18.325: Finite Random Matrix Theory
Jacobians of Matrix Transforms (with wedge products)

Professor Alan Edelman

Handout #3, Tuesday, February 15, 2005

There is a wedge product notation that can facilitate the computation of matrix Jacobians. In point of fact, it is never needed at all. The reader who can follow the derivation of the Jacobian in Handout #2 is well equipped to never use wedge products. The notation also expresses the concept of volume on curved surfaces.

For advanced readers who truly wish to understand exterior products from a full mathematical viewpoint, this handout contains the important details. We took some pains to write a readable account of wedge products at the cost of straying way beyond the needs of random matrix theory.

The steps to understanding how to use wedge products for practical calculations are very straightforward.

1. Learn how to wedge two quantities together. This is usually understood instantly.
2. Recognize that the wedge product is a formalism for computing determinants. This is understood instantly as well, and at this point many readers wonder what else is needed.
3. Practice wedging the order n^2 or mn or some such entries of a matrix together. This requires mastering three good examples.
4. Learn the mathematical interpretation of the wedge product of such quantities as the lower triangular part of $Q^T\mathrm{d}Q$ and how this relates to the Jacobian for QR or the symmetric eigendecomposition. We provide a thorough explanation.

1 The Mechanics of Wedging

The algebra of wedge products is very easy to master. We will wedge together differentials as illustrated in this example:

$$\begin{aligned}(2\mathrm{d}x + x^2\mathrm{d}y + 5\mathrm{d}w + 2\mathrm{d}z) \wedge (y\mathrm{d}x - x\mathrm{d}y) = \\ (-2x - x^2y)\mathrm{d}x \wedge \mathrm{d}y + 5y(\mathrm{d}w \wedge \mathrm{d}x) - 5x(\mathrm{d}w \wedge \mathrm{d}y) - 2y(\mathrm{d}x \wedge \mathrm{d}z) + 2x(\mathrm{d}y \wedge \mathrm{d}z)\end{aligned} \tag{1}$$

Formally the wedge product 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

$$(p\mathrm{d}u + q\mathrm{d}v) \wedge (r\mathrm{d}u + s\mathrm{d}v) = (ps - qr)(\mathrm{d}u \wedge \mathrm{d}v)$$

It therefore follows that

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

In general

$$\sum f_i(x)\,\mathrm{d}x_i \,\wedge\, \sum g_j(x)\,\mathrm{d}x_j = \sum_{i<j}(f_i(x)g_j(x) - f_j(x)g_i(x))\,\mathrm{d}x_i \wedge \mathrm{d}x_j = \sum_{i<j}\begin{vmatrix} f_i(x) & g_i(x) \\ f_j(x) & g_j(x)\end{vmatrix}\mathrm{d}x_i \wedge \mathrm{d}x_j\,.$$

We can wedge together more than two differentials. For example

$$2\mathrm{d}x \wedge (3\mathrm{d}x + 5\mathrm{d}y) \wedge 7(\mathrm{d}x + \mathrm{d}y + \mathrm{d}z) = 70\mathrm{d}x \wedge \mathrm{d}y \wedge \mathrm{d}z \tag{2}$$

Let us write down concretely the wedge product.

Rewriting the two examples in (1) and (2) in matrix notation, we have

$$F = \begin{pmatrix} 2 & y \\ x^2 & -x \\ 5 & 0 \\ 2 & 0 \end{pmatrix} \qquad \text{and} \qquad F = \begin{pmatrix} 2 & 3 & 7 \\ 0 & 5 & 7 \\ 0 & 0 & 7 \end{pmatrix}$$

In the first case, we have computed all 2×2 subdeterminants and in the second case we compute the one 3×3 determinant. In general, if $F \in \mathbb{R}^{n,p}$, we compute all $\binom{n}{p}$ subdeterminants of size p.

