Chapter 9

Jacobians of Matrix Transforms (with wedge products)

9.1 Wedge Products (The "$\wedge$" Notation)

The straightforward definition of the Jacobian for an $n \times n$ matrix function such as $Y = X^2$ indicates the construction of the $n^2 \times n^2$ matrix $\frac{\partial Y_{ij}}{\partial Y_{kl}}$ (perhaps this is best thought of as $(n \times n)$ by $(n \times n)$—a four dimensional construction) . Mathematicians have invented a notation, wedge products, that avoids the construction of this huge matrix and still formally achieves the same goal.

In this book we will wedge together differentials as illustrated in this example:

$$(2\mathrm{d}x + x^2\mathrm{d}y + 5\mathrm{d}w + 2\mathrm{d}z) \wedge (y\mathrm{d}x - x\mathrm{d}y) =$$
$$(-2x - x^2y)\mathrm{d}x \wedge \mathrm{d}y + 5y(\mathrm{d}w \wedge \mathrm{d}x) - 5x(\mathrm{d}w \wedge \mathrm{d}y) - 2y(\mathrm{d}x \wedge \mathrm{d}z) + 2x(\mathrm{d}y \wedge \mathrm{d}z)$$

Formally the wedge product is quite easy, it acts like multiplication except that it follows the anticommutative law

$$(\mathrm{d}u \wedge \mathrm{d}v) = (-\mathrm{d}v \wedge \mathrm{d}u)$$

Generally

$$(p\mathrm{d}u + q\mathrm{d}r) \wedge (r\mathrm{d}u + s\mathrm{d}r) = (pr - qs)(\mathrm{d}u \wedge \mathrm{d}r)$$

It therefore follows that

$$\mathrm{d}u \wedge \mathrm{d}u = 0\,.$$

In the next section, we will explore the algebra a little further. Suffice it to say that any linear combination of differentials is referred to as a 1-form. When we wedge two together, we get a 2-form, and so on. We are allowed to take linear combinations of 2-forms as in $\mathrm{d}x \wedge \mathrm{d}y + x2\mathrm{d}y \wedge \mathrm{d}z$ but we are not allowed to add 1-forms and 2-forms. $\mathrm{d}x + \mathrm{d}x \wedge \mathrm{d}y$ is not defined!

Example 1: differential elements in a 2×2 matrix

Let $Y = Y(X)$ be a 2×2 matrix, where $X = \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix}$

$$\mathrm{d}Y = \begin{bmatrix} \frac{\partial y_{11}}{\partial x_{11}}\mathrm{d}x_{11} + \frac{\partial y_{11}}{\partial x_{12}}\mathrm{d}x_{12} + \frac{\partial y_{11}}{\partial x_{21}}\mathrm{d}x_{21} + \frac{\partial y_{11}}{\partial x_{22}}\mathrm{d}x_{22} & \frac{\partial y_{12}}{\partial x_{11}}\mathrm{d}x_{11} + \frac{\partial y_{12}}{\partial x_{12}}\mathrm{d}x_{12} + \frac{\partial y_{12}}{\partial x_{21}}\mathrm{d}x_{21} + \frac{\partial y_{12}}{\partial x_{22}}\mathrm{d}x_{22} \\ \frac{\partial y_{21}}{\partial x_{11}}\mathrm{d}x_{11} + \frac{\partial y_{21}}{\partial x_{12}}\mathrm{d}x_{12} + \frac{\partial y_{21}}{\partial x_{21}}\mathrm{d}x_{21} + \frac{\partial y_{21}}{\partial x_{22}}\mathrm{d}x_{22} & \frac{\partial y_{22}}{\partial x_{11}}\mathrm{d}x_{11} + \frac{\partial y_{22}}{\partial x_{12}}\mathrm{d}x_{12} + \frac{\partial y_{22}}{\partial x_{21}}\mathrm{d}x_{21} + \frac{\partial y_{22}}{\partial x_{22}}\mathrm{d}x_{22} \end{bmatrix}$$

We have that the wedge product of all the elements is

$$\mathrm{d}y_{11} \wedge \mathrm{d}y_{21} \wedge \mathrm{d}y_{12} \wedge \mathrm{d}y_{22} = \det\left(\frac{\partial y_{ij}}{\partial x_{kl}}\right) \mathrm{d}x_{11} \wedge \mathrm{d}x_{21} \wedge \mathrm{d}x_{12} \wedge \mathrm{d}x_{22}\,.$$

Example 2: differential elements of an $n \times n$ matrix

If $Y = Y(X)$ is an $n \times n$ matrix, then

$$\bigwedge (\mathrm{d}Y_{ij}) = \det\left|\frac{\partial Y_{ij}}{\partial x_{kl}}\right| \left(\bigwedge \mathrm{d}x_{ij}\right)$$

Notice that we use "$\bigwedge$" to indicate the wedge product of the elements ignoring order.

Example 3: differential elements in an $n \times 1$ vector

If we were to compute the determinant of an upper triangular matrix U we might write

$$U = \begin{pmatrix} u_{11} & * & * & * & * \\ & u_{22} & * & * & * \\ & & u_{33} & * & * \\ & & & u_{44} & * \\ & & & & u_{55} \end{pmatrix}$$

here the $*$'s indicate "don't care's."

If U is expressing a relationship between different differential quantities: $\mathrm{d}y = U\mathrm{d}x$, then we can write

$$\mathrm{d}y_1 \wedge \mathrm{d}y_2 \wedge \mathrm{d}y_3 \wedge \mathrm{d}y_4 \wedge \mathrm{d}y_5 = (u_{11}\mathrm{d}x_1 + \cdots) \wedge (u_{22}\mathrm{d}x_2 + \cdots) \wedge (u_{33}\mathrm{d}x_3 + \cdots) \wedge (u_{44}\mathrm{d}x_4 + \cdots) \wedge u_{55}\mathrm{d}x_5$$

where the ellipses indicate "don't care's" because of the commutative property of the wedge product. We then have that

$$\mathrm{d}y_1 \wedge \mathrm{d}y_2 \wedge \mathrm{d}y_3 \wedge \mathrm{d}y_4 \wedge \mathrm{d}y_5 = u_{11}u_{22}u_{33}u_{44}u_{55}\, \mathrm{d}x_1 \wedge \mathrm{d}x_2 \wedge \mathrm{d}x_3 \wedge \mathrm{d}x_4 \wedge \mathrm{d}x_5\,.$$

Experience has shown that this wedge notation is superior for expressing such determinants.

