Lecture 2: Differentiation and Linear Algebra

Joan Lasenby

[with thanks to: Chris Doran & Anthony Lasenby (book), Hugo Hadfield & Eivind Eide (code), Leo Dorst (book)........ and of course, David Hestenes]

Signal Processing Group, Engineering Department, Cambridge, UK
and
Trinity College, Cambridge
`jl221@cam.ac.uk`, `www-sigproc.eng.cam.ac.uk/` $\sim$ `jl`

July 2018

Contents: GA Course I, Session 2

- The vector derivative and examples of its use.
- Multivector differentiation: examples.
- Linear algebra
- Geometric algebras with non-Euclidean metrics.

The contents follow the notation and ordering of *Geometric Algebra for Physicists [C.J.L. Doran and A.N. Lasenby]* and the corresponding course the book was based on.

The Vector Derivative

A vector x can be represented in terms of coordinates in two ways:

$$x = x^k e_k \quad \text{or} \quad x = x_k e^k$$

(Summation implied). Depending on whether we expand in terms of a given frame $\{e_k\}$ or its reciprocal $\{e^k\}$. The coefficients in these two frames are therefore given by

$$x^k = e^k \cdot x \quad \text{and} \quad x_k = e_k \cdot x$$

Now define the following derivative operator which we call the vector derivative

$$\nabla = \sum_k e^k \frac{\partial}{\partial x^k} \equiv e^k \frac{\partial}{\partial x^k}$$

..this is clearly a vector!

The Vector Derivative, cont...

$$\nabla = \sum_k e^k \frac{\partial}{\partial x^k}$$

This is a definition so far, but we will now see how this form arises.

Suppose we have a function acting on vectors, $F(x)$. Using standard definitions of rates of change, we can define the directional derivative of F, evaluated at x, in the direction of a vector a as

$$\lim_{\epsilon \to 0} \frac{F(x + \epsilon a) - F(x)}{\epsilon}$$

The Vector Derivative cont....

Now, suppose we want the directional derivative in the direction of one of our frame vectors, say e_1, this is given by

$$\lim_{\epsilon \to 0} \frac{F((x^1+\epsilon)e_1 + x^2 e_2 + x^3 e_3) - F(x^1 e_1 + x^2 e_2 + x^3 e_3)}{\epsilon}$$

which we recognise as

$$\frac{\partial F(x)}{\partial x^1}$$

ie the derivative with respect to the first coordinate, keeping the second and third coordinates constant.

The Vector Derivative cont....

So, if we wish to define a gradient operator, ∇, such that $(a \cdot \nabla) F(x)$ gives the directional derivative of F in the a direction, we clearly need:

$$e_i \cdot \nabla = \frac{\partial}{\partial x^i} \quad \text{for} \quad i = 1, 2, 3$$

...which, since $e_i \cdot e^j \frac{\partial}{\partial x^j} = \frac{\partial}{\partial x^i}$, gives us the previous form of the vector derivative:

$$\nabla = \sum_k e^k \frac{\partial}{\partial x^k}$$

The Vector Derivative cont....

It follows now that if we dot ∇ with a, we get the directional derivative in the a direction:

$$a \cdot \nabla F(x) = \lim_{\epsilon \to 0} \frac{F(x + \epsilon a) - F(x)}{\epsilon}$$

The definition of ∇ is in fact independent of the choice of frame.

Operating on Scalar and Vector Fields

Operating on:

A Scalar Field $\phi(x)$: it gives $\nabla\phi$ which is the gradient.

A Vector Field $J(x)$: it gives ∇J. This is a geometric product

Scalar part gives divergence

Bivector part gives curl

$$\nabla J = \nabla \cdot J + \nabla \wedge J$$

Very important in electromagnetism.

The Multivector Derivative

Recall our definition of the directional derivative in the a direction

$$a\cdot\nabla\, F(x) = \lim_{\epsilon\to 0} \frac{F(x+\epsilon a) - F(x)}{\epsilon}$$

We now want to generalise this idea to enable us to find the derivative of $F(X)$, in the A 'direction' – where X is a general mixed grade multivector (so $F(X)$ is a general multivector valued function of X).

