at any point is the real vector space $\mathbb{X}$ of finite dimension.
Let $\mathbb{V}$ be an open subset of $\mathbb{M}$, such that for any $x \in \mathbb{X}$, the line segment joining x to p lies in $\mathbb{V}$.
This line segment has the natural parametrization $x+t(p-x)$, where $t \in [0,1]$.
Let $f : \mathbb{V} \to \mathbb{Y}$ be a smooth function taking values in a real vector space $\mathbb{Y}$ of finite dimensions.

For any positive integer k and for any $x \in \mathbb{V}$, put:

$$g_k(x) = \int_0^1 \frac{t^{k-1}}{(k-1)!} f^{(k)}(x + t(p-x))\, dt,$$

$$h_k(x) = g_k(x)((x-p)^k),$$

$$(x-p)^k = (x-p)^{\otimes k} = (x-p) \otimes (x-p) \cdots \otimes (x-p) \ (k \text{ factors of } (x-p) \text{ in this tensor product}).$$

Note that g_k is well-defined and smooth on $\mathbb{V}$ and takes values in $(\mathbb{X}^*)^{\odot k} \otimes \mathbb{Y}$.
Also we have the formulas, using the Fundamental Theorem of Calculus and the Chain Rule:

$$g_k(p) = \frac{f^{(k)}(p)}{(k-1)!} \int_0^1 t^{k-1}\, dt = \frac{f^{(k)}(p)}{k!},$$

$$h_1(x) = g_1(x)(x-p) = \int_0^1 f^{(1)}(x + t(p-x))(x-p)\, dt$$

$$= -\int_0^1 \left(\frac{d}{dt}\left(f(x + t(p-x)) \right) \right) dt = -\left[f(x+t(p-x)) \right]_{t=0}^{t=1} = f(x) - f(p).$$

Also, again by the Fundamental Theorem of Calculus and the Chain Rule, we have, for any positive integer k and any $x \in \mathbb{V}$:

$$h_k(x) - h_{k+1}(x) = \int_0^1 \left(\frac{t^{k-1}}{(k-1)!} f^{(k)}(x + t(p-x))((x-p)^k) \ - \ \frac{t^k}{k!} f^{(k+1)}(x + t(p-x))((x-p)^{k+1}) \right) dt$$

$$= \int_0^1 \frac{d}{dt}\left(\frac{t^k}{k!} f^{(k)}(x + t(p-x))((x-p)^k) \right) dt$$

$$= \left[\frac{t^k}{k!} f^{(k)}(x + t(p-x))((x-p)^k) \right]_{t=0}^{t=1} = \frac{1}{k!} f^{(k)}(p)((x-p)^k).$$

Summing this relation from $k = 1$ to n, the sum on the left-hand side collapses and we get, for any $x \in \mathbb{V}$ and any positive integer n, the formula:

$$h_1(x) - h_{n+1}(x) = \sum_{k=1}^{n} \frac{1}{k!} f^{(k)}(p)((x - p)^k).$$

Rearranging, and using the formula $h_1(x) = f(x) - f(p)$, proved above, we have proved Taylor's theorem for smooth functions, valid for any $x \in \mathbb{V}$ and any non-negative integer n:

$$f(x) = g_{n+1}(x)((x - p)^{n+1}) + \sum_{k=0}^{n} \frac{1}{k!}(f^{(k)}(p)((x - p)^k),$$

$$g_{n+1}(x) = \frac{1}{n!} \int_0^1 t^n f^{(n+1)}(x + t(p - x)) \, dt, \quad g_{n+1}(p) = \frac{f^{(n+1)}(p)}{(n+1)!}.$$

Here $g_{n+1} : \mathbb{V} \to (\mathbb{X}^*)^{\odot(n+1)} \otimes \mathbb{Y}$ is smooth.

Note that the formula, for a given non-negative integer n, is also valid, as written, when f is only $\mathcal{C}^{n+1}$ on $\mathbb{V}$, except that the remainder function g_{n+1} is only continuous, not necessarily smooth.

Finally note that if we write $f^{(k)} = D^k f$, where D is the derivative operator, we have formally (non-rigorously), using the series for e^u:

$$(e^{(x-p).D} f)(p) - f(x) = \left[(e^{t(x-p).D} f)(x + t(p - x)) \right]_{t=0}^{t=1}$$

$$= \int_0^1 \frac{d}{dt} \left((e^{t(x-p).D} f)(x + t(p - x)) \right) \, dt = \sum_{k=0}^{\infty} \int_0^1 \frac{d}{dt} \left(\frac{t^k}{k!} f^{(k)}(x + t(p - x))((x - p)^{\odot k}) \right) \, dt$$

$$= \sum_{k=1}^{\infty} (h_k(x) - h_{k+1}(x)) + \int_0^1 \frac{d}{dt} \left[f(x + t(p - x)) \right] \, dt = h_1(x) + f(p) - f(x) = 0.$$

Here we needed to assume that $h_n(x) \to 0$ as $n \to \infty$.
So formally, we have the equation $(e^{(x-p).D} f)(p) = f(x)$.
This formula can be rigorously justified in the analytic case for x in the range of convergence around the base point p.
Note that in the smooth case, however, it is generally false, since, for example in one-dimension, for the smooth function $f(x) = e^{-\frac{1}{x}}$, for $x > 0$ and $f(x) = 0$, for $x \leq 0$, we have $(e^{x.D} f)(0) = 0 \neq f(x)$, since f and all its derivatives vanish at the origin.

Vector fields, first-order homogeneous differential operators and derivations

We work in the smooth category.

Fix a real affine space $\mathbb{M}$ of finite dimension, whose tangent space at any point is the real finite dimensional vector space $\mathbb{X}$.

Also fix $\mathbb{U}$ a non-empty open subset of $\mathbb{M}$.

The co-ordinate vector function x of $\mathbb{U}$ assigns to any $x \in \mathbb{U}$ the position vector $x \in \mathbb{X}$, relative to a chosen origin for $\mathbb{M}$, so x can be regarded as a smooth function on $\mathbb{U}$ with values in $\mathbb{X}$.

In particular we have, for its (constant) derivative:

$$x' = \delta \in \mathbb{X}^* \otimes \mathbb{X}.$$

Here δ is the Kronecker delta, which gives the identity endomorphism of $\mathbb{X}$.

A derivation v of $\mathcal{C}^\infty(\mathbb{U})$ is by definition a map $v : \mathcal{C}^\infty(\mathbb{U}) \to \mathcal{C}^\infty(\mathbb{U})$, such that:

- v kills constants, $v(c) = 0$, for any constant function c on $\mathbb{U}$.

- v is an abelian group homomorphism: $v(f + g) = v(f) + v(g)$, for any f and g in $\mathcal{C}^\infty(\mathbb{U})$.

- v obeys a Liebniz rule: for any f and g in $\mathbb{U}$, we have:

$$v(fg) = v(f)g + fv(g).$$

Intuitively v is an infinitesimal automorphism of $\mathcal{C}^\infty(\mathbb{U})$.

Denote by $\mathcal{T}(\mathbb{U})$ the collection of all derivations of $\mathcal{C}^\infty(\mathbb{U})$.

Then it is clear that $\mathcal{T}(\mathbb{U})$ is a $\mathcal{C}^\infty(\mathbb{U})$-module, under left multiplication by elements of $\mathcal{C}^\infty(\mathbb{U})$.

- By definition, a smooth vector field on $\mathbb{U}$ is a derivation of $\mathcal{C}^\infty(\mathbb{U})$.

We want to find a concrete description of $\mathcal{T}(\mathbb{U})$.

Finally note that we can introduce the derivative derivation ∂:

$$\partial : \mathbb{C}^\infty(\mathbb{U}) \to \mathbb{X}^* \otimes \mathbb{C}^\infty(\mathbb{U}).$$

Here ∂ is defined by the succinct formula, valid for any $f \in \mathbb{C}^\infty(\mathbb{U})$:

$$\partial(f) = f'.$$

To this end, we introduce the derivative derivation ∂:

$$\partial : \mathbb{C}^\infty(\mathbb{U}) \to \mathbb{X}^* \otimes \mathbb{C}^\infty(\mathbb{U}).$$

Here ∂ is defined by the succinct formula, valid for any $f \in \mathbb{C}^\infty(\mathbb{U})$:

$$\partial(f) = f'.$$

Note that the properties of the derivative immediately give that ∂ is a derivation, as required. Then if w is a smooth function on $\mathbb{U}$ with values in $\mathbb{X}$, w determines an operator on $\mathcal{C}^\infty(\mathbb{U})$, denoted $w.\partial : \mathcal{C}^\infty(\mathbb{U}) \to \mathcal{C}^\infty(\mathbb{U})$ given by the formula, valid for any $f \in \mathcal{C}^\infty(\mathbb{U})$ and any $x \in \mathbb{U}$:

$$(w.\partial(f))(x) = f'(x)(w(x)).$$

So the pairing $(w, \partial) \to w.\partial$ is the usual dual pairing of a dual vector with a vector. Normally one would write this as $\partial(w)$, but we want to make clear that the derivative operator is not acting on w, so we keep it to the left of the derivative operator ∂, which acts to the right. Note that immediately from its definition, by the standard properties of the derivative, we have $w.\partial \in \mathcal{T}(\mathbb{U})$.

Also note that since x' is the Kronecker delta, we have:

$$w.\partial(x) = w(x).$$

This means, in particular, that given the operator $w.\partial$, the smooth function w is uniquely determined.

- By definition, a first-order smooth homogeneous differential operator is an operator on $\mathcal{C}^\infty(\mathbb{U})$, of the form $w.\partial$, for some w, a smooth function on $\mathbb{U}$ with values in $\mathbb{X}$.

(A general smooth first-order operator is then of the form $f \to w.\partial(f) + gf$, where g is a smooth function on $\mathbb{U}$).

The main theorem we want to prove is:

- An operator on $\mathcal{C}^\infty(\mathbb{U})$ is a derivation, if and only if it is a smooth vector field, if and only if it is a first-order smooth homogeneous differential operator.

Everything in this theorem is immediate from the definitions and the discussion above, except for the following, which we have to prove:

- Given a derivation v of $\mathcal{C}^\infty(\mathbb{U})$, we must show that for any $f \in \mathcal{C}^\infty(\mathbb{U})$ and for any $x \in \mathbb{U}$, we have the formula:

$$(v(f))(x) = f'(x)(v(x)).$$

The first key result is the following:

- Let $\phi \neq \mathbb{V} \subset \mathbb{U}$, with $\mathbb{V}$ open.
  If $f \in \mathcal{C}^\infty(\mathbb{U})$ is such that $f|_\mathbb{V}$ vanishes, then $v(f)|_\mathbb{V} = 0$.