Notation:

Let

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n\,.$$

We have seen that it is natural to form $\mathrm{d}x_1 \wedge \mathrm{d}x_2 \wedge \cdots \wedge \mathrm{d}x_n$ which we may denote $(\mathrm{d}x)$, reserving $\mathrm{d}x$ for the vector of differentials (1-forms). Notationally,

$$\mathrm{d}x = \begin{pmatrix} \mathrm{d}x_1 \\ \vdots \\ \mathrm{d}x_n \end{pmatrix} \quad \text{while} \quad (\mathrm{d}x) = \bigwedge_{i=1}^{n} \mathrm{d}x_i = \mathrm{d}x_1 \wedge \ldots \wedge \mathrm{d}x_n\,.$$

We will avoid any potential conflict in notation by using the parentheses in a manner consistent with the description above. (It is rare that a matrix or vector is parenthesized by itself in mathematics.)

Similarly if $A \in \mathbb{R}^{n,n}$,

$$\mathrm{d}A = \begin{pmatrix} \mathrm{d}a_{11} & \cdots & \mathrm{d}a_{1n} \\ \vdots & & \vdots \\ \mathrm{d}a_{n1} & \cdots & \mathrm{d}a_{nn} \end{pmatrix}$$

is a matrix of differentials, and we may form the exterior product

$$(\mathrm{d}A) = \mathrm{d}a_{11} \wedge \mathrm{d}a_{12} \wedge \cdots \wedge \mathrm{d}a_{nn} = \bigwedge_{i,j=1,\ldots,n} \mathrm{d}a_{ij} \ .$$

This suggests a general notation. If Φ is some object, $\mathrm{d}\Phi$ is a similar object of differentials, and $(\mathrm{d}\Phi)$ is the exterior product of independent elements in Φ.

Thus if U is upper triangular, $(\mathrm{d}U) = \bigwedge_{i \le j} \mathrm{d}u_{ij}$ or if L is *unit* lower triangular, $(\mathrm{d}L) = \bigwedge_{i>j} \mathrm{d}l_{ij}$. Similarly,

If S is symmetric, $(\mathrm{d}S) = \bigwedge_{i \ge j} \mathrm{d}s_{ij}$.

If Λ is diagonal, $(\mathrm{d}\Lambda) = \bigwedge_i \mathrm{d}\lambda_i$.

If A is anti-symmetric, $(\mathrm{d}A) = \bigwedge_{i<j} \mathrm{d}a_{ij}$ (Sign will not matter).

We may also get sloppy and write $(\mathrm{d}A)(\mathrm{d}B)$ when strictly speaking we should write $(\mathrm{d}A) \wedge (\mathrm{d}B)$.

9.2 Exterior Products (The Algebra)

Let V be an n-dimensional vector space over $\mathbb{R}$. For $p = 0, 1, \ldots, n$ we define the pth exterior product. For $p = 0$ it is $\mathbb{R}$ and for $p = 1$ it is V. For $p = 2$, it consists of formal sums of the form

$$\sum_i a_i (u_i \wedge w_i),$$

where $a_i \in \mathbb{R}$ and $u_i, w_i \in V$. (We say "u_i wedge v_i.") Additionally, we require that $(au + v) \wedge w = a(u \wedge w) + (v \wedge w)$, $u \wedge (bv + w) = b(u \wedge v) + (u \wedge w)$ and $u \wedge u{=}0$. A consequence of the last relation is that $u \wedge w = -w \wedge u$ which we have referred to previously as the anti-commutative law. We further require that if $e_1, \ldots, e_n$ constitute a basis for V, then $e_i \wedge e_j$ for $i < j$, constitute a basis for the second exterior product.

Proceeding analogously, if the e_i form a basis for V we can produce formal sums

$$\sum_\gamma c_\gamma (e_{\gamma_1} \wedge e_{\gamma_2} \wedge \cdots \wedge e_{\gamma_p}),$$

where γ is the multi-index $(\gamma_1, \ldots, \gamma_p)$, where $\gamma_1 < \cdots < \gamma_p$. The expression is multilinear, and the signs change if we transpose any two elements.

The table below lists the exterior products of a vector space $V = \{c_i e_i\}$.

p	pth Exterior Product	Dimension
0	$V^0 = \mathbb{R}$	1
1	$V^1 = V = \{c_i e_i\}$	n
2	$V^2 = \left\{\sum_{i<j} c_{ij} e_i \wedge e_j\right\}$	$n(n-1)/2$
3	$V^3 = \left\{\sum_{i<j<k} c_{ijk} e_i \wedge e_j \wedge e_k\right\}$	$n(n-1)(n-2)/6$
$\vdots$	$\vdots$	$\vdots$
p	$V^p = \left\{\sum_{i_1<i_2<\ldots<i_p} c_{i_1 i_2 \ldots i_p} e_{i_1} \wedge e_{i_2} \wedge \ldots \wedge e_{i_p}\right\}$	$\binom{n}{p}$
$\vdots$	$\vdots$	$\vdots$
n	$V^n = \{c e_1 \wedge e_2 \wedge \ldots \wedge e_n\}$	1
$n+1$	$V^{n+1} = \{0\}$	0

In this book $V = \{\sum c_i \mathrm{d}x_i\}$, i.e. the 1-forms. Then V^p consists of the p-forms, i.e. the rank p exterior differential forms.

9.3 Integration Using Differential Forms

One nice property of our differential form notation is that if $y = y(x)$ is some function from (a subset of) $\mathbb{R}^n$ to $\mathbb{R}^n$, then the formula for changing the volume element is built into the identity

$$\int_{y(S)} f(y)\mathrm{d}y_1 \wedge \ldots \wedge \mathrm{d}y_n = \int_S f(y(x))\mathrm{d}x_1 \wedge \ldots \wedge \mathrm{d}x_n,$$

because the Jacobian emerges when we write the exterior product of the $\mathrm{d}y$'s in terms of the $\mathrm{d}x$'s.