Let us use $*$ to denote taking the scalar part, ie $P * Q \equiv \langle PQ \rangle$. Then, provided A has same grades as X, it makes sense to define:

$$A * \partial_X F(X) = \lim_{\tau\to 0} \frac{F(X+\tau A) - F(X)}{\tau}$$

The Multivector Derivative cont...

Let $\{e_J\}$ be a basis for X – ie if X is a bivector, then $\{e_J\}$ will be the basis bivectors.

With the definition on the previous slide, $e_J * \partial_X$ is therefore the partial derivative in the e_J direction. Giving

$$\partial_X \equiv \sum_J e^J e_J * \partial_X$$

[since $e_J * \partial_X \equiv e_J * \{e^I(e_I * \partial_X\}$].

Key to using these definitions of multivector differentiation are several important results:

The Multivector Derivative cont...

If $P_X(B)$ is the projection of B onto the grades of X (ie $P_X(B) \equiv e^J \langle e_J B \rangle$), then our first important result is

$$\partial_X \langle XB \rangle = P_X(B)$$

We can see this by going back to our definitions:

$$e_J * \partial_X \langle XB \rangle = \lim_{\tau \to 0} \frac{\langle (X + \tau e_J)B \rangle - \langle XB \rangle}{\tau} = \lim_{\tau \to 0} \frac{\langle XB \rangle + \tau \langle e_J B \rangle - \langle XB \rangle}{\tau}$$

$$\lim_{\tau \to 0} \frac{\tau \langle e_J B \rangle}{\tau} = \langle e_J B \rangle$$

Therefore giving us

$$\partial_X \langle XB \rangle = e^J (e_J * \partial_X) \langle XB \rangle = e^J \langle e_J B \rangle \equiv P_X(B)$$

Other Key Results

Some other useful results are listed here (proofs are similar to that on previous slide and are left as exercises):

$$\partial_X \langle XB \rangle = P_X(B)$$

$$\partial_X \langle \tilde{X}B \rangle = P_X(\tilde{B})$$

$$\partial_{\tilde{X}} \langle \tilde{X}B \rangle = P_{\tilde{X}}(B) = P_X(B)$$

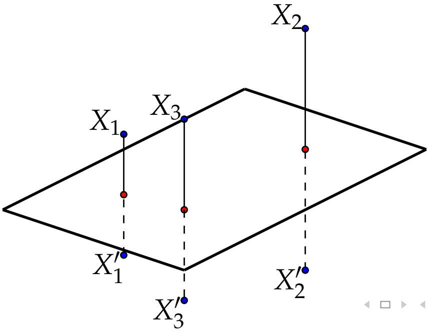
$$\partial_\psi \langle M\psi^{-1} \rangle = -\psi^{-1} P_\psi(M) \psi^{-1}$$

X, B, M, ψ all general multivectors.

A Simple Example

Suppose we wish to fit a set of points $\{X_i\}$ to a plane Φ – where the X_i and Φ are conformal representations (vector and 4 vector respectively).

One possible way forward is to find the plane that minimises the sum of the squared perpendicular distances of the points from the plane.



Plane fitting example, cont....

Recall that $\Phi X \Phi$ is the reflection of X in Φ, so that $-X \cdot (\Phi X \Phi)$ is the distance between the point and the plane. Thus we could take as our cost function:

$$S = -\sum_i X_i \cdot (\Phi X_i \Phi)$$

Now use the result $\partial_X \langle XB \rangle = P_X(B)$ to differentiate this expression wrt Φ

$$\partial_\Phi S = -\sum_i \partial_\Phi \langle X_i \Phi X_i \Phi \rangle = -\sum_i \dot{\partial}_\Phi \langle X_i \dot{\Phi} X_i \Phi \rangle + \dot{\partial}_\Phi \langle X_i \Phi X_i \dot{\Phi} \rangle$$

$$= -2 \sum_i P_\Phi(X_i \Phi X_i) = -2 \sum_i X_i \Phi X_i$$

$\Longrightarrow$ solve (via linear algebra techniques) $\sum_i X_i \Phi X_i = 0$.