Proof:

Let $p \in \mathbb{V}$.
We must show that $v(f)(p) = 0$.
Let $b \in \mathcal{C}^\infty(\mathbb{U})$ be a bump function, supported in $\mathbb{V}$, such that there exists an open subset $\mathbb{W} \subset \mathbb{V}$, with $p \in \mathbb{W}$, such that $b|_\mathbb{W} = 1$ (the constant function). We know that such a function b exists, by work done earlier this term.

Then we have $bf = 0$ on $\mathbb{U}$, since b vanishes outside $\mathbb{V}$ and f vanishes inside. We write the Liebniz rule:

$$0 = v(0) = v(bf) = bv(f) + v(b)f.$$

Evaluate at the point p.
We get, since $b(p) = 1$ and $f(p) = 0$:

$$0 = b(p)v(f)(p) + v(b)(p)f(p) = v(f)(p) + 0 = v(f)(p).$$

So the required result holds and we are done.

We can rephrase this result as follows: for $f \in \mathbb{C}^\infty(\mathbb{U})$, denote by $\mathcal{Z}(f)$ the closure of the open subset of $\mathbb{U}$ on which f is non-zero. $\mathcal{Z}(f)$ is called the support of f. Note that f vanishes on the open set $\mathbb{U} - \mathcal{Z}(f)$. Then the result just proved implies that for any $f \in \mathbb{C}^\infty(\mathbb{U})$ and for any vector field $v \in \mathcal{T}(\mathbb{U})$, we have $\mathcal{Z}(v(f)) \subset \mathcal{Z}(f)$: i.e. the action of v on $\mathcal{C}^\infty(\mathbb{U})$ is support non-increasing.

Now fix $p \in \mathbb{U}$.

Pick $\mathbb{V}$, an open ball centered at p with $\mathbb{V} \subset \mathbb{U}$.

The ball $\mathbb{V}$ exists, since $\mathbb{U}$ is open.

Also if $x \in \mathbb{V}$, the line segment from p to x lies entirely within $\mathbb{V}$.

So we can apply the second-order Taylor Theorem to the set $\mathbb{V}$, based at p, a special case of the full Taylor Theorem proved above:

- We have the expansion, valid for any $x \in \mathbb{V}$:
$$f(x) = f(p) + f'(p)(x - p) + g(x)((x - p)^2).$$

  Here $g(x)$ is a smooth function on $\mathbb{V}$ with values in $\mathbb{X}^* \odot \mathbb{X}^*$ and $(x - p)^2 = (x-p) \otimes (x-p) \in \mathbb{X} \otimes \mathbb{X}$, with the property that $g(p) = 2^{-1} f''(p)$.

Let $b \in \mathcal{C}^\infty(\mathbb{U})$ be a bump function, supported in $\mathbb{V}$, such that there exists an open subset $\mathbb{W} \subset \mathbb{V}$, with $p \in \mathbb{W}$, such that $b|_\mathbb{W} = 1$ (the constant function). Again, by work done earlier this term, we know that such a function b exists. Then for any $x \in \mathbb{U}$, put:

- $G(x) = b(x)g(x)$, if $x \in \mathbb{V}$,

- $G(x) = 0$, if $x \notin \mathbb{V}$.

Then G is a smooth function on $\mathbb{U}$ with values in $\mathbb{X}^* \odot \mathbb{X}^*$.

Also, by construction, we have $G(x) = g(x)$, for any $x \in \mathbb{W}$.

Next define a function F by the formula, valid for any $x \in \mathbb{U}$:
$$F(x) = f(p) + f'(p)(x - p) + G(x)(x - p, x - p).$$

Then $F \in \mathcal{C}^\infty(\mathbb{U})$ and by construction, we have $(F - f)|_\mathbb{W} = 0$.

It follows in particular, by the support non-increasing property of v, proved above, that $v(F - f)(p) = 0$.

This gives the relation, using the derivation property of v:

$$v(f)(p) = v(F)(p) - v(F-f)(p) = v(F)(p) = v\left(f(p) + f'(p)(x - p) + G(x)(x - p, x - p)\right)(p)$$
$$= f'(p)(v(x - p))(p) + v(G(x)(x - p, x - p))(p)$$
$$= f'(p)(v(x)(p)) + (v(G(x))(x - p, x - p))(p) + 2(G(x)(v(x - p), x - p))(p)$$
$$= (f'(x)(v(x))(p) + v(G(x))(p)(p - p, p - p) + 2G(p)(v(x)(p), p - p)$$
$$= (f'(x)(v(x))(p) + v(G(x))(p)(0, 0) + 2G(p)(v(x)(p), 0) = (f'(x)(v(x))(p).$$

Since $p \in \mathbb{U}$ is arbitrary, we have $v(f)(x) = f'(x)(v(x))$, for any $x \in \mathbb{U}$ and any $f \in \mathcal{C}^\infty(\mathbb{U})$. So $v = \partial_w$, where $w(x) = v(x)$, for any $x \in \mathbb{U}$ and we are done.

6

We may rephrase the theorem just proved as follows:

- Every derivation V of $\mathbb{C}^\infty(\mathbb{U})$ is given by the formula $V = v(x).\partial$, for some smooth function v defined on $\mathbb{U}$ with values in $\mathbb{X}$. Also $v(x) = V(x)$.

More explicitly, we have, using co-ordinates $(x^1, x^2, \ldots, x^n)$ and their corresponding partial derivative operators $\partial_i = \dfrac{\partial}{\partial x^i}, i = 1, 2, \ldots, n$.

$$V = \sum_{i=1}^{j} v^i \partial_i.$$

Here the coefficients $v^i = V(x^i), i = 1, 2, \ldots, n$ are smooth functions. More generally, if $\{e_i, I = 1, 2, \ldots, k\}$ forms a basis of derivations, we can write:

$$V = \sum_{i=1}^{j} V^i e_i.$$

Here the coefficients V^i are smooth functions. Then if $\{f_i, I = 1, 2, \ldots, k\}$ forms another basis, we have transitions formula:

$$f_i = \sum_{j=1}^{k} m_i^j e_j, \quad e_i = \sum_{j=1}^{k} n_i^j f_j.$$

Here each of m_i^j and n_i^j is a smoothly invertible matrices of smooth functions on $\mathbb{U}$ and each is the inverse of the other:

$$\sum_{t=1}^{k} m_i^k n_k^j = \sum_{t=1}^{k} n_i^k m_k^j = \delta_i^j.$$

The special case where we have two systems of coordinates X and Y for $\mathbb{U}$. By definition these are smooth diffeomorphisms: $X : \mathbb{U} \to \mathbb{X}$ and $Y : \mathbb{U} \to \mathbb{Y}$ where $\mathbb{X}$ and $\mathbb{Y}$ lie in $\mathbb{R}^k$ and k is the dimension of $\mathbb{U}$. Then, given a vector field, V the list of functions $V(X)$ determines V uniquely, as does the list $V(Y)$ and they are connected by the chain rule:

$$V(Y^i) = \sum_{j=1}^{k} J_j^i(X) V(X^j).$$

Here J^i_j is the Jacobian matrix of the transformation. More explicitly, we have Y as a composite function of X and then the Jacobian is the Jacobian matrix of the transformation. For example:

- Let $\mathbb{U} = \{(x, y, z) \in \mathbb{R}^3 : \text{if } y = 0, \text{ then } x > 0\}$

- Let $X^{-1} : (r, t, w) \in \mathbb{R}^+ \times (-\pi, \pi) \times \mathbb{R} \to \mathbb{U}$ be given by the formula $X^{-1}(r, t, w) = (r\cos(t), r\sin(t), w)$.
  In co-ordinates, we have:

$$x = r\cos(t), \quad y = r\sin(t), \quad z = w,$$

$$r = \sqrt{x^2 + y^2}, \quad t = \arctan(yx^{-1}), \quad w = z.$$

- Let Y^{-1} map $(s, p, u) \in \mathbb{R}^+ \times (0, \pi) \times (-\pi, \pi) \to \mathbb{U}$ be given by the formula $Y^{-1}(s, u, p) = (s\cos(u)\sin(p), s\sin(u)\sin(p), s\cos(p))$.
  In co-ordinates, we have:

$$x = s\cos(u)\sin(p), \quad y = s\sin(u)\sin(p), \quad z = s\cos(p),$$

$$s = \sqrt{x^2 + y^2 + z^2}, \quad u = \arctan(yx^{-1}), \quad p = \arctan\left(\left(\frac{x^2 + y^2}{x^2 + y^2 + z^2}\right)^{\frac{1}{2}}\right).$$

Then we have:

$$Y \circ X^{-1}(r, t, w) = Y(r\cos(t), r\sin(t), w) = (s, u, p),$$

$$s = \sqrt{r^2 + w^2}, \quad u = t, \quad p = \arctan\left(\frac{r}{w}\right),$$

$$r = s\sin(p), \quad t = u, \quad w = s\cos(p).$$

The (x, y, z) coordinate basis of derivations $\partial = [\partial_x, \partial_y, \partial_z]$ is related to the (r, t, w) co-ordinate derivations $[\partial_r, \partial_t, \partial_w]$ by the invertible matrix transformations:

$$[\partial_r, \partial_t, \partial_w] = [\partial_x, \partial_y, \partial_z] \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial t} & \frac{\partial x}{\partial w} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial t} & \frac{\partial y}{\partial w} \\ \frac{\partial z}{\partial r} & \frac{\partial z}{\partial t} & \frac{\partial z}{\partial w} \end{vmatrix} = [\partial_x, \partial_y, \partial_z] \begin{vmatrix} \cos(t) & -r\sin(t) & 0 \\ \sin(t) & r\cos(t) & 0 \\ 0 & 0 & 1 \end{vmatrix},$$

$$= [\partial_x, \partial_y, \partial_z] \begin{vmatrix} x(x^2 + y^2)^{-\frac{1}{2}} & -y & 0 \\ y(x^2 + y^2)^{-\frac{1}{2}} & x & 0 \\ 0 & 0 & 1 \end{vmatrix}.$$