We will only concern ourselves with integration of n-forms on manifolds of dimension n. In fact, most of our manifolds will be flat (subsets of $\mathbb{R}^n$), or surfaces only slightly more complicated than spheres. For example, the Stiefel manifold $V_{m,n}$ of n by m orthogonal matrices Q ($Q^T Q = I_m$) which we shall introduce shortly. Exterior products will give us the correct volume element for integration.

If the x_i are Cartesian coordinates on a flat manifold[1], then $(\mathrm{d}x) \equiv \mathrm{d}x_1 \wedge \mathrm{d}x_2 \wedge \ldots \mathrm{d}x_n$ is the correct volume element. For simplicity, this may be written as $\mathrm{d}x_1\mathrm{d}x_2 \ldots \mathrm{d}x_n$ so as to correspond to the Lebesgue measure. Let q_i be the ith component of a unit vector $q \in \mathbb{R}^n$. Evidently, n parameters is one too many for specifying points on the sphere. Unless $q_n = 0$, we may use q_1 through q_{n-1} as local coordinates on the sphere, and then $\mathrm{d}q_n$ may be thought of as a linear combination of the $\mathrm{d}q_i$ for $i < n$. ($\sum_i q_i \mathrm{d}q_i = 0$ because $q^T q = 1$). However, the Cartesian volume element $\mathrm{d}q_1\mathrm{d}q_2 \ldots \mathrm{d}q_{n-1}$ is not correct for integrating functions on the sphere. It is as if we took a map of the Earth and used Latitude and Longitude as Cartesian coordinates, and then tried to make some presumption about the size of Greenland[2].

Integration:

[1] hyperplanes are flat, an ordinary cylinder is flat (because it can be rolled), but the sphere is not flat

[2] I do not think that I have ever seen a map of the Earth that uses Latitude and Longitude as Cartesian coordinates. The most familiar map, the Mercator map, takes a stereographic projection of the Earth onto the (complex) plane, and then takes the image of the entire plane into an infinite strip simply by taking the complex logarithm.

$\int_{x \in S} f(x)(\mathrm{d}x)$ or $\int_S f(\mathrm{d}x)$ and other related expressions will denote the "ordinary" integral over a region $S \in \mathbb{R}$.

Example. $\int_{\mathbb{R}^n} \exp(-||x||^2/2)(\mathrm{d}x) = (2\pi)^{n^2/2}$ and similarly $\int_{\mathbb{R}^{n,n}} \exp(-||x||^2_{F)/2}(\mathrm{d}A) = (2\pi)^{n^2/2}$. $(||A||_F^2 = \mathrm{tr}(A^TA) = \sum_{i,j} a_{ij}^2) =$"Frobenius norm" of A squared.

If an object has n parameters, the correct differential form for the volume element is an n-form. What about $x \in S^{n-1}$, i.e., $\{x \in \mathbb{R}^n : ||x|| = 1\}$? $\bigwedge_{i=1} \mathrm{d}x_i = (\mathrm{d}x) = 0$. We have $\sum x_i^2 = 1 \Rightarrow \sum x_i \mathrm{d}x_i = 0 \Rightarrow \mathrm{d}x_n = -\frac{1}{x_n}(x_1\mathrm{d}x_1 + \cdots + x_{n-1}\mathrm{d}x_{n-1})$. Whatever the correct volume element for a sphere is, it is not $(\mathrm{d}x)$.

As an example, we revisit spherical coordinates in the next section.

9.4 Better Spherical Coordinates

Students who have seen any integral on the sphere before probably have worked with traditional spherical coordinates or integrated with respect to something labeled "the surface element of the sphere." We mention certain problems with these notations. Before we do, we mention that the sphere is so symmetric and so easy, that these problems never manifest themselves very seriously on the sphere, but they become more serious on more complicated surfaces.

The first problem concerns spherical coordinates: the angles are not symmetric.

They do not interchange nicely. Often one wants to preserve the spherical symmetry by writing $x = qr$, where $r = \|x\|$ and $q = x/\|x\|$. Of course, q then has n components expressing $n-1$ parameters. The n quantities $\mathrm{d}q_1, \ldots, \mathrm{d}q_n$ are linearly dependent. Indeed differentiating $q^Tq = 1$ we obtain that $q^T\mathrm{d}q = \sum_{i=1}^n q_i \mathrm{d}q_i = 0$.

Writing the Jacobian from x to $q_i r$ is slightly awkward. One choice is to write the radial and angular parts separately. Since $\mathrm{d}x = q\mathrm{d}r + \mathrm{d}qr$,

$$q^T\mathrm{d}x = \mathrm{d}r \qquad \text{and} \qquad (I - qq^T)\mathrm{d}x = r\mathrm{d}q\,.$$

We then have that

$$\mathrm{d}x = \mathrm{d}r \wedge (r\mathrm{d}q) = r^{n-1}\mathrm{d}r(\mathrm{d}q),$$

where $(\mathrm{d}q)$ is the surface element of the sphere.

We introduce an explicit formula for the surface element of the sphere. Many readers will wonder why this is necessary. Experience has shown that one can need only set up a notation such as $\mathrm{d}S$ for the surface element of the sphere, and most integrals work out just fine. We have two reason for introducing this formula, both pedagogical. The first is to understand wedge products on a curved space in general. The sphere is one of the nicest curved spaces to begin working with. Our second reason, is that when we work on spaces of orthogonal matrices, both square and rectangular, then it becomes more important to keep track of the correct volume element. The sphere is an important stepping stone for this understanding.

The second problem concern the surface element. If one thinks about how this element is used, it is never particularly quantified. Mostly one uses the symmetry of the sphere to make any issue about its size melt away.

Alternatively, we can derive an expression for the surface element on a sphere. We introduce an orthogonal and symmetric matrix H such that $Hq = e_1$ and $He_1 = q$, where e_1 is the first column of the identity. Then

$$H\mathrm{d}x = e_1\mathrm{d}r + H\mathrm{d}q\, r = \begin{pmatrix} \mathrm{d}r \\ r(H\mathrm{d}q)_2 \\ r(H\mathrm{d}q)_3 \\ \vdots \\ r(H\mathrm{d}q)_n \end{pmatrix}.$$

Thus

$$(\mathrm{d}x) = (H\mathrm{d}x) = r^{n-1}\mathrm{d}r \bigwedge_{i=2}^{n} (H\mathrm{d}q)_i\,.$$

We can conclude that the surface element on the sphere is $(\mathrm{d}q) = \bigwedge_{i=2}^{n}(H\mathrm{d}q)_i$.