Differentiation cont....

Of course we can extend these ideas to other geometric fitting problems and also to those without closed form solutions, using gradient information to find solutions.

Another example is differentiating wrt rotors or bivectors.

Suppose we wished to create a Kalman filter-like system which tracked bivectors (not simply their components in some basis) – this might involve evaluating expressions such as

$$\partial_{B_n} \sum_{i=1}^{L} \langle v_n^i R_n u_{n-1}^i \tilde{R}_n \rangle$$

where $R_n = e^{-B_n}$, u, v s are vectors.

Differentiation cont....

Using just the standard results given, and a page of algebra later (but one only needs to do it once!) we find that

$$\partial_{B_n} \langle v_n R_n u_{n-1} \tilde{R}_n \rangle = -\Gamma(B_n) + \frac{1}{|B_n|^2} \langle B_n \Gamma(B_n) \tilde{R}_n B_n R_n \rangle_2$$

$$+ \frac{\sin(|B_n|)}{|B_n|} \left\langle \frac{B_n \Gamma(B_n) \tilde{R}_n B_n}{|B_n|^2} + \Gamma(B_n) \tilde{R}_n \right\rangle_2$$

where $\Gamma(B_n) = \frac{1}{2}[u_{n-1} \wedge \tilde{R}_n v_n R_n] R_n$.

Linear Algebra

A linear function, f, mapping vectors to vectors asatisfies

$$\mathrm{f}(\lambda a + \mu b) = \lambda \mathrm{f}(a) + \mu \mathrm{f}(b)$$

We can now extend f to act on any order blade by (outermorphism)

$$\mathrm{f}(a_1 \wedge a_2 \wedge ... \wedge a_n) = \mathrm{f}(a_1) \wedge \mathrm{f}(a_2) \wedge .... \wedge \mathrm{f}(a_n)$$

Note that the resulting blade has the same grade as the original blade. Thus, an important property is that these extended linear functions are grade preserving, ie

$$\mathrm{f}(A_r) = \langle \mathrm{f}(A_r) \rangle_r$$

Linear Algebra cont....

Matrices are also linear functions which map vectors to vectors. If F is the matrix corresponding to the linear function f, we obtain the elements of F via

$$\mathrm{F}_{ij} = e_i \cdot \mathrm{f}(e_j)$$

Where $\{e_i\}$ is the basis in which the vectors the matrix acts on are written.

As with matrix multiplication, where we obtain a 3rd matrix (linear function) from combining two other matrices (linear functions), ie $\mathrm{H} = \mathrm{FG}$, we can also write

$$\mathrm{h}(a) = \mathrm{f}[\mathrm{g}(a)] = \mathrm{fg}(a)$$

The product of linear functions is associative.

We now need to verify that

$$\mathrm{h}(A) = \mathrm{f}[\mathrm{g}(A)] = \mathrm{fg}(A)$$

for any multivector A.

First take a blade $a_1 \wedge a_2 \wedge ... \wedge a_r$ and note that

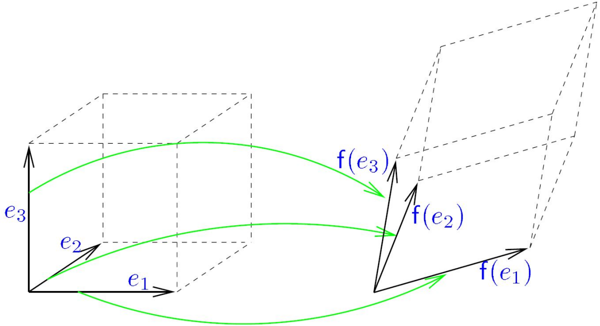
$$\mathrm{h}(a_1 \wedge a_2 \wedge ... \wedge a_r) = \mathrm{fg}(a_1) \wedge \mathrm{fg}(a_2) \wedge ... \wedge \mathrm{fg}(a_r)$$

$$= \mathrm{f}[\mathrm{g}(a_1) \wedge \mathrm{g}(a_2) \wedge ..... \wedge \mathrm{g}(a_r)] = \mathrm{f}[\mathrm{g}(a_1 \wedge a_2 \wedge ... \wedge a_r)]$$

from which we get the first result.