If $F(x) \in \mathbb{R}^{n,p}$ and $\mathrm{d}x$ is the vector $(\mathrm{d}x_1, \mathrm{d}x_2, \ldots, \mathrm{d}x_n)^T$, then we can wedge together the elements of $F(x)^T\mathrm{d}x$. The result is

$$\bigwedge_{i=1}^{p} (F(x)^T\mathrm{d}x)_i = \sum_{i_1 < i_2 < \ldots < i_p} F\begin{pmatrix} i_1 & i_2 & \ldots & i_p \\ 1 & 2 & \ldots & p \end{pmatrix} \mathrm{d}x_{i_1} \wedge \cdots \wedge \mathrm{d}x_{i_p}\,,$$

where $F\begin{pmatrix} i_1 & i_2 & \ldots & i_p \\ 1 & 2 & \ldots & p \end{pmatrix}$ denotes the subdeterminant of F obtained by taking rows $i_1, i_2, \ldots, i_p$ and all p columns. In almost MATLAB notation, $F\begin{pmatrix} i_1 & i_2 & \ldots & i_p \\ 1 & 2 & \ldots & p \end{pmatrix} = \det(F[i_1\ i_2\ \ldots\ i_p, :])$. In simple English, if we wedge together p differentials which we can store in the columns of F, then the wedge product computes all $p \times p$ subdeterminants.

We use the notation

$$(F(x)^T\mathrm{d}x)^\wedge \equiv \bigwedge_{i=1}^{p} (F(x)^T\mathrm{d}x)_i\,,$$

i.e., "$(\qquad)^\wedge$" denotes wedge together all the elements of the vector inside the parentheses.

A notational note: We are inventing the "$(\qquad)^\wedge$" notation. Books such as [1] use only parentheses to indicate the wedge product of the components inside. Unfortunately, we have found that parenthesis notation is ambiguous especially when we want to wedge together the elements of a complicated expression. Once we decided on placing the wedge symbol "$\wedge$" somewhere, we experimented with upper/lower left and upper/lower right, finally settling on the upper right notation as above.

We will extend the "$(\qquad)^\wedge$" notation from vectors to matrices of differentials. We will only wedge together the independent elements. For example

$$(\mathrm{d}M)^\wedge = \bigwedge_{\substack{1\le i\le m \\ 1\le j\le n}} \mathrm{d}M_{ij} \qquad M \in \mathbb{R}^{mn}\,.$$

We use subscripts to indicate which elements to wedge over. For example $(\mathrm{d}S)^\wedge_{i\ge j} = \bigwedge_{i\le j} \mathrm{d}S_{ij}$. When unnecessary as in the cases below, we omit the subscripts.

$$\begin{array}{llllll}
(\mathrm{d}S)^\wedge & = & \bigwedge_{1\le i\le j\le n} \mathrm{d}S_{ij} & S \in \mathbb{R}^{nn} & \text{symmetric} \\
(\mathrm{d}A)^\wedge & = & \bigwedge_{1\le i< j\le n} \mathrm{d}A_{ij} & A \in \mathbb{R}^{nn} & \text{antisymmetric} \\
(\mathrm{d}\Lambda)^\wedge & = & \bigwedge_{1\le i\le n} \mathrm{d}\Lambda_{ii} & \Lambda \in \mathbb{R}^{nn} & \text{diagonal} \\
(\mathrm{d}U)^\wedge & = & \bigwedge_{1\le i\le j\le n} \mathrm{d}U_{ij} & U \in \mathbb{R}^{nn} & \text{upper triangular} \\
(\mathrm{d}U)^\wedge & = & \bigwedge_{1\le i< j\le n} \mathrm{d}U_{ij} & U \in \mathbb{R}^{nn} & \text{strictly upper triangular} \\
\text{etc.} & & & &
\end{array}$$

Definitional note: We have not specified the order. Therefore the wedge product of elements of a matrix is defined up to "+" or "−" sign.

Notational note: We write $(\mathrm{d}M_1)^\wedge(\mathrm{d}M_2)^\wedge$ for $(\mathrm{d}M_1)^\wedge \wedge (\mathrm{d}M_2)^\wedge$.

2 Jacobians with wedge products

2.1 Wedge Notation not useful ($Y = X^2$)

Consider the 2×2 case of $Y = X^2$ as in Example 1 of Handout #2.