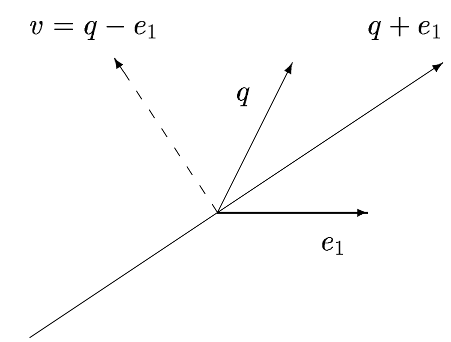
Householder Reflectors

We can explicitly construct and H as described above so as to be the Householder reflector. Choose $v = e_1 - q$ the external angle bisector and

$$H = I - 2\frac{vv^T}{v^Tv}\,.$$

See figure (9.4) which illustrate that H is a reflection through the internal angle bisector of $q + e_1$.

Notice that $(H\mathrm{d}q)_1 = 0$ and every other component $\sum_{j=1}^{n} H_{ij}\mathrm{d}q_j$ $(i \neq 1)$ may be thought of as a tangent on the sphere. $H = H^T, Hq = e_1, He_1 = q_1, H^2 = I$, and H is orthogonal.



Application

Surface Area of Sphere Computation

We directly use the formula $(\mathrm{d}x) = r^{n-1}\mathrm{d}r(\mathrm{d}q)$:

$$\begin{aligned}(2\pi)^{\frac{n}{2}} &= \int_{x\in\mathbb{R}^n} e^{-\frac{1}{2}\|x\|^2}\mathrm{d}x &= \int_{r=0}^{\infty} r^{n-1}e^{-\frac{1}{2}r^2}\mathrm{d}r \int(\mathrm{d}q) \\ &= 2^{\frac{n-2}{2}}\Gamma\left(\frac{n}{2}\right)\int(\mathrm{d}q) \quad \text{or} & \int(\mathrm{d}q) = \frac{2\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2})} = A_n,\end{aligned}$$

and A_n is the surface of the sphere of radius 1. For example,

$$\begin{aligned} A_2 &= 2\pi \quad \text{(circumference of a circle)} \\ A_3 &= 4\pi \quad \text{(area of a sphere in 3d)} \\ A_4 &= 2\pi^2 \end{aligned}$$

Geometrical Interpretation: A Rotating Coordinate System

We introduce the notation $(\mathrm{d}x)$ for $\mathrm{d}x^1 \wedge \ldots \wedge \mathrm{d}x^n$. We will ignore signs, because most of our integrations will be over positive quantities. It then follows that

$$(\mathrm{d}x) = (H\mathrm{d}x) = r^{n-1} dr \wedge_{i=2}^{n} (H\mathrm{d}q)_i.$$

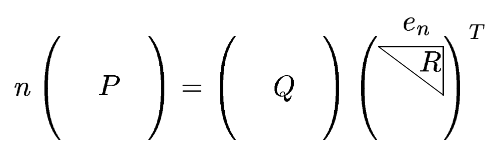
This is a perfect time to review the χ_n distribution, usually seen in the form χ_n^2 ("chi-squared"). If $x \in \mathbb{R}^n$ is a vector of random variables that are independent standard normals, then the joint density for x is

$$(2\pi)^{-n/2} \exp(-\|x\|^2/2)(\mathrm{d}x).$$

The random variable $\|x\|$ is said to have the χ_n distribution. Its probability density is readily computed to be

$$\frac{2^{n/2-1}}{\Gamma(n/2)} e^{-r^2/2} r^{n-1}.$$

9.5 Jacobians for Matrix Factorizations



$$\begin{aligned} H_p \ldots H_1 A &= R \\ A &= (H_1 H_2 \ldots H_p) R \\ A &= HR \\ \mathrm{d}A &= \mathrm{d}HR + H\mathrm{d}R \\ H^T\mathrm{d}A &= (H^T\mathrm{d}H)R + \mathrm{d}R \end{aligned}$$

9.5.1 The QR Decomposition

We now have the framework to compute Jacobians for arbitrary matrix factorizations.

Let $O(m)$ denote the "orthogonal group" of $m \times m$ matrices Q such that $Q^TQ = I$. We have seen $(Q^T\mathrm{d}Q) = \bigwedge_{i>j} q_i^T \mathrm{d}q_j$ is the natural volume element on $O(M)$. Also, notice that $O(n)$ has two connected components. When $m = 2$, we may take

$$Q = \begin{bmatrix} c & -s \\ s & c \end{bmatrix} (c^2 + s^2 = 1)$$

for *half* of $O(2)$. This gives

$$Q^T\mathrm{d}Q = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} -\sin\theta\mathrm{d}\theta & -\cos\theta\mathrm{d}\theta \\ \cos\theta\mathrm{d}\theta & -\sin\theta\mathrm{d}\theta \end{pmatrix} = \begin{pmatrix} 0 & -\mathrm{d}\theta \\ \mathrm{d}\theta & 0 \end{pmatrix}$$

in terms of $\mathrm{d}\theta$. Therefore $(Q^T\mathrm{d}Q) = \mathrm{d}\theta$.

QR Decomposition

Let $A \in \mathbb{R}^{n,m}\,(m \le n)$. Taking $x = a_i$, we see $\exists H_1$ such that $H_1 A$ has the form $\begin{pmatrix} x \\ 0 \\ \vdots \\ 0 \end{pmatrix}$. We can then construct an $H_2 = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & \tilde{H}_2 & \\ 0 & & & \end{pmatrix}$ so that $H_2 H_1 A = \begin{pmatrix} x & x \\ 0 & x \\ \vdots & \vdots \\ 0 & 0 \end{pmatrix}$. Continuing $H_m \cdots H_1 A = \begin{pmatrix} 0 & & \\ & \ddots & \\ & & 0 \\ 0 & \cdots & 0 \end{pmatrix}$ or $A = (H_1 \cdots H_m) \begin{pmatrix} R \\ O \end{pmatrix}$, where R is $m \times m$ upper triangular (with positive diagonals). let $Q =$ the first m columns of $H_1 \cdots H_m$. Then $A = QR$ as desired.