The Determinant

Consider the action of a linear function f on an orthogonal basis in 3d:



The unit cube $I = e_1 \wedge e_2 \wedge e_3$ is transformed to a parallelepiped, V

$$\mathsf{V} = \mathsf{f}(e_1) \wedge \mathsf{f}(e_2) \wedge \mathsf{f}(e_3) = \mathsf{f}(I)$$

The Determinant cont....

So, since $\mathsf{f}(I)$ is also a pseudoscalar, we see that if V is the magnitude of V, then

$$\mathsf{f}(I) = VI$$

Let us define the determinant of the linear function f as the volume scale factor V. So that

$$\boxed{\mathsf{f}(I) = \det(\mathsf{f})\; I}$$

This enables us to find the form of the determinant explicitly (in terms of partial derivatives between coordinate frames) very easily in any dimension.

A Key Result

As before, let $\mathrm{h} = \mathrm{fg}$, then

$$\mathrm{h}(I) = \det(\mathrm{h})\, I = \mathrm{f}(\mathrm{g}(I)) = \mathrm{f}(\det(\mathrm{g})\, I)$$

$$= \det(\mathrm{g})\, \mathrm{f}(I) = \det(\mathrm{g}) \det(\mathrm{f})(I)$$

So we have proved that

$$\det(\mathrm{fg}) = \det(\mathrm{f}) \det(\mathrm{g})$$

A very easy proof!

The Adjoint/Transpose of a Linear Function

For a matrix F and its transpose, F^T we have (for any vectors a, b)

$$a^T F b = b^T F^T a = \phi \quad (scalar)$$

In GA we can write this in terms of linear functions as

$$a \cdot \mathrm{f}(b) = \bar{\mathrm{f}}(a) \cdot b$$

This reverse linear function, $\bar{\mathrm{f}}$, is called the adjoint.

The Adjoint cont....

It is not hard to show that the adjoint extends to blades in the expected way

$$\bar{f}(a_1 \wedge a_2 \wedge ... \wedge a_n) = \bar{f}(a_1) \wedge \bar{f}(a_2) \wedge .... \wedge \bar{f}(a_n)$$

See exercises to show that

$$a \cdot f(b \wedge c) = f[\bar{f}(a) \cdot (b \wedge c)]$$

This can now be generalised to

$$A_r \cdot \bar{f}(B_s) = \bar{f}[f(A_r) \cdot B_s] \quad r \leq s$$

$$f(A_r) \cdot B_s = f[A_r \cdot \bar{f}(B_s)] \quad r \geq s$$

The Inverse

$$A_r \cdot \bar{\mathrm{f}}(B_s) = \bar{\mathrm{f}}[\mathrm{f}(A_r) \cdot B_s] \quad r \leq s$$

Now put $B_s = I$ in this formula:

$$A_r \cdot \bar{\mathrm{f}}(I) = A_r \cdot \det(\mathrm{f})(I) = \det(\mathrm{f})(A_r I)$$

$$= \bar{\mathrm{f}}[\mathrm{f}(A_r) \cdot I] = \bar{\mathrm{f}}[\mathrm{f}(A_r) I]$$

We can now write this as

$$A_r = \bar{\mathrm{f}}[\mathrm{f}(A_r) I] I^{-1} [\det(\mathrm{f})]^{-1}$$

The Inverse cont...