8

The inverse transformations are:

$$[\partial_x, \partial_y, \partial_z] = [\partial_r, \partial_t, \partial_w] \begin{vmatrix} \frac{\partial r}{\partial x} & \frac{\partial r}{\partial y} & \frac{\partial r}{\partial z} \\ \frac{\partial t}{\partial x} & \frac{\partial t}{\partial y} & \frac{\partial t}{\partial z} \\ \frac{\partial w}{\partial x} & \frac{\partial w}{\partial y} & \frac{\partial w}{\partial z} \end{vmatrix} = [\partial_r, \partial_t, \partial_w] \begin{vmatrix} x(x^2+y^2)^{-\frac{1}{2}} & y(x^2+y^2)^{-\frac{1}{2}} & 0 \\ -y(x^2+y^2)^{-1} & x(x^2+y^2)^{-1} & 0 \\ 0 & 0 & 1 \end{vmatrix}$$

$$= [\partial_r, \partial_t, \partial_w] \begin{vmatrix} \cos(t) & \sin(t) & 0 \\ -r^{-1}\sin(t) & r^{-1}\cos(t) & 0 \\ 0 & 0 & 1 \end{vmatrix}.$$

Similarly, we have:

$$[\partial_s, \partial_p, \partial_u] = [\partial_x, \partial_y, \partial_z] \begin{vmatrix} \frac{\partial x}{\partial s} & \frac{\partial x}{\partial p} & \frac{\partial x}{\partial u} \\ \frac{\partial y}{\partial s} & \frac{\partial y}{\partial p} & \frac{\partial y}{\partial u} \\ \frac{\partial z}{\partial s} & \frac{\partial z}{\partial p} & \frac{\partial z}{\partial u} \end{vmatrix} = [\partial_x, \partial_y, \partial_z] \begin{vmatrix} \cos(u)\sin(p) & s\cos(u)\cos(p) & -s\sin(u)\sin(p) \\ \sin(u)\sin(p) & s\sin(u)\cos(p) & s\cos(u)\sin(p) \\ \cos(p) & -s\sin(p) & 0 \end{vmatrix}$$

$$= [\partial_x, \partial_y, \partial_z] \begin{vmatrix} x(x^2+y^2+z^2)^{-\frac{1}{2}} & zx(x^2+y^2)^{-\frac{1}{2}} & -y \\ y(x^2+y^2+z^2)^{-\frac{1}{2}} & zy(x^2+y^2)^{-\frac{1}{2}} & x \\ z(x^2+y^2+z^2)^{-\frac{1}{2}} & -(x^2+y^2)^{\frac{1}{2}} & 0 \end{vmatrix}.$$

The inverse transformations are:

$$[\partial_x, \partial_y, \partial_z] = [\partial_s, \partial_p, \partial_u] \begin{vmatrix} \frac{\partial s}{\partial x} & \frac{\partial s}{\partial y} & \frac{\partial s}{\partial z} \\ \frac{\partial p}{\partial x} & \frac{\partial p}{\partial y} & \frac{\partial p}{\partial z} \\ \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} & \frac{\partial u}{\partial z} \end{vmatrix}$$

$$= [\partial_s, \partial_p, \partial_u] \begin{vmatrix} x(x^2+y^2+z^2)^{-\frac{1}{2}} & y(x^2+y^2+z^2)^{-\frac{1}{2}} & z(x^2+y^2+z^2)^{-\frac{1}{2}} \\ zx(x^2+y^2)^{-\frac{1}{2}}(x^2+y^2+z^2)^{-1} & yz(x^2+y^2)^{-\frac{1}{2}}(x^2+y^2+z^2)^{-1} & -(x^2+y^2)^{\frac{1}{2}}(x^2+y^2+z^2)^{-1} \\ -y(x^2+y^2)^{-1} & x(x^2+y^2)^{-1} & 0 \end{vmatrix}.$$

Similarly, we have the inverse pair of transformations:

$$[\partial_r, \partial_t, \partial_w] = [\partial_s, \partial_p, \partial_u] \begin{vmatrix} \frac{\partial s}{\partial r} & \frac{\partial s}{\partial t} & \frac{\partial s}{\partial w} \\ \frac{\partial p}{\partial r} & \frac{\partial p}{\partial t} & \frac{\partial p}{\partial w} \\ \frac{\partial u}{\partial r} & \frac{\partial u}{\partial t} & \frac{\partial u}{\partial w} \end{vmatrix} = [\partial_s, \partial_p, \partial_u] \begin{vmatrix} r(r^2+w^2)^{-\frac{1}{2}} & 0 & w(r^2+w^2)^{-\frac{1}{2}} \\ w(r^2+w^2)^{-1} & 0 & -r(r^2+w^2)^{-1} \\ 0 & 1 & 0 \end{vmatrix}$$

$$= [\partial_s, \partial_p, \partial_u] \begin{vmatrix} \sin(p) & 0 & \cos(p) \\ \cos(p)s^{-1} & 0 & -\sin(p)s^{-1} \\ 0 & 1 & 0 \end{vmatrix},$$

$$[\partial_s, \partial_p, \partial_u] = [\partial_r, \partial_t, \partial_w] \begin{vmatrix} \frac{\partial r}{\partial s} & \frac{\partial r}{\partial p} & \frac{\partial r}{\partial u} \\ \frac{\partial t}{\partial s} & \frac{\partial t}{\partial p} & \frac{\partial t}{\partial u} \\ \frac{\partial w}{\partial s} & \frac{\partial w}{\partial p} & \frac{\partial w}{\partial u} \end{vmatrix} = [\partial_r, \partial_t, \partial_w] \begin{vmatrix} \sin(p) & s\cos(p) & 0 \\ 0 & 0 & 1 \\ \cos(p) & -s\sin(p) & 0 \end{vmatrix} = [\partial_r, \partial_t, \partial_w] \begin{vmatrix} r(r^2+w^2)^{-\frac{1}{2}} & w & 0 \\ 0 & 0 & 1 \\ w(r^2+w^2)^{-\frac{1}{2}} & -r & 0 \end{vmatrix}.$$

It may be checked that these matrix transformations are compatible.

Finally note that smooth vector fields have the sheaf property: if we have an open subset $\mathbb{V}$ of $\mathbb{U}$, then a smooth vector field V on $\mathbb{U}$ restricts naturally to a vector field, denoted $V|_{\mathbb{V}}$, on $\mathbb{V}$, simply by restricting the coefficient functions v^i.

More invariantly, if we are given $f \in \mathcal{C}^\infty(\mathbb{V})$ and $v \in \mathcal{T}(\mathbb{U})$ and we wish to compute $v|_{\mathbb{V}}(f)$, we can do this point-wise as follows:

Given $x \in \mathbb{V}$, pick a smooth bump function b_x on $\mathbb{U}$ which has the constant value 1 in an open set containing x and is zero outside $\mathbb{V}$ (we know how to do this). Then the product function $g = fb_x$ is smooth on $\mathbb{V}$ and extends naturally, by zero, to a smooth function on $\mathbb{U}$, still called g. Then $V(g)$ is then well-defined as a smooth function on $\mathbb{U}$. Now put $(V|_{\mathbb{V}}(f))(x) = (v(g))(x)$. It is easy to see that this number is independent of the various choices involved. As x varies in $\mathbb{V}$ this gives the required smooth function $(V|_{\mathbb{V}}(f))$ on $\mathbb{V}$ and then the map $f \to (V|_{\mathbb{V}}(f))$ for $f \in \mathcal{C}^\infty(\mathbb{V})$ is the required restriction of the derivation V. Finally this restriction operation makes the collection of derivations into a sheaf of $\mathcal{C}^\infty$ modules over the space $\mathbb{M}$.

In particular we can assign a "value" to the vector field V at a point p: if $V = \sum_{i=1}^k v^i(x)\partial_i$, then its value at p is $V_p = \sum_{i=1}^k v^i(p)\partial_i$. This is called a tangent vector at p. The space of all such tangent vectors at p is a vector space of dimension k, denoted $\mathcal{T}_p$. Note that a tangent vector acts on smooth functions by $V_p(f) = V(f)(p)$. This can be shown easily to be independent of the choices made and depends only on the image of f in the space $\mathcal{T}_p^* = \mathcal{Z}_p/\mathcal{Z}_p^2$ (called the co-tangent space at p), where $\mathcal{Z}_p$ is the space of smooth functions defined near p which vanish at p and $\mathcal{Z}_p^2$ is the space spanned by all squares of elements of $\mathcal{Z}_p$. Then the image of f in $\mathcal{T}_p^*$ is called $(df)_p$ and we have:

$$V_p = \sum_{i=1}^k v^i(p)\partial_i, \quad (df)_p = \sum_{i=1}^k (\partial_i f)(p)dx^i,$$

$$V_p(f) = \sum_{i=1}^k v^i(p)(\partial_i f)(p).$$

So the tangent space $\mathcal{T}_p$ and the co-tangent space $\mathcal{T}_p^*$ at p are dual.

The Lie bracket

For $v \in \mathcal{T}(\mathbb{U})$ and $w \in \mathcal{T}(\mathbb{U})$ and $f \in \mathcal{C}^\infty(\mathbb{U})$ define:

$$[v,w](f) = v(w(f)) - w(v(f)).$$

As f varies this gives a map $[v,w] : \mathcal{C}^\infty(\mathbb{U}) \to \mathcal{C}^\infty(\mathbb{U})$.

- Then $[v,w] \in \mathcal{T}(U)$.

The vector field $[v,w]$ is called the Lie bracket of the vector fields v and w.

Proof:
The only non-trivial property to be checked is the Liebniz rule.
So let f and g be smooth functions.
Then we have:

$$[v,w](fg) = v(w(fg)) - (w(v(fg))$$
$$= v(w(f)g + fw(g)) - (w(v(f)g + fv(g))$$
$$= v(w(f)g) + v(fw(g)) - w(v(f)g) - w(fv(g))$$
$$= v(w(f))g + w(f)v(g) + v(f)w(g) + fv(w(g)) - w(v(f))g - v(f)w(g) - w(f)v(g) - fw(v(g))$$
$$= (v(w(f)) - w(v(f)))g + f(v(w(g)) - w(v(g)))$$
$$= ([v,w](f))g + f[v,w](g).$$

From its definition, we see that the Lie bracket is skew symmetric and linear in each argument over the reals:

$$[v,w] = -[w,v],$$
$$[v+x, w+y] = [v,w] + [x,w] + [v,y] + [x,y],$$
$$[av, bw] = ab[v,w].$$

Here v, w, x and y are in $\mathcal{T}(U)$ and a and b are real constants.
Also we have a Liebniz rule for multiplication by functions:

$$[fv, gw] = fg[v,w] + fv(g)w - gw(f)v.$$

This is valid for any f and g in $\mathcal{C}^\infty(\mathbb{U})$ and any v and w in $\mathcal{T}(\mathbb{U})$.
The proof is direct calculation: for any $h \in \mathcal{C}^\infty(\mathbb{U})$, we have:

$$[fv, gw](h) = fv(gw(h)) - gw(fv(h)) = fv(g)w(h) + fgv(w(h)) - gw(f)v(h) - gfw(v(h))$$
$$= fg(v(w(h)) - w(v(h))) + fv(g)w(h) - gw(f)v(h)$$
$$= (fg[v,w] + fv(g)w - gw(f)v)(h).$$

Using this rule we see that if $v = \sum_{i=1}^{n} v^i \partial_i$ and $w = \sum_{j=1}^{n} w^j \partial_j$, where the $\partial_i, i = 1, 2, \ldots, n$ are the partial derivative derivations and where the coefficient functions v^i and w^j are smooth on $\mathbb{U}$, then we have:

$$[v, w] = \sum_{k=1}^{n} z^k \partial_k,$$

$$z^k = v(w^k) - w(v^k) = \sum_{i=1}^{n} \left(v^i \partial_i w^k - w^i \partial_i v^k \right).$$

Note that the partial derivative derivations commute:

$$[\partial_i, \partial_j] = 0.$$

Finally, we have the Jacobi identity for any three vector fields v, w and x in $\mathcal{T}(\mathbb{U})$:
$$[v, [w, x]] + [w, [x, v]] + [x, [v, w]] = 0.$$

Again this is checked by direct calculation.