The Stiefel Manifold

The set of $Q \in \mathbb{R}^{n,m}$ such that $Q^T Q = I_m$ is denoted $V_{m,n}$ and is known as the Stiefel manifold. Considering the Householder construction, $\exists H_1, \cdots, H_m$ such that.

$$H_m H_{m-1} \cdots H_1 Q = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix} \; (Why?)$$

so that $Q =$ 1st m columns of $H_1 H_2 \cdots H_{m-1} H_m$.

Corollary 9.1. *The Stiefel manifold may be specified by* $(n-1) + (n-2) + \cdots + (n-m) = mn - 1/2m(m+1)$ *parameters. You may think of this as the* mn *arbitrary parameters in an* $n \times m$ *matrix reduced by the* $m(m+1)/2$ *conditions that* $q_i^T q_j = \delta_{ij}$ *for* $i \ge j$. *You might also think of this as*

$$\dim\{\mathrm{Q}\} = \underset{\substack{\uparrow \\ mn}}{\dim\{A\}} - \underset{\substack{\uparrow \\ m(m+1)/2}}{\dim\{R\}}$$

It is no coincidence that it is more economical in numerical computations to store the Householder parameters than to compute out the Q.

This is the prescription that we would like to follow for the QR decomposition for the n by m matrix A. If $Q \in \mathbb{R}^{m,n}$ is orthogonal, let H be an orthogonal m by m matrix such that $H^T Q$ is the first m columns of I. Actually H may be constructed by applying the Householder process to Q. Notice that Q is simply the first m columns of H.

As we proceed with the general case, notice how this generalizes the situation when $m = 1$. If $A = QR$, then $\mathrm{d}A = Q\mathrm{d}R + \mathrm{d}QR$ and $H^T \mathrm{d}A = H^T Q \mathrm{d}R + H^T \mathrm{d}QR$. Let $H = [h_1, \ldots, h_n]$. The matrix $H^T Q \mathrm{d}R$ is an n by m upper triangular matrix. While $H^T \mathrm{d}Q$ is (rectangularly) antisymmetric. ($h_i^T h_j = 0$ implies $h_i^T \mathrm{d}h_j = -h_j^T \mathrm{d}h_i$)

In matrix form:

$$H^T\mathrm{d}A = \begin{pmatrix} \mathrm{d}r_{11} & \mathrm{d}r_{12} & \dots & \mathrm{d}r_{1m} \\ & \mathrm{d}r_{22} & \dots & \mathrm{d}r_{2m} \\ & & \ddots & \vdots \\ & & & r_{nm} \\ & & & \\ & & & \\ & & & \end{pmatrix} + \begin{pmatrix} 0 & -h_2^T\mathrm{d}h_1 & \dots & -h_m^T\mathrm{d}h_1 \\ h_2^T\mathrm{d}h_1 & 0 & \dots & -h_m^T\mathrm{d}h_2 \\ & & \ddots & \\ h_m^T\mathrm{d}h_1 & h_m^T\mathrm{d}h_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ h_{n-1}^T\mathrm{d}h_1 & h_{n-1}^T\mathrm{d}h_2 & \dots & h_{n-1}^T\mathrm{d}h_m \\ h_n^T\mathrm{d}h_1 & h_n^T\mathrm{d}h_2 & \dots & h_n^T\mathrm{d}h_m \end{pmatrix} R$$

By $(\mathrm{d}R)$ we mean $\bigwedge_{i\leq j}\mathrm{d}r_{ij}$. For the antisymmetric $H^T\mathrm{d}Q$ we take the exterior derivative of all the elements below the diagonal:

$$(H^T\mathrm{d}Q) = \bigwedge_{i>j} h_i^T\mathrm{d}h_j.$$

Theorem 9.2. *The Jacobian of the change of variables $A = QR$ is*

$$(\mathrm{d}A) = \prod_{i=1}^{m} r_{ii}^{n-i}(\mathrm{d}R)(H^T\mathrm{d}Q),$$

where $(H^T\mathrm{d}Q) = \bigwedge_{i>j} h_i^T\mathrm{d}h_j$.

Proof. We first take the exterior derivative of the elements in $H^T\mathrm{d}QR$ that are below the diagonal, one column at a time from left to right. Because of the multiplication of each entry by r_{11}, the first column is $r_{11}^{n-1}\bigwedge_{j=2}^{n} h_j^T\mathrm{d}h_1$. (As it was when we only had a sphere.) The second column is multiplied by r_{22} and then r_{12} times the first column is added to it. However the addition of the first column makes no further contribution to the exterior product because identical differential forms "wedge out" to zero. This pattern continues: r_{jj} multiplies the entries in the jth column $n-j$ of which are below the diagonal, and sums do not make a contribution. This gives $r_{jj}^{n-j}\bigwedge_{i>j} h_i^T\mathrm{d}h_j$. The next step is to take the exterior product with the elements of $\mathrm{d}R + H^T\mathrm{d}QR$ that are on or above the diagonal. This is easy since the terms in $H^T\mathrm{d}QR$ make no further contribution. □

9.5.2 Haar Measure and Volume of the Stiefel Manifold

It is evident that the volume element in mn dimensional space decouples into a term due to the upper triangular component and a term due to the orthogonal matrix component. The differential form

$$(H^T\mathrm{d}Q) = \bigwedge_{j=1}^{m}\bigwedge_{i=j+1}^{n} h_i^T\mathrm{d}h_j$$

is the natural volume element on the Stiefel manifold.

We may define

$$\mu(S) = \int_S (H^T\mathrm{d}Q).$$

This represent the surface area (volume) of the region S on the Stiefel manifold. This "measure" μ is known as Haar measure when $m = n$. It is invariant under orthogonal rotations.