Repeat this here:

$$A_r = \bar{\mathsf{f}}[\mathsf{f}(A_r)I]I^{-1}[\det(\mathsf{f})]^{-1}$$

The next stage is to put $A_r = \mathsf{f}^{-1}(B_r)$ in this equation:

$$\mathsf{f}^{-1}(B_r) = \bar{\mathsf{f}}[B_r I]I^{-1}[\det(\mathsf{f})]^{-1}$$

This leads us to the important and simple formulae for the inverse of a function and its adjoint

$$\boxed{\mathsf{f}^{-1}(A) = [\det(\mathsf{f})]^{-1}\bar{\mathsf{f}}[AI]I^{-1}}$$

$$\boxed{\bar{\mathsf{f}}^{-1}(A) = [\det(\mathsf{f})]^{-1}\mathsf{f}[AI]I^{-1}}$$

An Example

Let us see if this works for rotations

$$\mathrm{R}(a) = Ra\tilde{R} \quad \text{and} \quad \bar{\mathrm{R}}(a) = \tilde{R}aR$$

So, putting this in our inverse formula:

$$\mathrm{R}^{-1}(A) = [\det(\mathrm{R})]^{-1}\bar{\mathrm{R}}(AI)I^{-1}$$

$$= [\det(\mathrm{R})]^{-1}\tilde{R}(AI)RI^{-1} = \tilde{R}AR$$

since $\det(R) = 1$. Thus the inverse is the adjoint ... as we know from $R\tilde{R} = 1$.

A Second Example

In the Gauge Theory of Gravity, there is a choice of gauge (linear function) which makes light paths straight in the vicinity of a black hole:

$$\bar{h}(a) = a + \frac{M}{r}(a \cdot e_-)e_- \quad e_- = e_t - e_r$$

The inverse of this gauge function is needed – given the formula for the inverse it is not too hard to work this out!

More Linear Algebra...

That we won't look at.....

- The idea of eigenblades – this becomes possible with our extension of linear functions to act on blades.
- Symmetric ($\mathsf{f}(a) = \bar{\mathsf{f}}(a)$) and antisymmetric ($\mathsf{f}(a) = -\bar{\mathsf{f}}(a)$) functions. In particular, antisymmetric functions are best studied using bivectors.
- Decompositions.
- Tensors - we can think of tensors as linear functions mapping r-blades to s-blades. Thus we retain some physical intuition that is generally lost in index notation.

Functional Differentiation

We will only touch on this briefly, but it is crucial to work in physics and has hardly been used at all in other fields.

$$\partial_{\mathfrak{f}(a)}(\mathfrak{f}(b)\cdot c) = (a\cdot b)c$$

In engineering, this, in particular, enables us to differentiate wrt to structured matrices in a way which is very hard to do otherwise.

Space-Time Algebra

Here we have 3 dimensions of space and 1 dimension of time – described by 4 orthogonal vectors, $\{\gamma_0, \gamma_1, \gamma_2, \gamma_3\}$, such that ($i, j = 1, 2, 3$):

$$\gamma_0^2 = 1, \quad \gamma_0 \cdot \gamma_i = 0, \quad \gamma_i \cdot \gamma_j = -\delta_{ij}$$

With γ_0 picking out the time axis. Call this the Spacetime Algebra (STA)

Since it is a 4d algebra, we know it has $2^4 = 16$ elements, 1 scalar, 4 vectors, 6 bivectors, 4 trivectors, 1 4-vector/pseudoscalar. We now look at the bivector algebra of the STA:

$$\gamma_i \wedge \gamma_0, \quad \gamma_i \wedge \gamma_j \;\; (i \neq j)$$

The Spacetime Bivectors

Consider the set of bivectors of the form $\gamma_i \wedge \gamma_0$. Write these as:

$$\sigma_i = \gamma_i \wedge \gamma_0 = \gamma_i \gamma_0$$

The σ_i satisfy:

$$\begin{aligned} \sigma_i \cdot \sigma_j &= \frac{1}{2}\left(\gamma_i\gamma_0\gamma_j\gamma_0 + \gamma_j\gamma_0\gamma_i\gamma_0\right) \\ &= \frac{1}{2}\left(-\gamma_i\gamma_j - \gamma_j\gamma_i\right) = \delta_{ij} \end{aligned}$$

These bivectors can be seen as relative vectors and generate a 3d Euclidean algebra – which is called a relative space in the rest frame of γ_0.