For any $h \in \mathcal{C}^\infty(\mathbb{U})$, we have:

$$([v, [w, x]] + [w, [x, v]] + [x, [v, w]])(h) = [v, [w, x]](h) + [w, [x, v]](h) + [x, [v, w]](h)$$

$$= v([w, x](h)) - [w, x](v(h)) + w([x, v](h)) - [x, v](w(h)) + x([v, w](h)) - [v, w](x(h))$$

$$= v(w(x(h)) - x(w(h))) - (w(x(v(h))) - x(w(v(h)))) + w(x(v(h)) - v(x(h)))$$

$$- (x(v(w(h))) - v(x(w(h)))) + x(v(w(h)) - w(v(h))) - (v(w(x(h))) - (w(v(x(h)))))$$

$$= v(w(x(h))) - v(x(w(h))) - w(x(v(h))) + x(w(v(h))) + w(x(v(h))) - w(v(x(h)))$$

$$- x(v(w(h))) + v(x(w(h))) + x(v(w(h))) - x(w(v(h))) - v(w(x(h))) + w(v(x(h))) = 0.$$

For each $v \in \mathcal{T}(\mathbb{U})$, we define the Lie derivative operator $\mathcal{L}_v$ acting on smooth functions and on vector fields, by the formulas, valid for any vector field $w \in \mathcal{T}(\mathbb{U})$ and any smooth function f on $\mathbb{U}$:

- $\mathcal{L}_v f = v(f)$,

- $\mathcal{L}_v w = [v, w]$.

Then we have, rewriting the above formulas, valid for any f and g in $\mathcal{C}^\infty(\mathbb{U})$ and any smooth vector fields v, w and x on $\mathbb{U}$ and for any real constant function c on $\mathbb{U}$:

$$\mathcal{L}_{fv} = f\mathcal{L}_v,$$

$$\mathcal{L}_{v+w} = \mathcal{L}_v + \mathcal{L}_w,$$

$$\mathcal{L}_v w = -\mathcal{L}_w v,$$

$$\mathcal{L}_v c = 0,$$

$$\mathcal{L}_v(f+g) = \mathcal{L}_v f + \mathcal{L}_v g,$$

$$\mathcal{L}_v(fg) = (\mathcal{L}_v f)g + f\mathcal{L}_v g,$$

$$\mathcal{L}_v(w+x) = \mathcal{L}_v w + \mathcal{L}_v x,$$

$$\mathcal{L}_v(fw) = (\mathcal{L}_v f)w + f\mathcal{L}_v w,$$

$$\mathcal{L}_v\mathcal{L}_w - \mathcal{L}_w\mathcal{L}_v = \mathcal{L}_{[v,w]}.$$

The contravariant tensor algebra

The contravariant tensor algebra of $\mathbb{U}$ is the (associative) tensor algebra denoted $\mathcal{T}(\mathbb{U})$, generated by $\mathcal{T}^0(\mathbb{U}) = \mathcal{C}^\infty(\mathbb{U})$ and by $\mathcal{T}^1(\mathbb{U})$, the space of smooth vector fields on $\mathbb{U}$ (note the change of notation here, the space $\mathcal{T}^1(\mathbb{U})$, by itself, was previously called $\mathcal{T}(\mathbb{U})$), subject to the relations that multiplication of tensors is linear in each argument, over the ring $\mathcal{C}^\infty(\mathbb{U})$. So each tensor is a (finite) sum of words $v = fv_1 v_2 \ldots v_k$ (called a word of k letters), with $f \in \mathcal{C}^\infty(\mathbb{U})$ and $v_i \in \mathcal{T}^1(\mathbb{U}), i = 1, 2, \ldots, k$, for k a non-negative integer, with the product of the word v with the word $w = gw_1 w_2 \ldots w_m$, with $g \in \mathcal{C}^\infty(\mathbb{U})$ and $w_i \in \mathcal{T}^1(\mathbb{U}), i = 1, 2, \ldots, m$, for m a non-negative integer, being the $k + m$ word $x = vw = hx_1 x_2 \ldots x_{k+m}$, where $h = fg$ and $x_i = v_i$, for $1 \leq i \leq k$ and $x_i = w_{i-k}$, for $k + 1 \leq i \leq k + m$. The subspace of $\mathcal{T}(\mathbb{U})$ spanned by words of k letters, for a fixed k is denoted $\mathcal{T}^k(\mathbb{U})$ and its elements are called tensors of type k. Then the product maps $\mathcal{T}^p(\mathbb{U}) \times \mathcal{T}^q(\mathbb{U})$ into $\mathcal{T}^{p+q}(\mathbb{U})$, for any nonnegative integers p and q.

In co-ordinates, every tensor τ of $\mathcal{T}^p(\mathbb{U})$ can be written uniquely:

$$\tau = \sum_{i_1=1}^n \sum_{i_2=1}^n \cdots \sum_{i_p=1}^n \tau^{i_1 i_2 \ldots i_p}(x)\partial_{i_1}\partial_{i_2}\ldots\partial_{i_p}.$$

13

Here each coefficient $\tau^{i_1 i_2 \ldots i_p}(x)$ is a smooth function on $\mathbb{U}$. To avoid confusion with differential operators, it is traditional to introduce the notation $\otimes$ for the tensor product, so τ is then written:

$$\tau = \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \cdots \sum_{i_p=1}^{n} \tau^{i_1 i_2 \ldots i_p}(x) \partial_{i_1} \otimes \partial_{i_2} \otimes \cdots \otimes \partial_{i_p}.$$

So the n^p words $\partial_{i_1} \otimes \partial_{i_2} \otimes \cdots \otimes \partial_{i_p}$ with $1 \leq i_j \leq n$, for each $j = 1, 2, \ldots, p$, form a basis for $\mathcal{T}^p(\mathbb{U})$, over the ring $\mathcal{C}^\infty(\mathbb{U})$, called the coordinate basis.

The symmetric tensor algebra $\mathcal{ST}(\mathbb{U})$ is the algebra obtained from the tensor algebra, by requiring that all letters commute, so the multiplication in this algebra is commutative and indeed is a standard commutative algebra of polynomials in n-variables over $\mathcal{C}^\infty(\mathbb{U})$. So there is a surjective algebra homomorphism from $\mathcal{T}(\mathbb{U})$ to $\mathcal{ST}(\mathbb{U})$. which maps each word to itself.

For clarity the symmetric tensor product is often represented by the notation $\odot$. The image of the co-ordinate basis for $\mathcal{T}^p(\mathbb{U})$, gives the following coordinate basis of words for the image $\mathcal{ST}^p(\mathbb{U})$ of $\mathcal{T}^p(\mathbb{U})$:

$$\partial_{i_1} \odot \partial_{i_2} \odot \cdots \odot \partial_{i_p}, \text{ for } 1 \leq i_1 \leq i_2 \cdots \leq i_p \leq n.$$

This basis has $\binom{n+p-1}{p}$ elements.

The skew-symmetric tensor algebra $\Omega\mathcal{T}(\mathbb{U})$ is the algebra obtained from the tensor algebra, by requiring that all letters representing vector fields anti-commute, so the multiplication in this algebra is graded-commutative: words with an even number of letters commute with everything, words with an odd number of letters anti-commute with each other. There is a surjective algebra homomorphism from $\mathcal{T}(\mathbb{U})$ to $\Omega\mathcal{T}(\mathbb{U})$, which maps each word to itself.

For clarity the skew-symmetric tensor product is often represented by the notation $\wedge$. The image of the co-ordinate basis for $\mathcal{T}^p(\mathbb{U})$, gives the following coordinate basis for the image $\mathcal{ST}^p(\mathbb{U})$ of $\mathcal{T}^p(\mathbb{U})$:

$$\partial_{i_1} \wedge \partial_{i_2} \wedge \cdots \wedge \partial_{i_p}, \text{ for } 1 < i_1 < i_2 \cdots < i_p < n.$$

This basis has $\binom{n}{p}$ elements.

In particular, we have $\Omega\mathcal{T}^p(\mathbb{U}) = 0$ if $p > n$ and the total dimension of $\Omega\mathcal{T}(\mathbb{U})$ over $\mathcal{C}^\infty(\mathbb{U})$ is finite and is $\sum_{p=0}^{n} \binom{n}{p} = 2^n$.

If v is a vector field, the Lie derivative along v, $\mathcal{L}_v$ extends naturally to $\mathcal{T}(\mathbb{U})$ such that the Lie derivative acts on vector fields and smooth functions as given above and for any tensors τ and υ any smooth vector fields v and w and any smooth function f on $\mathbb{U}$, we have:

$$\mathcal{L}_v(\tau + \upsilon) = \mathcal{L}_v(\tau) + \mathcal{L}_v(\upsilon),$$

$$\mathcal{L}_v(\tau\upsilon) = \mathcal{L}_v(\tau)\upsilon + \tau\mathcal{L}_v(\upsilon),$$

$$\mathcal{L}_{v+w} = \mathcal{L}_v + \mathcal{L}_w,$$

$$\mathcal{L}_{fv} = f\mathcal{L}_v,$$

$$\mathcal{L}_v\mathcal{L}_w - \mathcal{L}_w\mathcal{L}_v = \mathcal{L}_{[v,w]}.$$

Acting on the word $w = gw_1w_2 \ldots w_k$, we have:

$$\mathcal{L}_v w = v(g)w_1w_2 \ldots w_k + g[v, w_1]w_2 \ldots w_k + gw_1[v, w_2] \ldots w_k + \cdots + gw_1w_2 \ldots [v, w_k].$$

In particular $\mathcal{L}_v$ preserves the type of a tensor. This action is compatible with the passages to the symmetric and skew symmetric tensor algebras, so the Lie derivative also acts naturally on these algebras obeying the same rules as above.