Exercise. *Let* $\Gamma_m(a) \equiv \pi^{m(m-1)/4} \prod_{i=1}^{m} \Gamma[a - \frac{1}{2}(i-1)]$*. Show that the volume of* $V_{m,n}$ *is*

$$\text{Vol } (V_{m,n}) = \frac{2^m \pi^{mn/2}}{\Gamma_m(\frac{1}{2}n)} \ .$$

Exercise. *What is this expression when* $n = 2$*? Why is this number twice what you might have thought it would be?*

Exercise. *Let* A *have independent elements from a standard normal distribution. Prove that* Q *and* R *are independent, and that* Q *is distributed uniformly with respect to Haar measure. How are the elements on the strictly upper triangular part of* R *distributed. How are the diagonal elements of* R *distributed? Interpret the QR algorithm in a statistical sense. (This may be explained in further detail in class).*

Readers who may never have taken a course in advanced measure theory might enjoy a loose general description of Haar measure. If G is any group, then we can define the map on ordered pairs that sends (g, h) to $g^{-1}h$. If G is also a manifold (or some kind of Hausdorff topological space), and if this map is continuous, we have a topological group. An additional assumption one might like is that every $g \in G$ has an open neighborhood whose closure is compact. This is a locally compact topological group. The set of square nonsingular matrices or the set of orthogonal matrices are good examples. A measure $\mu(E)$ is some sort of volume defined on E which may be thought of as nice ("measurable") subsets of G. The measure is a Haar measure if $\mu(gE) = \mu(E)$, for every $g \in G$. In the example of orthogonal n by n matrices, the condition that our measure be Haar is that

$$\int_{Q \in S} f(Q)(Q^T \mathrm{d}Q) = \int_{Q \in Q_0^{-1} S} f(Q_0 Q)(Q^T \mathrm{d}Q).$$

In other words, Haar measure is symmetrical, no matter how we rotate our sets, we get the same answer. The general theorem is that on every locally compact topological group, there exists a Haar measure μ.

9.5.3 Symmetric Eigenproblem

If $S \in \mathbb{R}^{n,n}$ is symmetric it may be written

$$S = Q \Lambda Q^T, Q \in O(n), \Lambda = \dim(\lambda_1, \cdots, \lambda_n)$$

the columns of Q are the eigenvectors, while the λ_i are the eigenvalues.

Computing the Jacobian is performed by the product rule for differentiation and conjugating by Q:

$$\mathrm{d}S = \mathrm{d}Q \Lambda Q^T + Q \mathrm{d}\Lambda Q^T + Q \Lambda \mathrm{d}Q^T$$

$$Q^T \mathrm{d}SQ = [Q^T \mathrm{d}Q]\Lambda - \Lambda[Q^T \mathrm{d}Q] + \mathrm{d}\Lambda$$

Notice that $Q^T \mathrm{d}SQ$ is a symmetric matrix of differentials with $\mathrm{d}\lambda_i$ on the diagonal and $q_j^T \mathrm{d}q_i(\lambda_i - \lambda_j)$ as the i, jth entry in the upper triangular part.

Therefore $(Q^T \mathrm{d}SQ) = \prod_{i<j} |\lambda_i - \lambda_j| (\mathrm{d}\Lambda)(Q^T \mathrm{d}Q)$. We have from Example 3 of Section 8.3 that $(Q^T \mathrm{d}SQ) = (\mathrm{d}S)$.

9.5.4 Singular Values

The singular values of a matrix are far more important than the eigenvalues of a matrix, and yet eigenvalues are taught as part of every standard linear algebra course, and singular values are still not mentioned as often as they should. Singular values have a long history, but their important status was raised by Gene Golub at Stanford, who you may see driving around with his "DR SVD" or "PROF SVD" California license plate.

The singular values may just as readily be defined for matrices $A \in \mathbb{R}^{n,m}$. One definition is simply that σ_i^2 is the ith eigenvalue of the positive definite matrix A^TA. The σ_i are defined to be non-negative. If A is square, but not symmetric, there is little connection between the eigenvalues and the singular values of A. It may be an unfortunate accident of mathematical development that the concept of eigenvalues remains more familiar than the concept of singular values. Because of this, statisticians refer to the matrices A^TA more often than they need to, and proofs become unnecessarily cumbersome.

An alternative definition is that the singular values are the lengths of the semi-axes of the image of the unit ball under the transformation A.

The most useful algebraic definition is the singular value decomposition

$$A = U\Sigma V^T,$$

where U is orthogonal $\in \mathbb{R}^{m,n}$, Σ is diagonal with σ_i in the i,i th position, and V is a square orthogonal matrix $\in \mathbb{R}^{n,n}$. We will assume $n \geq m$, though there is an obvious modification for $n < m$. There are various expanded forms, for instance U could be square, and $\Sigma \in \mathbb{R}^{n,m}$, etc.

If the singular values of A are distinct, then the singular vectors V are defined up to sign, as the eigenvectors of A^TA. If the singular values are positive, this uniquely determines U as $AV^T\Sigma^{-1}$. Therefore we may say that generically the singular value decomposition covers A a total of 2^m times.

$$\begin{aligned} A &= U\Sigma V^T \\ \mathrm{d}A &= \mathrm{d}U\Sigma V^T + U\mathrm{d}\Sigma V^T + U\Sigma \mathrm{d}V^T \\ U^T\mathrm{d}AV &= U^T\mathrm{d}V\Sigma + \mathrm{d}\Sigma\Sigma V^T\mathrm{d}V \end{aligned}$$

9.6 Advanced Differential Forms

9.6.1 Multilinear Functions

Mathematically, the wedge notation $w_1 \wedge \cdots \wedge w_k$ may be identified with a real linear function on n by k matrices. Specifically, let the matrix $W = [w_1 \ldots w_k]$, and consider the linear function $T_W(V) \equiv \det(W^TV)$. A moment's thought will convince the reader that this is a multilinear function, and if we interchange two columns of W, we negate the function. Real combinations of such functions are in one to one correspondence with real combinations of wedge products.