Note that: $\sigma_1\sigma_2\sigma_3 = (\gamma_1\gamma_0)(\gamma_2\gamma_0)(\gamma_3\gamma_0) = \gamma_0\gamma_1\gamma_2\gamma_3 = I$

Our STA and our 3d relative space have the same pseudoscalar.

Lorentz Transformations

Lorentz transformations: how the coordinates (local rest frame) of events seen by one observer (x, y, z, t) are related to those seen by another observer, (x', y', z', t').

Consider one frame moving at velocity βc along the e_3 axis. We find that basis vectors in our two frames are in fact related by a rotor transformation, similar to rotations in Euclidean space:

$$e'_\mu = R e_\mu \tilde{R} \quad \text{with} \quad R = \mathrm{e}^{\alpha e_3 e_0 / 2}$$

with $\mu = 0, 1, 2, 3$ and $\tanh(\alpha) = \beta$.

Thus, boosts in relativity are achieved simply by rotors.

The Conformal Model of 3d Space (CGA)

Another example, which Leo will expand upon in the following sessions, is the algebra $\mathcal{G}_{4,1}$.

We take the basis of 3d Euclidean space and add on two more basis vectors which square to +1 and -1:

$$\{e_1, e_2, e_3, e, \bar{e}\}, \;\; e_i^2 = 1, \;\; e^2 = 1, \;\; \bar{e}^2 = -1$$

$$e_i \cdot e_j = \delta_{ij}, \;\; e_i \cdot e = 0, \;\; e_i \cdot \bar{e} = 0, \;\; e \cdot \bar{e} = 0$$

Look for transformations that keep $n = e + \bar{e}$ invariant – to get the special conformal group as rotors.

Exercises 1

1. By noting that $\langle XB \rangle = \langle (XB)^{\sim} \rangle$, show the second key result
$$\partial_X \langle \tilde{X} B \rangle = P_X(\tilde{B})$$

2. Key result 1 tells us that $\partial_{\tilde{X}} \langle \tilde{X} B \rangle = P_{\tilde{X}}(B)$. Verify that $P_{\tilde{X}}(B) = P_X(B)$, to give the 3rd key result.

3. to show the 4th key result
$$\partial_\psi \langle M \psi^{-1} \rangle = -\psi^{-1} P_\psi(M) \psi^{-1}$$
use the fact that $\partial_\psi \langle M \psi \psi^{-1} \rangle = \partial_\psi \langle M \rangle = 0$. Hint: recall that XAX has the same grades as A.

Exercises 2

1. For a matrix F

$$F = \begin{pmatrix} \mathrm{F}_{11} & \mathrm{F}_{12} \\ \mathrm{F}_{21} & \mathrm{F}_{22} \end{pmatrix}$$

   Verify that $\mathrm{F}_{ij} = e_i \cdot \mathrm{f}(e_j)$, where $e_1 = [1, 0]^T$ and $e_2 = [0, 1]^T$, for $i, j = 1, 2$.

2. Rotations are linear functions, so we can write $\mathrm{R}(a) = Ra\tilde{R}$, where R is the rotor. If A_r is an r-blade, show that

$$RA_r\tilde{R} = (Ra_1\tilde{R}) \wedge (Ra_2\tilde{R}) \wedge \ldots \wedge (Ra_r\tilde{R})$$

   Thus we can rotate any element of our algebra with the same rotor expression.

Exercises 3

1. For any vectors p, q, r, show that

$$p\cdot(q\wedge r) = (p\cdot q)r - (p\cdot r)q$$

2. By using the fact that $a\cdot \mathrm{f}(b\wedge c) = a\cdot[\mathrm{f}(b)\wedge \mathrm{f}(c)]$, use the above result to show that

$$a\cdot \mathrm{f}(b\wedge c) = (\bar{\mathrm{f}}(a)\cdot b)\mathrm{f}(c) - (\bar{\mathrm{f}}(a)\cdot c)\mathrm{f}(b)$$

and simplify to get the final result

$$a\cdot \mathrm{f}(b\wedge c) = \mathrm{f}[\bar{\mathrm{f}}(a)\cdot(b\wedge c)]$$