With $X = \begin{bmatrix} p & q \\ r & s \end{bmatrix}$

$$\mathrm{d}Y = X\mathrm{d}X + \mathrm{d}XX = \begin{bmatrix} p & q \\ r & s \end{bmatrix} \begin{bmatrix} \mathrm{d}p & \mathrm{d}q \\ \mathrm{d}r & \mathrm{d}s \end{bmatrix} + \begin{bmatrix} \mathrm{d}p & \mathrm{d}q \\ \mathrm{d}r & \mathrm{d}s \end{bmatrix} \begin{bmatrix} p & q \\ r & s \end{bmatrix}.$$

Thus

$$\begin{aligned} \mathrm{d}Y_{11} &= 2p\mathrm{d}p + q\mathrm{d}r + r\mathrm{d}q \\ \mathrm{d}Y_{21} &= r\mathrm{d}p + (p+s)\mathrm{d}r + r\mathrm{d}s \\ \mathrm{d}Y_{12} &= q\mathrm{d}p + (p+s)\mathrm{d}q + q\mathrm{d}s \\ \mathrm{d}Y_{22} &= q\mathrm{d}r + r\mathrm{d}q + 2s\mathrm{d}s\,. \end{aligned} \tag{3}$$

If we wedge together all of the elements of Y, we obtain the determinant of the Jacobian:

$$(\mathrm{d}Y)^\wedge = 4(p+s)^2(sp - qr)\,.$$

The reader should compare the notation with that from the previous chapter:

$$J = \begin{array}{cccccc} & \partial p & \partial r & \partial q & \partial s & \\ \left[\begin{array}{c} \\ \\ \\ \\ \end{array}\right. & \begin{array}{c} 2p \\ r \\ q \\ 0 \end{array} & \begin{array}{c} q \\ p+s \\ 0 \\ q \end{array} & \begin{array}{c} r \\ 0 \\ p+s \\ r \end{array} & \begin{array}{c} 0 \\ r \\ q \\ 2s \end{array} & \left.\begin{array}{c} \\ \\ \\ \\ \end{array}\right] \begin{array}{c} \partial Y_{11} \\ \partial Y_{21} \\ \partial Y_{12} \\ \partial Y_{22} \end{array} \end{array}$$

We conclude that for these examples the notation is of little value.

2.2 Square Matrix = Antisymmetric + Upper Triangular

Given any matrix $M \in \mathbb{R}^{n,n}$, let

$$A = \mathrm{lower}(M) - \mathrm{lower}(M)^T,$$

where lower(M) is the strictly lower triangular part of M. Further let

$$R = \mathrm{upper}(M) + \mathrm{lower}(M)^T,$$

where upper(M) includes the diagonal. We have the decomposition

$$\boxed{M} = \boxed{A} + \diagdown R = (\text{antisymmetric}) + (\text{upper triangular})\,.$$

Therefore

$$\mathrm{d}M = \begin{pmatrix} \mathrm{d}r_{11} & \mathrm{d}r_{12} - \mathrm{d}a_{21} & \mathrm{d}r_{13} - \mathrm{d}a_{31} & \cdots & \mathrm{d}r_{1n} - \mathrm{d}a_{n1} \\ \mathrm{d}a_{21} & \mathrm{d}r_{22} & \mathrm{d}r_{23} - \mathrm{d}a_{32} & & \mathrm{d}r_{2n} - \mathrm{d}a_{n2} \\ \mathrm{d}a_{31} & \mathrm{d}a_{32} & \mathrm{d}r_{33} & & \mathrm{d}r_{3n} - \mathrm{d}a_{n3} \\ \vdots & \vdots & \vdots & & \vdots \\ \mathrm{d}a_{n1} & \mathrm{d}a_{n2} & \mathrm{d}a_{n3} & & \mathrm{d}r_{nn} \end{pmatrix}. \tag{4}$$

We can easily see that the $-\mathrm{d}a_{ij}$ play no role in the wedge product:

$$(\mathrm{d}M)^\wedge = (\mathrm{lower}\ (\mathrm{d}A) + \mathrm{d}R)^\wedge = (\mathrm{d}A)^\wedge(\mathrm{d}R)^\wedge\,.$$

We feel this notation is already somewhat more mechanical than what was needed in the last chapter. Without wedges, we would have concentrated on the $n(n+1)/2$ parameters in R and the $n(n-1)/2$ in lower(A) and would have written a block two by two matrix as we did in Handout #2.