Finally note that there are natural collections of tensors that maps isomorphically to each of the symmetric and skew symmetric algebras.

- In the case of the symmetric algebra it is the space spanned by all words of the form fw^p, for f smooth and w a vector field.

More explicitly there is an idempotent symmetrization map which we describe as follows. Recall the permutation group $\mathbb{S}_n$ defined for any given $n \in \mathbb{N}$ as the group of bijections from $\mathbb{N}_n = \{1, 2, \ldots, n\} \subset \mathbb{N}$ to itself, with the group multiplication being composition. Then $\mathbb{S}_n$ has $n!$ elements.

Given a word $w = fw_1w_2 \ldots w_p$ in $\mathcal{T}^p(\mathbb{U})$ as above and given a permutation σ in $\mathbb{S}_p$, we define:

$$w_\sigma = fw_{\sigma(1)}w_{\sigma(2)} \ldots w_{\sigma(p)}.$$

Note that if ρ is another permutation in $\mathbb{S}_p$, we have:

$$(w_\sigma)_\rho = fw_{\rho(\sigma(1))}w_{\rho(\sigma(2))} \ldots w_{\rho(\sigma(p))}$$

$$= fw_{(\rho \circ \sigma)(1))}w_{(\rho \circ \sigma)(2))} \ldots w_{\rho \circ \sigma)(p))}$$

$$= w_{\rho \circ \sigma}.$$

Then the assignment $w \to w_\sigma$ for each word extends naturally to give a homomorphism, $\tau \to \tau_\sigma$ from $\mathcal{T}^p(\mathbb{U})$ to itself and we have $(\tau_\sigma)_\rho = \tau_{\rho \circ \sigma}$, for any permutations ρ and σ in $\mathbb{S}_p$.

Then given any tensor $\tau \in \mathcal{T}^p(\mathbb{U})$, we define:

$$S(\tau) = \frac{1}{p!} \sum_{\sigma \in \mathbb{S}_p} \tau_\sigma.$$

This gives an endomorphims of $\mathcal{T}^p(\mathbb{U})$. We prove that it is idempotent:

$$S^2(\tau) = S\left(\frac{1}{p!}\sum_{\sigma\in\mathbb{S}_p}\tau_\sigma\right) = \frac{1}{p!}\sum_{\rho\in\mathbb{S}_p}\sum_{\sigma\in\mathbb{S}_p}(\tau_\sigma)_\rho$$

$$= \frac{1}{p!}\sum_{\rho\in\mathbb{S}_p}\sum_{\sigma\in\mathbb{S}_p}\tau_{\rho\circ\sigma}$$

$$= \frac{1}{p!}\sum_{\rho\in\mathbb{S}_p}\sum_{\sigma\in\mathbb{S}_p}\tau_{\rho\circ(\rho^{-1}\circ\sigma)}$$

$$= \frac{1}{p!}\sum_{\rho\in\mathbb{S}_p}\sum_{\sigma\in\mathbb{S}_p}\tau_\sigma$$

$$= \sum_{\sigma\in\mathbb{S}_p}\tau_\sigma = S(\tau).$$

Here we used the fact that as σ runs once through all permutations in $\mathbb{S}_p$, so does $\rho^{-1}\circ\sigma$, for each fixed permutation ρ in $\mathbb{S}_p$. The image of S is then called the space of symmetric tensors. On this space we can define the symmetric tensor product by the formula $A\odot B = S(AB)$. This makes the space of symmetric tensors into an algebra naturally isomorphic to the algebra $\mathcal{ST}(\mathbb{U})$.

- In the case of the skew symmetric tensors algebra, we use the fact, easily proved by induction, that there is a group epimorphism s_n of the symmetric group $\mathbb{S}_n$ to $\mathbb{Z}_2$, the multiplicative group of two elements 1 and -1, with 1 the identity and $(-1)^2 = 1$. For a permutation σ, the quantity $s_n(\sigma)$ is called its sign and σ is called even if its sign is 1 and odd otherwise. A simple transposition is a permutation which fixes all but two element of $\mathbb{N}_n$. Every permutation is a composition of simple transpositions and $s_n(\sigma)$ is then even if and only if there are an even number of terms in this composition.

Derivations of differential forms

Given a smooth function f on $\mathbb{U}$, its

The algebra of forms $\Omega(\mathbb{U})$ is the graded algebra over $\mathcal{C}^\infty(\mathbb{U})$ generated by the zero-forms $\mathcal{C}^\infty(\mathbb{U})$ and by exact differential one forms df, for $f \in \mathcal{C}^\infty(\mathbb{U})$,

such that the map $d : \mathcal{C}^\infty(\mathbb{U}) \to \Omega^1(\mathbb{U})$, $f \to df$, for any $f \in \mathcal{C}^\infty(\mathbb{U})$, is a derivation. Also we have $(df)(dg) = -(dg)(df)$, for any f and g in $\mathcal{C}^\infty(\mathbb{U})$.

A derivation v of $\Omega(\mathbb{U})$ is by definition a map $v : \Omega(\mathbb{U}) \to \Omega(\mathbb{U})$, such that:

- v kills constants, $v(c) = 0$, for any constant function c on $\mathbb{U}$.

- v is an abelian group homomorphism: $v(\alpha + \beta) = v(\alpha) + v(\beta)$, for any α and β in $\Omega(\mathbb{U})$.

- v obeys a Liebniz rule: for any integers k and l and any α and β in $\Omega(\mathbb{U})$, where α has degree k and β degree l, we have:

$$v(\alpha\beta) = v(\alpha)\beta + (-1)^{kl}v(\beta)\alpha.$$

The operator $d : \mathcal{C}^\infty(\mathbb{U}) \to \Omega^1(\mathbb{U})$ extends naturally and uniquely to a derivation of $\Omega(\mathbb{U})$, still called d, such that $d^2 = 0$.
Then d is called the exterior derivative.

Denote by $\mathcal{D}(\mathbb{U})$ the space of all derivations of $\Omega(\mathbb{U})$.
Note that if v is in $\mathcal{D}(\mathbb{U})$, then so is γv, the multiplication of v on the left by any element γ of $\Omega(\mathbb{U})$, so $\mathcal{D}(\mathbb{U})$ is naturally an $\Omega(\mathbb{U})$-module.

A derivation v is said to be local if and only if it annihilates $\Omega^0(\mathbb{U})$.
Define the $\mathbb{X}^*$-valued derivation δ of $\Omega(\mathbb{U})$ by the formula, valid for any smooth function f:
$$\delta(f) = 0, \quad \delta(df) = f'.$$
Then δ is a derivation of degree minus one; also every local derivation V of $\Omega(\mathbb{U})$ has the unique expression:

$$V = v.\delta.$$

Let V be any derivation of $\Omega(\mathbb{U})$. Then its action on $C^\infty(\mathbb{U})$ is that of a derivation so it may be given by the formula:

$$V(f) = v.f'.$$

Here v is a unique smooth form on $\mathbb{U}$, with values in $\mathbb{X}$.
Put $W = v.\delta$, so W is a local derivation. Then we have, since W annihilates functions, so also does its even and odd parts:

$$(V - [W, d])(f) = v.f' - v.\delta(df) = 0.$$

So we have the unique decomposition of any derivation:

$$V = [W, d] + T, \text{with } T \text{ and } W \text{ local derivations.}$$

Notes on the theorem of Archimedes, Newton, Green, Stokes, Ostragodsky, Poincaré, de Rham, etc.

The theorem is roughly stated as follows:

- Let $\mathcal{R}$ be an oriented smooth n-manifold with boundary $\partial\mathcal{R}$, at least piecewise smooth, where $\partial\mathcal{R}$ acquires the induced orientation from that of $\mathcal{R}$.

  Let ω be an $(n-1)$-form, smoothly defined on $\mathcal{R}$ and smooth up to and including the boundary $\partial\mathcal{R}$.

  Then we have:
  $$\int_{\mathcal{R}} d\omega = \int_{\partial\mathcal{R}} \omega.$$

For us these manifolds will always be embedded inside some Euclidean space. A fairly simple case and, as we shall see, essentially the base case for a reverse induction proof of the general case, is when $\mathcal{R}$ is a bounded region in $\mathbb{R}^n$, with a smooth boundary, $\partial\mathcal{R}$, given by a single equation $f(x) = 0$, where f is a smooth function on $\mathbb{R}^n$, such that $\mathcal{R}$ is the region given by the inequality $f(x) \leq 0$ and on the boundary the differential $df \neq 0$ (by the rank theorem this guarantees that the boundary is a smooth manifold). Also $x = (x^1, x^2, \ldots x^n)$, where $x^j \in \mathbb{R}$, for each $j = 1, 2, \ldots, n$. The gradient vector ∇f is then non-vanishing on $\partial\mathcal{R}$ and is everywhere outward pointing. Denote by N the unit outward pointing, normal vector, so the unit vector in the direction of ∇f, so we have the relations:

$$\nabla f = N\sqrt{(\nabla f).(\nabla f)}, \quad N.N = 1.$$

In the following we will often have occasion to divide by a partial derivative of f with respect to a coordinate: one or more of these derivatives may vanish at a point, but always there is at least one that is non-vanishing, so the relevant formula will be understood to apply only to them.

Introduce the volume n-form:

$$\Sigma = dx^1 dx^2 \ldots dx^n.$$

Note (though we will not really need this) that we can write this invariantly as:

$$\theta^{a_1} \theta^{a_2} \ldots \theta^{a_n} = \Sigma \epsilon^{a_1 a_2 \ldots a_n}.$$

Here θ^a gives the canonical isomorphism between dual vectors and one-forms: $\alpha_a \to \alpha_a \theta^a$. Also the θ^a anti-commute: $\theta^a \theta^b = -\theta^b \theta^a$. Then $\epsilon^{a_1 a_2 \ldots a_n}$ is the orientation tensor and is totally skew. Dual to θ^b is the derivation δ_a of degree minus one, which annihilates functions and obeys $\delta_a \theta^b = \delta_a^b$. More explicitly, the components of δ_a give n derivations, each of degree minus one, $\underline{\delta} = (\delta_1, \delta_2, \ldots, \delta_n)$, where $\delta_i(dx^j) = \delta_i^j$. Then put:

$$\underline{\Sigma} = \underline{\delta}\Sigma = (\delta_1 \Sigma, \delta_2 \Sigma, \ldots, \delta_n \Sigma)$$

$$= \left(dx^2 dx^3 dx^4 \ldots dx^n, \ -dx^1 dx^3 dx^4 \ldots dx^n, \ dx^1 dx^2 dx^4 \ldots dx^n, \ \ldots, \ (-1)^{n-1} dx^1 dx^2 \ldots dx^{n-1} \right).$$

For example:

- When $n = 1$, we have $\underline{\Sigma} = (1)$.