Perhaps we ought to define a slightly more general object: a rank k tensor. Let $T(v_1, \ldots, v_k)$ denote a real valued multilinear function of the k vectors $v_i \in \mathbb{R}^n$. This means that if α and β are scalars, then for each i

$$T(v_1, \ldots, \alpha v_i + \beta v_i', \ldots, v_k) = \alpha T(v_1, \ldots, v_i, \ldots, v_k) + \beta T(v_1, \ldots, v_i', \ldots, v_k).$$

In terms of the components, any multilinear function may be written as

$$T(v) = \sum T_{i_1,\ldots,i_k} v_{i_1,1} \ldots, v_{i_k,k}.$$

If we have a collection of k vectors such as $v_1, \ldots, v_k$ in $\mathbb{R}^n$, it is notationally convenient to construct the matrix V whose jth column is the jth vector. ($V = [v_1 \ldots v_k]$)

Therefore, a multilinear function may be thought of as a a "square" k-dimensional array of n^k numbers, i.e., an $n \times n \times \ldots \times n$ array of numbers. It is clear that the set of multilinear functions form a vector space of dimension n^k, with the usual definition for linear combinations of functions:

$$(\alpha T_1 + \beta T_2)(V) = \alpha T_1(V) + \beta T_2(V).$$

When $k = 0$, T is scalar. When $k = 1$, given any vector $w \in \mathbb{R}^n$, we have $T_w(v) = w^T v$. When $k = 2$, given any n by n matrix A, we have $T_A(v_1, v_2) = v_1^T A v_2$. For $k > 2$ elementary linear algebra notation breaks down, but the idea remains straightforward. (Is that because we are three dimensional creatures used to writing on two dimensional paper?)

Elementary linear algebra notations, however, is just perfect for the multilinear functions that we are considering: Let W be an n by k matrix and define $T_W(V)$ (here $V = [v_1, \ldots, v_k]$) as $\det(W^T V)$. Notice that we are taking the determinant of a k by k matrix.

Exercise. *Prove that T_W is indeed a rank k multilinear function using nothing other than familiar properties of the determinant.*

When $k = n$, it follows from the identity $\det(W^T V) = \det(W)\det(V)$ that $T_W = (\det W) T_I$, where I denotes the n by n identity matrix. When $k > n$ we are taking the determinant of a matrix of dimension greater than n, but of rank at most n. Therefore $T_W = 0$ when $k > n$.

Since the tensors of the form T_W are a subset of an n^k dimensional vector space, we may form the vector space Anti generated by all possible linear combinations of the tensors T_W. This space is known either as the as the set of **antisymmetric** tensors or **alternating** tensors. This space is isomorphic to the kth exterior product.

Antisymmetric tensors T have the property that if V has two identical columns, then $T(V) = 0$, and further if we interchange two columns of V to create a V', then $T(V') = -T(V)$. To prove this, note that this statement follows from the determinant for the tensors of the form T_W, and therefore this property holds for linear combinations of such tensors.

Now we turn to the algebra of differential forms. So far, all you have seen are algebraic objects, tensors or in particular antisymmetric tensors. By the magic of switching notation, but using no further tricks, we will create an object that looks like a volume element for integration.

Let W be a matrix each of whose columns contains $n-1$ zeros, and one value 1. We may write $W = [e_{i_1}, \ldots, e_{i_k}]$, where e_j denotes the jth unit vector (i.e., jth column of the identity matrix.)

As a matter of notation, we will write the tensor T_W as

$$\mathrm{d}x_{i_1} \wedge \mathrm{d}x_{i_2} \wedge \ldots \wedge \mathrm{d}x_{i_k}$$

and we will start to forget that this was once a tensor. The reader may verify that if any of the two i_j's are equal, then we have the 0 tensor, and if we interchange two of the i's say i_1 and i_2, then we negate the tensor.

Exercise. *Show that the tensors of the form*

$$\mathrm{d}x_{i_1} \wedge \mathrm{d}x_{i_2} \wedge \ldots \wedge \mathrm{d}x_{i_k},$$

for $i_1 < i_2 < \ldots < i_k$ form a basis for antisymmetric tensors. Therefore, it is a vector space of dimension $\binom{n}{k}$.

For any matrix T_W we may write the tensor in this notation as

connection
not clear
Too late

$$\left(\sum_i w_{i1}\mathrm{d}x_i\right) \wedge \left(\sum_i w_{i2}\mathrm{d}x_i\right) \wedge \ldots \wedge \left(\sum_i w_{ik}\mathrm{d}x_i\right) .$$

It becomes very easy to algebraically expand this in terms of our basis. We simply assume that sums distribute, and wedges alternate. In particular, when $n = k$, we see directly that $T_W = (\det W)T_I$. This could have been our definition of wedge products. It would have been a bit simpler, but then you might not have seen what all this has to do with tensors. (This approach is taken in Muirhead's book, for example.) What is this?

9.6.2 Differential Forms

It is easy to see that every rank 0 and rank 1 multilinear function is trivially antisymmetric. A rank two tensor: $T(v_1, v_2) = v_1^T A v_2$ is antisymmetric if and only if, A is an antisymmetric matrix, i.e., $A^T = -A$. The reader should notice that we defined rank 2 alternating tensors in terms of linear combinations of functions derived from n by 2 matrices W, and now we are noting that all alternating tensors may be expressed as antisymmetric matrices. Compare both definitions closely.

What about rank 3 antisymmetric tensors? Since such a tensor is a multilinear function, it may be represented as an $n \times n \times n$ cubical array of entries T_{ijk}, $1 \leq i, j, k \leq n$. If you can imagine holding this cube by the two corners at the 1,1,1 entry and the n, n, n entry, then the array of numbers is invariant under a 120 degree rotation through this axis. Other symmetries (in fact reflections) of the cube preserving these two points, negate the entries. Thus we see the generalization of transposing, and antisymmetric matrices. Just an an antisymmetric matrix is determined by its $\binom{n}{2}$ entries in the upper triangular part, a rank 2 tensor is determined by the $\binom{n}{3}$ entries in an upper tetrahedral part covering nearly one sixth of the array. This idea generalizes as well.

We now digress onto a brief discussion of how differential forms fit into other areas of mathematics and physics. The reader primarily interested in eigenvalues of random matrices may safely omit this section.