The reader should think carefully about the following sentence: The realization that the upper $\mathrm{d}a_{ji}$'s play no role in $(\mathrm{d}M)^\wedge$, is equivalent to the realization that the matrix X_{12} plays no role in the determinant of the 2×2 block matrix

$$X = \begin{bmatrix} X_{11} & X_{12} \\ 0 & X_{21} \end{bmatrix}.$$

2.3 Triangular $\longleftrightarrow$ Symmetric

Let $S = U + U^T =$ ◹ $+$ ◺. If we consider the map from U to upper(S), it is easy to see that the Jacobian determinant is 2^n.

$$\begin{aligned} (\mathrm{d}S)^\wedge &= (\mathrm{diag}(\mathrm{d}S))^\wedge \wedge (\text{strictly-upper}(\mathrm{d}S))^\wedge \\ &= 2^n (\mathrm{diag}(\mathrm{d}U))^\wedge \wedge (\text{strictly-upper}(\mathrm{d}U))^\wedge \\ &= 2^n (\mathrm{d}U)^\wedge. \end{aligned}$$

2.4 General matrix functions

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

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

The wedge product of all the elements is

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

In general, if $Y = Y(X)$ is an $n \times n$ matrix, then

$$(\mathrm{d}Y)^\wedge = (\det J)(\mathrm{d}X)^\wedge,$$

where $\det J$ denotes the Jacobian determinant $\left(\frac{\partial y_{ij}}{\partial x_{kl}}\right)$.

2.5 Eigenproblem Jacobians

Case I: S Symmetric

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

$$S = Q\Lambda Q^T, Q^T Q = I, \Lambda = \mathrm{diag}(\lambda_1, \cdots, \lambda_n).$$

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

Differentiating,

$$\begin{aligned} \mathrm{d}S &= \mathrm{d}Q\Lambda Q^T + Q\mathrm{d}\Lambda Q^T + Q\Lambda \mathrm{d}Q^T, \\ Q^T\mathrm{d}SQ &= (Q^T\mathrm{d}Q)\Lambda - \Lambda(Q^T\mathrm{d}Q) + \mathrm{d}\Lambda. \end{aligned}$$

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

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

$$(\mathrm{d}S)^\wedge = \prod_{i<j} |\lambda_i - \lambda_j| (\mathrm{d}\Lambda)^\wedge (Q^T\mathrm{d}Q)^\wedge.$$

Many readers will wonder about the meaning of $(Q^T\mathrm{d}Q)^\wedge$. We did. We will discuss this at length in later handouts.

Case II: A **General**

If $A \in \mathbb{R}^{n,n}$ is diagonalizable if may be written

$$A = X\Lambda X^{-1}, \qquad \mathrm{diag}(X^TX) = I, \qquad \Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n),$$

the columns of X are the normalized eigenvectors, while the λ_i are the eigenvalues.

Differentiating,

$$\begin{array}{rcl} \mathrm{d}A & = & \mathrm{d}X\Lambda X^{-1} + X\mathrm{d}\Lambda X^{-1} - X\Lambda X^{-1}\mathrm{d}XX^{-1}, \\ X^{-1}\mathrm{d}AX & = & (X^{-1}\mathrm{d}X)\Lambda - \Lambda(X^{-1}\mathrm{d}X) + \mathrm{d}\Lambda\,. \end{array}$$

Notice that $X^{-1}\mathrm{d}AX$ is a matrix of differentials with $\mathrm{d}\lambda_i$ on the diagonal and $(X^{-1}\mathrm{d}X)_{ij}(\lambda_i - \lambda_j)$ as the i, j entry off the diagonal.

Therefore $(X^{-1}\mathrm{d}AX)^\wedge = \prod_{i \neq j}(\lambda_i - \lambda_j)(\mathrm{d}\Lambda)^\wedge_{i=j}(X^{-1}\mathrm{d}X)^\wedge_{i \neq j}$. We have from before, that $(X^{-1}\mathrm{d}AX)^\wedge = (\mathrm{d}A)^\wedge$ so finally

$$(\mathrm{d}A)^\wedge = \prod_{i<j}(\lambda_i - \lambda_j)^2(\mathrm{d}\Lambda)^\wedge(X^{-1}\mathrm{d}X)^\wedge_{i \neq j}$$

Case III: Generalized Eigenvalue Problem

Consider $S, T \in \mathbb{R}^{n,n}$ symmetric and T positive definite. We consider the "generalized" eigenvalue problem.

need to work out

$$SX = TX\Lambda\,.$$

2.6 The QR Factorization

Any matrix A may be written as QR (or YR when not square), where $Q^TQ = I$ ($Y^TY = I$), and R is upper triangular. For readers unfamiliar with the construction we outline it