- When $n = 2$, with Cartesian coordinates (x, y), we have:

$$\underline{\Sigma} = (dy, \ -dx).$$

- When $n = 3$, with Cartesian coordinates (x, y, z), we have:

$$\underline{\Sigma} = (dy\,dz, \ dz\,dx, \ dx\,dy).$$

- When $n = 4$, with Cartesian coordinates (t, x, y, z), we have:

$$\underline{\Sigma} = (dx\,dy\,dz, \ -dt\,dy\,dz, \ -dt\,dz\,dx, \ -dt\,dx\,dy).$$

Notice that we have the identity:

$$\theta^a \Sigma_b = \delta_b^a \Sigma.$$

Proof:
Using the derivation property of δ_b, we have:

$$\theta^a \Sigma_b = \theta^a \delta_b \Sigma = -\delta_b(\theta^a \Sigma) + (\delta_b \theta^a)\Sigma = \delta_b^a \Sigma.$$

Here we used also that $\theta^a \Sigma = 0$, which is true because the left-hand side is an $(n+1)$-form, which automatically vanishes in n dimensions.

For example, we have in the case $n = 3$:

$$dx\,\underline{\Sigma} = dx\,(dy\,dz,\,dz\,dx,\,dx\,dy) = (\Sigma, 0, 0),$$

$$dy\,\underline{\Sigma} = dy\,(dy\,dz,\,dz\,dx,\,dx\,dy) = (0, \Sigma, 0),$$

$$dz\,\underline{\Sigma} = dz\,(dy\,dz,\,dz\,dx,\,dx\,dy) = (0, 0, \Sigma).$$

Now we may write the smooth $(n-1)$-form ω in terms of the basis of $(n-1)$-forms given by $\underline{\Sigma}$ as $\omega = v.\underline{\Sigma} = \sum_{i=1}^{n} v^i \Sigma_i$, where $v = \sum_{i=1}^{n} v^i \partial_i$ is a smooth vector field on $\mathcal{R}$. Then, since $d\Sigma_i = 0$, $dx^i \Sigma_j = \delta_j^i \Sigma$ and $d = \Sigma_{i=1}^{n} dx^I \partial_i$, we have:

$$d\omega = d\left(\sum_{i=1}^{n} v^i \Sigma_i\right) = \sum_{i=1}^{n} (dv^i)\Sigma_i = \sum_{i,j=1}^{n} dx^i(\partial_i v^j)\Sigma_j = \sum_{i,j=1}^{n} \delta_j^i(\partial_i v^j)\Sigma = (\nabla.v)\Sigma.$$

So the Stokes' Theorem can be rewritten as the divergence theorem:

$$\int_{\partial\mathcal{R}} v.\underline{\Sigma} = \int_{\mathcal{R}} (\nabla.v)\Sigma.$$

Now, restricted to the space $\partial\mathcal{R}$, any $(n-1)$-forms should be linearly dependent, since the space of $(n-1)$-forms in $(n-1)$ dimensions is one-dimensional. In terms of the differentials of the co-ordinates, we have the non-trivial linear relation on restriction to the surface $f = 0$:

$$0 = df = (\nabla f).\theta = f_1 dx^1 + f_2 dx^2 + \cdots + f_n dx^n.$$

This allows us to reduce the space of differentials to a space of dimension $n-1$. Indeed, consider the $(n-1)$-form Σ_i at a point where the differential df is non-zero. Then we can consider the differential df as the first element of a basis of forms at that point, so we have the expression $\Sigma_i = \lambda_i \beta + (df)\mu_i$, for some $(n-1)$-form β and some $(n-2)$-forms, μ_i and some scalars λ_i (here β is the product of the other forms of the basis). Multiplying this equation by df we get the relation:

$$(df)\Sigma_i = \lambda_i(df)\beta = \Sigma_{j=1}^{n} f_j dx^j \Sigma_i = f_i \Sigma.$$

But the form $\underline{\Sigma} \neq 0$, so at least one of the $f_i\Sigma$ is not zero, so neither can the form $(df)\beta$ be zero, so we get $\lambda_i = kf_i = mN_i$, for some non-zero scalars k and m. Put $\alpha = m\beta$ giving the relation:

$$\underline{\Sigma}_i = N_i\alpha + (df)\mu_i$$

In particular, on restriction to the hypersurface $f = 0$, since then $df = 0$, we get the relation, for some $(n-1)$-form α:

$$\underline{\Sigma} = N\alpha.$$

Taking the dot product of this equation with the unit vector N we get $\alpha = N.\underline{\Sigma}$. Accordingly, put $\alpha = N.\underline{\Sigma}$. Then restricted to the surface $f = 0$, we have the relation:

$$\underline{\Sigma} = N\alpha = \frac{(\nabla f)\alpha}{\sqrt{(\nabla f).(\nabla f)}}.$$

This gives the relations, valid on restriction to the boundary surface $f = 0$:

$$\alpha = \frac{\Sigma_i\sqrt{(\nabla f).(\nabla f)}}{\partial_i f}, \quad \underline{\Sigma}_i = \frac{(\partial_i f)(\nabla f).\underline{\Sigma}}{(\nabla f).(\nabla f)}.$$

So, for example, in two dimensions on $f = 0$, we have:

$$\alpha = N.\underline{\Sigma} = \frac{(f_x, f_y).(dy, -dx)}{\sqrt{f_x^2 + f_y^2}} = \frac{f_x dy - f_y dx}{\sqrt{f_x^2 + f_y^2}} = \frac{dy\sqrt{f_x^2 + f_y^2}}{f_x} = -\frac{dx\sqrt{f_x^2 + f_y^2}}{f_y}.$$

Here we used that on the curve $f = 0$, we have $df = 0$, so we have:

$$f_x dx + f_y dy = 0, \quad \frac{dx}{f_y} = -\frac{dy}{f_x}, \quad \frac{dy}{dx} = -\frac{f_x}{f_y}, \quad \frac{dx}{dy} = -\frac{f_y}{f_x}.$$

In particular we have that the measures $ds = \dfrac{|dx|\sqrt{f_x^2 + f_y^2}}{|f_y|}$ and $ds = \dfrac{|dy|\sqrt{f_x^2 + f_y^2}}{|f_x|}$ are equal and each gives the standard measure of length along the curve:

$$ds^2 = dx^2 + dy^2 = dx^2\left(1 + \left(\frac{dy}{dx}\right)^2\right) = dx^2\left(1 + \left(\frac{f_x}{f_y}\right)^2\right)$$

$$= \frac{dx^2}{f_y^2}\left(f_x^2 + f_y^2\right) = \frac{dy^2}{f_x^2}\left(f_x^2 + f_y^2\right),$$

$$|\alpha| = ds.$$

23

Now consider the proof of the divergence theorem in two dimensions, where we put $v = [P, Q]$:

$$\int_{\partial \mathcal{R}} v.\underline{\Sigma} = \int_{\mathcal{R}} (\nabla.v)\Sigma,$$

$$\int_{\partial \mathcal{R}} [P, Q].[dy, -dx] = \int_{\mathcal{R}} (P_x + Q_y)dxdy,$$

$$\int_{\partial \mathcal{R}} (Pdy - Qdx) = \int_{\mathcal{R}} (P_x + Q_y)dxdy,$$

This is linear over the reals in the functions P and Q, so to prove the theorem, it suffices to prove the following two relations:
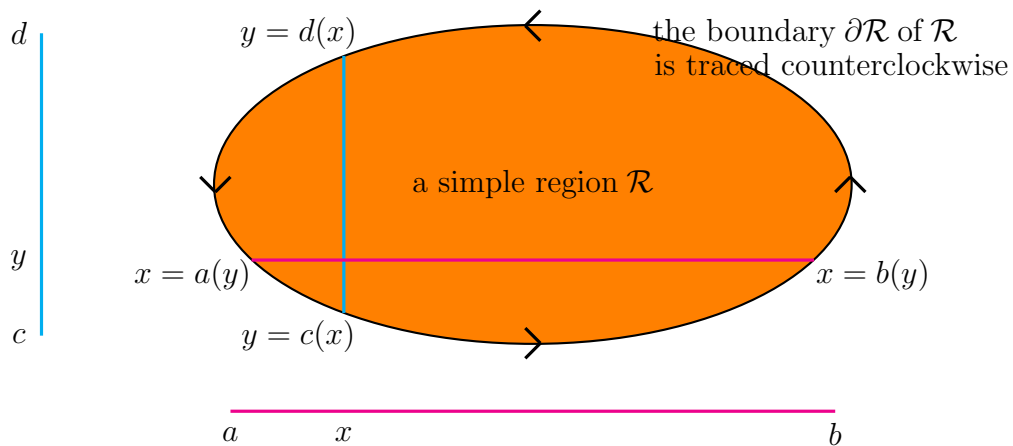
$$\int_{\partial \mathcal{R}} Pdy = \int_{\mathcal{R}} P_x dxdy, \quad -\int_{\partial \mathcal{R}} Qdx = \int_{\mathcal{R}} Q_y dxdy.$$

Here the two-dimensional integrals are properly oriented, so they reduce to ordinary double integrals.

We will prove the theorem for a simple region, $\mathcal{R}$, simple in the sense that it can be parametrized easily to cover the region in two ways:

- For $c \leq y \leq d$, we have $a(y) \leq x \leq b(y)$.

- For $a \leq x \leq b$, we have $c(x) \leq y \leq d(x)$.

See the sketch:

24

Then using the first parametrization, we have:

$$\int_{\mathcal{R}} P_x dxdy = \int_c^d \int_{a(y)}^{b(y)} P_x dxdy = \int_c^d [P(x,y)]_{a(y)}^{b(y)} dy$$

$$= \int_c^d P(b(y),y)dy + \int_d^c P(a(y),y)dy = \int_{\partial R} P(x,y)dy.$$

The point here is that the right-hand-part of the boundary ∂R is parametrized by $x = b(y)$, for y from c to d, whereas the left-hand part is parametrized by $x = a(y)$, for y from d to c.

Next, using the second parametrization, we have:

$$\int_{\mathcal{R}} Q_y dxdy = \int_a^b \int_{c(x)}^{d(x)} Q_y dydx = \int_a^b [Q(x,y)]_{c(x)}^{d(x)} dx$$

$$= -\int_a^b Q(x,c(x))dx - \int_b^a Q(x,d(x))dx = -\int_{\partial R} Q(x,y)dx.$$

The point here is that the lower part of the boundary ∂R is parametrized by $y = c(x)$, for x from a to b, whereas the upper part is parametrized by $y = d(x)$, for x from b to a.
So the required result is proved.