An *exterior differential form* of rank or degree k may be thought of as an antisymmetric multilinear function at every point $x \in \mathbb{R}^n$:

$$\phi = \phi(x) = \sum_{i_1 < \ldots < i_k} f_{i_1,\ldots,i_k}(x)\mathrm{d}x_{i_1} \wedge \ldots \wedge \mathrm{d}x_{i_k}.$$

Usually the coefficient functions $f_{i_1,\ldots,i_k}(x)$ are taken to be sufficiently differentiable or analytic for whatever purpose one has in mind. The simplest example is a rank 0 form, which is nothing other than a function $f(x)$ defined on $\mathbb{R}^n$. A rank 1 form may be thought of nothing other than a function from $\mathbb{R}^n$ to $\mathbb{R}^n$. We may associate, $v(x)$ with $v_1(x)\mathrm{d}x_1 + \cdots + v_n(x)\mathrm{d}x_n$. If f is a differentiable function, we may consider its gradient as

$$\frac{\partial f}{\partial x_1}\mathrm{d}x_1 + \cdots + \frac{\partial f}{\partial x_n}\mathrm{d}x_n.$$

Thus the action of taking a gradient turns a 0 form into a 1 form.

In general, if ϕ is a differential k form, we can form a differential $k+1$ form $\mathrm{d}\phi$ by generalizing the idea of the gradient:

$$\mathrm{d}\phi = \sum_{j=0}^{n} \sum_{i_1 < \ldots < i_k} \frac{\partial f_{i_1,\ldots,i_k}}{\partial x_j}\mathrm{d}x_j \wedge \mathrm{d}x_{i_1} \wedge \ldots \wedge \mathrm{d}x_{i_k}.$$

Let $n = 3$. If ϕ is a 0-form, i.e., a function, $\mathrm{d}\phi$ is its gradient. As a tensor, $\mathrm{d}\phi(v)$ computes a directional derivative at x in the direction v. If ϕ is a 1-form, i.e., a vector function of $\mathbb{R}^3$, sometimes called a vector field, then $\mathrm{d}\phi$ is an object you may recognize: it is the curl of the vector field. Lastly, if ϕ is the 2-form $\phi = f_1 \mathrm{d}x_2 \wedge \mathrm{d}x_3 + f_2 \mathrm{d}x_3 \wedge \mathrm{d}x_1 + f_3 \mathrm{d}x_1 \wedge \mathrm{d}x_2$, then $\mathrm{d}\phi$ is what you might remember from your advanced calculus days as the divergence of $(f_1, f_2, f_3)^T$.

Exercise. *Prove that* $\mathrm{d}\mathrm{d}\phi = 0$ *always.*

A form ϕ is called closed if $\mathrm{d}\phi = 0$, while ϕ is called exact if $\phi = \mathrm{d}\theta$ for some form θ. Exercise 1 may be reworded: prove that if ϕ is exact, then ϕ must be closed. Under the right assumptions, the converse also holds. (See an advanced book on calculus on manifolds.) Exercise 1?

The idea of a differential form can be well defined on arbitrary manifolds, but this is beyond the scope of this course. The basic idea remains the same as you see here, but it is necessary to first define coordinates on the manifold, then define differential forms on these coordinates in a consistent manner.

9.6.3 Differential Forms in Physics

For readers curious how differential forms are used in physics, we express Maxwell's equations in differential form.

At every point $x \in \mathbb{R}^3$ can be found an electrical field vector $E(x) \in \mathbb{R}^3$ and a magnetic field vector $B(x) \in \mathbb{R}^3$. The electrical field vector describes how a charged particle will be influenced at that point by electrical attraction and repulsion, and the magnetic field describes the influence of the magnetic field on the same particle. If we add on x_4 as the time coordinate, then $E(x)$ and $B(x)$ describe the fields at a particular point in space at a particular time.

Let F denote the antisymmetric matrix

$$\begin{pmatrix} 0 & -E_1 & -E_2 & -E_3 \\ E_1 & 0 & B_3 & -B_2 \\ E_2 & -B_3 & 0 & B_1 \\ E_3 & B_2 & -B_1 & 0 \end{pmatrix}$$

or the associated tensor $F = -E_1 \mathrm{d}x_1 \wedge \mathrm{d}x_2 - E_2 \mathrm{d}x_1 \wedge \mathrm{d}x_3 - E_3 \mathrm{d}x_1 \wedge \mathrm{d}x_4 + B_3 \mathrm{d}x_2 \wedge \mathrm{d}x_3 - B_2 \mathrm{d}x_2 \wedge \mathrm{d}x_4 + B_1 \mathrm{d}x_3 \wedge \mathrm{d}x_4$.

Special relativity combines all the electrical and magnetic forces on an electron into one matrix multiply Fu where u is the relativized velocity vector: $u_i = \frac{v_i}{\sqrt{1-v^2}}$, for $i = 1, 2, 3$, and $u_4 = \frac{1}{\sqrt{1-v^2}}$ in units such that the speed of light is 1.

Two of Maxwell's equations may be obtained from the equation $\mathrm{d}F = 0$. These equations are known as the magnetostatic and magnetodynamic equations. Feel free to derive them for yourself, and as a check, find someone with a T-shirt that has Maxwell's equations written in differential form.

The other operator on differential forms is the divergence. The divergence turns k forms into $k-1$ forms. It is linear and defined on $\phi = g(x) \mathrm{d}x_{i_1} \wedge \ldots \wedge \mathrm{d}x_{i_k}$ as

$$\nabla \cdot \phi = \sum_{j=1}^{k} \frac{\partial g(x)}{\partial x_{i_j}} \bigwedge_{m \neq j} \mathrm{d}x_{i_m}.$$

The bigwedge notation indicates the term $\mathrm{d}x_{i_1} \wedge \mathrm{d}x_{i_2} \wedge \ldots \wedge \mathrm{d}x_{i_k}$ with the $\mathrm{d}x_{i_j}$ term omitted.

If we may introduce another physical quantity $J(x) \in \mathbb{R}^4$, whose first three components are the current density, and whose last component in the charge density at $x \in \mathbb{R}^4$, then the other two Maxwell's equations are $\nabla \cdot F = J$. These are the electrostatic and electrodynamic equations.

In summary, Maxwell's equations are

$$\mathrm{d}F = 0 \quad \text{and} \quad \nabla \cdot F = J \ .$$

These two equations can be combined into one Poisson like equation $\nabla \cdot \mathrm{d}A = J$, where $\mathrm{d}A = F$, but perhaps we have digressed enough.

One can also integrate differential forms over appropriate manifolds generalizing the famous Stokes' and Green's formulas.