Example
By this theorem, the area $\mathcal{A}$ of the region $\mathcal{R}$ can be given by any of the formulas:

$$\mathcal{A} = \int_{\mathcal{R}} dxdy = \int_{\partial \mathcal{R}} xdy = -\int_{\partial \mathcal{R}} ydx = \frac{1}{2} \int_{\partial \mathcal{R}} (xdy - ydx).$$

In polar co-ordinates, we have:

$$x = r\cos(\theta), \ dx = \cos(\theta)dr - \sin(\theta)rd\theta,$$

$$y = r\sin(\theta), \ dy = \sin(\theta)dr + \cos(\theta)rd\theta,$$

$$xdy - ydx = r\cos(\theta)(\sin(\theta)dr + \cos(\theta)rd\theta) - r\sin(\theta)(\cos(\theta)dr - \sin(\theta)rd\theta) = r^2 d\theta,$$

$$dxdy = \frac{1}{2}d(xdy - ydx) = \frac{1}{2}d(r^2 d\theta) = rdrd\theta.$$

So the area is also given by the formulas:

$$\mathcal{A} = \int_{\mathcal{R}} rdrd\theta = \frac{1}{2} \int_{\partial \mathcal{R}} r^2 d\theta.$$

The proof of the divergence theorem in all dimensions is basically the same. In $(n + 1)$-dimensions, let a simple region be one whose projection to each coordinate hyperplane is itself a simple region in n-dimensions.

In particular, if we write the $(n+1)$-co-ordinates as $(t, \underline{x})$ where $\underline{x} = (x^1, x^2, \ldots, x^n)$, by the linearity of Stokes' Theorem in the vector field, v, it suffices to take the vector field v to have only one non-zero component, denoted w, in the t-direction, in which case $\nabla.v = w_t$, so we just need to prove:

$$\int_{\mathcal{R}} w_t dt d^n x = \int_{\partial \mathcal{R}} v.N d^n S.$$

Here N is the outward pointing normal and $d^n S$ is the hyper-surface volume element. The point here is that once we have proved this case, the other cases, where v has just one non-zero component in another co-ordinate direction, follow by appropriate permutation of the coordinates and then the general case, where v is arbitrary is obtained by summation of these special case, using linearity over the reals. Now the region $\mathcal{R}$ is described by the range t goes from $t = a(x)$ to $t = b(x)$, as x ranges over $\mathcal{S}$, the projection of the region $\mathcal{R}$ onto the hyper-plane $t = 0$.

By the Fundamental Theorem of Calculus, the integral on the left-hand side is:

$$\int_{x \in \mathcal{S}} \int_{a(x)}^{b(x)} w_t dt d^n x = \int_{x \in \mathcal{S}} [w]_{a(x)}^{b(x)} d^n x = \int_{x \in \mathcal{S}} (w(b(x), x) - w(a(x), x)) d^n x$$

On the upper part of the region, given by the equation $f(t, x) = t - b(x) = 0$, the outward pointing (non-normalized) gradient vector is the vector:

$$\nabla(t - b(x)) = (1, -\nabla b).$$

On the lower part of the region, given by the equation $f(t, x) = a(x) - t = 0$, the outward pointing (non-normalized) gradient vector is the vector:

$$\nabla(a(x) - t) = (-1, \nabla a).$$

So, on the upper boundary, we have, for the surface element, $d^n S$, in terms of the coordinates $x^1, x^2, \ldots x_n$:

$$d^n S = \frac{\sqrt{f_t^2 + (\nabla b).(\nabla b)}}{|f_t|} d^n x = \sqrt{1 + (\nabla b).(\nabla b)} d^n x.$$

26

Similarly on the lower boundary, we have:

$$d^n S = \frac{\sqrt{f_t^2 + (\nabla a).(\nabla a)}}{|f_t|} d^n x = \sqrt{1 + (\nabla a).(\nabla a)} d^n x.$$

Also for the integrand $v.N$ we have the expression, since $v = (w, 0, 0, \ldots, 0)$ on the upper boundary where $t = b(x)$:

$$v.N = \frac{w(b(x), x)}{\sqrt{1 + (\nabla b).(\nabla b)}},$$

$$v.N d^n S = w(b(x), x) d^n x.$$

On the lower boundary, where $t = a(x)$, we have instead:

$$v.N = -\frac{w(a(x), x)}{\sqrt{1 + (\nabla a).(\nabla a)}},$$

$$v.N dS = -w(a(x), x) d^n x.$$

So we get:

$$\int_{\partial \mathcal{R}} v.N d^n S = \int_{t=b(x), x \in \mathcal{S}} v.N d^n S + \int_{t=a(x), x \in \mathcal{S}} v.N d^n S$$

$$= \int_{x \in \mathcal{S}} w(b(x), x) d^n x - \int_{x \in \mathcal{S}} w(a(x), x) d^n x.$$

$$= \int_{x \in \mathcal{S}} (w(b(x), x) - w(a(x), x)) d^n x.$$

This agrees with the left-hand-side of Stokes' Theorem, calculated above, so we are done: the divergence theorem holds.

Next we consider the case where the manifold of integration lies in a higher dimensional space. Stokes' Theorem is additive in the regions under consideration, so using the Rank theorem, we may take for the integration region $\mathcal{R}$, the space $(t(x), x)$ as x varies over a simple region $\mathcal{S}$ in $\mathbb{R}^n$ and the boundary $\partial\mathcal{R}$ is given by the same formula $(t(x), x)$, for $x \in \partial\mathcal{S}$. Here the map t is a smooth map from $\mathcal{S}$ to $\mathbb{R}^m$, for some m, the total dimension of the space inside which the manifold lies being $m + n$. Then the pull-back formula for integration tells us that $\int_\mathcal{R} d\omega = \int_\mathcal{S} p^*(d\omega) = \int_\mathcal{S} d(p^*(\omega))$, where $p : \mathcal{S} \to \mathcal{R}$ is the bijection $x \to p(x) = (t(x), x)$. But applied to the boundary, we have also the pull-back formula:

$$\int_{\partial\mathcal{R}} \omega = \int_{\partial\mathcal{S}} p^*(\omega).$$

Bu the divergence theorem, proved above, we have

$$\int_{\partial\mathcal{S}} p^*(\omega) = \int_\mathcal{S} d(p^*(\omega)).$$

So Stokes' Theorem follows. An alternative and perhaps instructive way of thinking of this case, without using the pull-back argument, is to restrict to the case that the boundary of $\mathcal{S}$ is piecewise flat. Then we form a closed domain, whose (curved) "roof" is the region $\mathcal{R}$, flat "floor" is $\mathcal{S}$ and such that the boundary pieces each lie in a hyperplane, so form (flat) vertical "walls", with base the edges of the boundary of $\mathcal{S}$ and tops the corresponding curved pieces of the boundary of $\mathcal{R}$. Then we proceed by reverse induction on the dimension n: we assume that Stokes' theorem holds in all higher dimensions, the base case being the highest possible dimension, where the theorem becomes the divergence theorem, which we have already proved. Denote by $\mathcal{T}$ the "room" with "ceiling" $\mathcal{R}$, "floor" $\mathcal{S}$ and a finite number of "walls", the boundary regions $\mathcal{W}_i, i = 1, 2, \ldots, k$, each with outward pointing normal, above the boundary pieces $\mathcal{S}_i, i = 1, 2, \ldots k$ of $\mathcal{S}$. These various regions form the boundary of the room $\mathcal{T}$. Specifically in the sense of integration, where we orient both $\mathcal{S}$ and $\mathcal{R}$ upwards, we have:

$$\partial\mathcal{T} = \mathcal{R} - \mathcal{S} + \Sigma_{j=1}^k \mathcal{W}_j,$$

$$0 = \partial^2\mathcal{T} = \partial\mathcal{R} - \partial\mathcal{S} + \sum_{j=1}^k \partial\mathcal{W}_j, \quad \partial\mathcal{R} = \partial\mathcal{S} - \sum_{j=1}^k \partial\mathcal{W}_j.$$

The minus signs in front of $\mathcal{S}$ in the first equation is there because the orientation chosen for $\mathcal{S}$ points inside the region $\mathcal{T}$.

Then we have by the induction hypothesis, for any form α defined on $\mathcal{T}$:

$$\int_{\mathcal{T}} d\alpha = \int_{\partial\mathcal{T}} \alpha = \int_{\mathcal{R}} \alpha - \int_{\mathcal{S}} \alpha + \sum_{j=1}^{k} \int_{\mathcal{W}_j} \alpha.$$

So now take the case that $\alpha = d\omega$.

Then $d\alpha = 0$, so we get, using the inductive hypothesis:

$$\int_{\mathcal{T}} d\alpha = 0 = \int_{\partial\mathcal{T}} \alpha = \int_{\partial\mathcal{T}} d\omega = \int_{\mathcal{R}} d\omega - \int_{\mathcal{S}} d\omega + \sum_{j=1}^{k} \int_{\mathcal{W}_j} d\omega,$$

$$\int_{\mathcal{R}} d\omega = \int_{\mathcal{S}} d\omega - \sum_{j=1}^{k} \int_{\mathcal{W}_j} d\omega.$$

But each of the regions $\mathcal{S}$ and $\mathcal{W}_j$, $j = 1, 2, \ldots, k$ are ordinary flat regions, so the usual divergence theorem applies to each of them and so Stokes' theorem holds for them, giving:

$$\int_{\mathcal{R}} d\omega = \int_{\partial\mathcal{S}} \omega - \sum_{j=1}^{k} \int_{\partial\mathcal{W}_j} \omega = \int_{\partial\mathcal{S} - \sum_{j=1}^{k} \partial\mathcal{W}_j} \omega = \int_{\partial\mathcal{R}} \omega.$$

So by reverse induction on dimension, the required theorem holds.

Notice that here we are exploiting the idea that d on forms and ∂ on regions are dual to each other: then dual to the idea that $d^2 = 0$ is the idea that $\partial^2 = 0$ or " boundary of a boundary is empty". This idea is codified in the theory of singular homology.

Summarizing, if we represent the integral in the style of Paul Dirac: $\langle \mathcal{R} | \omega \rangle$ denoting the integral of ω over the region $\mathcal{R}$, then Stokes' Theorem says that the adjoint of d is ∂:

$$\langle \partial\mathcal{R} | \omega \rangle = \langle \mathcal{R} | d\omega \rangle.$$

It follows that the adjoint of d^2 is ∂^2:

$$\langle \partial^2 \mathcal{R} | \omega \rangle = \langle \partial(\partial\mathcal{R}) | \omega \rangle = \langle \partial\mathcal{R} | d\omega \rangle = \langle \mathcal{R} | d(d\omega) \rangle = \langle \mathcal{R} | d^2\omega \rangle.$$

So perhaps it is not surprising that the fact that $d^2 = 0$ is related to the fact that $\partial^2 = 0$. Finally we can build up Stokes' Theorem for more complicated regions, by splitting it appropriately into simple regions, using the additivity of both sides of the Stokes' Theorem under the decomposition of regions.

Special cases of Stokes' Theorem

In one dimension, the theorem is just the Fundamental Theorem of Calculus:

$$\int_a^b df = f(b) - f(a).$$

The same theorem is true in any higher dimension, where f is a function and the integration region is a path $\mathcal{R}$ connecting its boundary points $\partial\mathcal{R} = B - A$ (the minus sign coming from the orientation):

$$\int_{\mathcal{R}} df = \int_{\partial\mathcal{R}} f = f(B) - f(A).$$

In two dimensions, we may write the theorem in divergence form with $v = [P, Q]$ and $\nabla.v = P_x + Q_y$:

$$\int_{\mathcal{R}} (\nabla.v) dx dy = \int_{\mathcal{R}} (P_x + Q_y) dx dy = \int_{\partial\mathcal{R}} P dy - Q dx = \int_{\partial\mathcal{R}} [P, Q].N ds.$$

Alternatively, we may replace P by Q and Q by $-P$ in this formula giving the more natural looking:

$$\int_{\mathcal{R}} (\nabla \times v) dx dy = \int_{\mathcal{R}} (Q_x - P_y) dx dy = \int_{\partial\mathcal{R}} P dx + Q dy = \int v.T ds.$$

Here $v = [P, Q]$, $\nabla \times v = Q_x - P_y$ and T is the unit tangent vector to the curve $\partial\mathcal{R}$; note that $T ds = P dx + Q dy$.

In three dimensions, we have the divergence theorem, when $\mathcal{R}$ is three-dimensional:

$$\int_{\mathcal{R}} (\nabla.v) dx dy dz = \int_{\mathcal{R}} (P_x + Q_y + R_z) dx dy dz$$

$$= \int_{\partial\mathcal{R}} v.\underline{\Sigma} = \int_{\partial\mathcal{R}} (P dy dz + Q dz dx + R dx dy) = \int_{\partial\mathcal{R}} v.N d^2 S.$$

Here $v = [P, Q, R]$, $\underline{\Sigma} = [dy dz, dz dx, dx dy] = N d^2 S$.

The lower dimensional Stokes' Theorem reads, for the case that $\mathcal{R}$ is two-dimensional:

$$\int_{\partial \mathcal{R}} (Pdx+Qdy+Rdz) = \int_{\mathcal{R}} ((R_y-Q_z)dydz+(P_z-R_x)dzdx+(Q_x-P_y)dxdy),$$

$$\int_{\partial \mathcal{R}} v.Tds = \int_{\partial \mathcal{R}} v.d\underline{x} = \int_{\mathcal{R}} (\nabla \times v).\underline{\Sigma} = \int_{\mathcal{R}} [N, \nabla, v]d^2S.$$

Here the notation $[A, B, C]$ denotes the vector triple product, the determinant of its vector arguments, A, B and C, so $[A, B, C] = A.(B \times C) = (A \times B).C$, so $[N, \nabla, v] = N.(\nabla \times v) = (N \times \nabla).v$.

Also T is the unit tangent vector in the direction of the curve, so we have $Tds = d\underline{x} = [dx, dy, dz]$ and $Nd^2S = [dydz, dzdx, dxdy]$, as usual.

In four dimensions, with coordinates $x^a = (t, x, y, z) = (t, \underline{x})$, where $x^0 = t$ and $x^1 = x$, $x^2 = y$ and $x^3 = z$, there are two dual ways of writing a two-form:

- First we just write it out as a polynomial in dx^a:

$$F = \frac{1}{2}\Sigma_{a=1}^4 \Sigma_{b=1}^4 F_{ab}dx^a dx^b = dtE + B,$$

$$E = \underline{E}.d\underline{x} = F_{01}dx + F_{02}dy + F_{03}dz,$$

$$F_{01} = \underline{E}_1, F_{02} = \underline{E}_2, F_{03} = \underline{E}_3,$$

$$B = \underline{B}.\underline{\Sigma} = B_1 dx^2 dx^3 + B_2 dx^3 dx^1 + B_3 dx^1 dx^2 = B_1 dydz + B_2 dzdx + B_3 dxdy.$$

- Dually, we write:

$$F^* = \frac{1}{2}\Sigma_{a=1}^4 \Sigma_{b=1}^4 F^{ab}\delta_a \delta_b \Omega = \frac{1}{2}\Sigma_{a=1}^4 \Sigma_{b=1}^4 F^{ab}\Sigma_{ab}.$$

Here, if we are in space-time, $[F^{01}, F^{02}, F^{03}] = -\underline{E}$ and $[F^{23}, F^{31}, F^{12}] = \underline{B}$. Also $\Omega = dtdxdydz$. Then we have:

$$\Sigma_{0i} = \delta_0 \delta_i \Omega = -\delta_i(dxdydz) = -\Sigma_i,$$

$$\Sigma_{23} = \delta_2 \delta_3 \Omega = -dtdx, \Sigma_{31} = \delta_3 \delta_1 \Omega = -dtdy, \Sigma_{12} = \delta_1 \delta_2 \Omega = -dtdz,$$

$$F^* = -dt\underline{B}.d\underline{x} + \underline{E}.\underline{\Sigma} - dt\underline{B}.d\underline{x}.$$

31

So the effect of going from F to F^* is the same as replacing $(\underline{E}, \underline{B})$ by $(-\underline{B}, \underline{E})$. The free electromagnetic field equations of James Clerk Maxwell are then:

$$0 = dF, 0 = dF^*.$$

Here the electromagnetic two-form in four dimensions, F encodes both the electric field $\underline{E}$ and the magnetic field $\underline{B}$.

Write $d = dt\partial_t + \partial$, where $\partial = dx\partial_x + dy\partial_y + dz\partial_z = d\underline{x}.\nabla$.

Then we have, with $\Sigma = dxdydz$ and $\underline{\Sigma} = [dydz, dzdx, dxdy]$ as usual:

$$dF = d(dtE + \underline{B}.\underline{\Sigma}) = -dt(\nabla \times \underline{E}).\underline{\Sigma} + dt(\partial_t\underline{B}).\underline{\Sigma}) + (\nabla.\underline{B})\Sigma,$$

$$= dt(-(\nabla \times \underline{E}).\underline{\Sigma} + \partial_t\underline{B}).\underline{\Sigma} + (\nabla.\underline{B})\Sigma,$$

$$dF^* = dt((\nabla \times \underline{B}).\underline{\Sigma} + \partial_t\underline{E}).\underline{\Sigma} + (\nabla.\underline{E})\Sigma.$$

The free Maxwell equations are then:

$$0 = dF, 0 = dF^*,$$

$$0 = \nabla.\underline{B}, \quad \partial_t\underline{B} = \nabla \times \underline{E},$$

$$0 = \nabla.\underline{E}, \quad \partial_t\underline{E} = -\nabla \times \underline{B},$$

Recall the derivative identity: $\nabla \times (\nabla \times \underline{u}) = \nabla(\nabla.\underline{u}) - \Delta\underline{u}$.

Here $\underline{u}$ is a vector field and $\Delta = \nabla.\nabla = \partial_x^2 + \partial_y^2 + \partial_z^2$ is called the three-dimensional Laplacian.

Then we get, using the free Maxwell equations:

$$\partial_t^2\underline{B} = \nabla \times \partial_t\underline{E} = -\nabla \times (\nabla \times \underline{B}) = \Delta\underline{B},$$

$$\partial_t^2\underline{E} = -\nabla \times \partial_t\underline{B} = -\nabla \times (\nabla \times \underline{E}) = \Delta\underline{E}.$$

$$\Box\underline{B} = 0, \quad \Box\underline{E} = 0.$$

Here $\Box$ is the wave operator $\Box = \partial_t^2 - \Delta$.

The fact that the free Maxwell equations obey the wave equation in four dimensions, is one of the cornerstones of Albert Einstein's theory of relativity. The four-dimensional perspective is due to Hermann Minkowski.

Maxwell's equations in the presence of currents are the following:

$$dF = 0, \quad dF^* = J.$$

Here J is the current three-form, which can be written $J = [\phi, \underline{j}].\underline{\Omega}$, where we have:

$$\underline{\Omega} = [\delta_0, \delta_1, \delta_2, \delta_3]\Omega = [dxdydz, -dtdydz, -dtdzdx, -dtdxdy].$$

So we have:

$$J = \phi\Sigma - dt\underline{j}.\underline{\Sigma}.$$

The equation of current conservation, is the equation $dJ = 0$, which follows by applying d to the equation $dF^* = J$:

$$0 = dJ = (dt\partial_t + \partial)(\phi dxdydz - dt\underline{j}.\underline{\Sigma})$$

$$= (\partial_t\phi + \nabla.\underline{j})dtdxdydz,$$

$$\partial_t\phi + \nabla.\underline{j} = 0.$$

The Maxwell equations now read:

$$dF = 0, \quad 0 = \nabla.\underline{B}, \quad \partial_t\underline{B} = \nabla \times \underline{E},$$

$$J = dF^*, \quad \phi\Sigma - dt\underline{j}.\underline{\Sigma} = dt((\nabla \times \underline{B}).\underline{\Sigma} + \partial_t\underline{E}).\underline{\Sigma} + (\nabla.\underline{E})\Sigma,$$

$$\phi = \nabla.\underline{E},$$

$$-\underline{j} = \partial_t\underline{E} + \nabla \times \underline{B}.$$

Now we have:

$$\partial_t^2\underline{E} = -\partial_t\underline{j} - \nabla \times \partial_t\underline{B} = -\partial_t\underline{j} - \nabla \times (\nabla \times \underline{E}) = \Delta\underline{E} - \partial_t\underline{j} - \nabla(\nabla.\underline{E})$$

$$\Box\underline{E} = -\partial_t\underline{j} - \nabla\phi,$$

$$\partial_t^2\underline{B} = \nabla \times \partial_t\underline{E} = -\nabla \times \underline{j} - \nabla \times (\nabla \times \underline{B}) = \Delta\underline{B} - \nabla \times \underline{j},$$

$$\Box\underline{B} = -\nabla \times \underline{j}.$$

So the electric and magnetic fields obey the wave equation with sources determined by the current.