Case I: $A \in \mathbb{R}^{n,n}$ ("Square Case")

Theorem 1. *If $A = QR$, where Q is orthogonal and R is upper triangular, then*

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

Proof. If $A = QR$, then $Q^T\mathrm{d}AR^{-1} = Q^T\mathrm{d}Q + \mathrm{d}RR^{-1}$. Differentially, we have decomposed $Q^T\mathrm{d}AR^{-1}$ into antisymmetric plus upper triangular as in Section 9.2.2. Thus

$$\begin{array}{ccccc} & & (Q^T\mathrm{d}AR^{-1})^\wedge & = & (Q^T\mathrm{d}Q)^\wedge(\mathrm{d}R\,R^{-1})^\wedge \\ & = & ((R^{-1} \otimes Q)^T\mathrm{d}A)^\wedge & & \\ & = & (\mathrm{d}A)^\wedge \prod r_{ii}^{-n} & = & \prod r_{ii}^{-i}(Q^T\mathrm{d}Q)^\wedge(\mathrm{d}R)^\wedge \end{array}$$

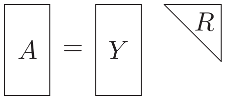
from which the result follows. Readers may wish to see written out $Q^T\mathrm{d}A = \mathrm{d}R + Q^T\mathrm{d}QR$:

$$Q^T\mathrm{d}A = \begin{pmatrix} \mathrm{d}r_{11} & \mathrm{d}r_{12} & \ldots & \mathrm{d}r_{1n} \\ & \mathrm{d}r_{22} & \ldots & \mathrm{d}r_{2n} \\ & & \ddots & \vdots \\ & & & r_{nn} \end{pmatrix} + \begin{pmatrix} 0 & -q_2^T\mathrm{d}q_1 & \ldots & -q_n^T\mathrm{d}q_1 \\ q_2^T\mathrm{d}q_1 & 0 & \ldots & -q_n^T\mathrm{d}q_2 \\ & & \ddots & \\ q_n^T\mathrm{d}q_1 & q_n^T\mathrm{d}q_2 & \ldots & 0 \end{pmatrix} R$$

Theorem 1 may be alternatively derived by wedging together the elements in the above matrix first by wedging the lower triangular part one column at a time from left to right. ☐

Case II: $A \in \mathbb{R}^{np}$ ("tall skinny") $(n \geq p)$

Suppose $A = YR$ where $Y^TY = I_p$ and R is upper triangular:



The reader can try to directly wedge the elements together.

Let $Q \in \mathbb{R}^{n,n}$ be any orthogonal matrix such that $Q^TY = I_{n,p}$, "the first p columns of the $n \times n$ identity." In other words, Q is a completion of the columns of Y to an orthogonal basis.

Theorem 2. *If* $A = YR$*, then* $(\mathrm{d}A)^\wedge = \prod_{i=1}^{p} r_{ii}^{n-i}(Q^T\mathrm{d}Y)^\wedge(\mathrm{d}R)^\wedge$.

Note: $Q^T\mathrm{d}Y$ is a tall-skinny matrix with anti-symmetric top p rows. Thus $(Q^T\mathrm{d}Y)^\wedge$ is a wedge product over $p(p-1)/2 + p(n-p)$ elements.

Proof. The proof is as in Theorem 9.1: $Q^T\mathrm{d}AR^{-1} = \mathrm{d}RR^{-1} + Q^T\mathrm{d}Y$ so that $(Q^T\mathrm{d}AR^{-1})^\wedge = \prod r_{ii}^{i}(Q^T\mathrm{d}Y)^\wedge(\mathrm{d}R)^\wedge = (\mathrm{d}A)^\wedge \prod_{i=1}^{p} r_{ii}^{\wedge}$, or explicitly:

$$Q^T\mathrm{d}A = \begin{pmatrix} \mathrm{d}r_{11} & \mathrm{d}r_{12} & \dots & \mathrm{d}r_{1p} \\ & \mathrm{d}r_{22} & \dots & \mathrm{d}r_{2p} \\ & & \ddots & \vdots \\ & & & r_{pp} \\ \\ \\ \end{pmatrix} + \begin{pmatrix} 0 & -q_2^T\mathrm{d}q_1 & \dots & -q_p^T\mathrm{d}q_1 \\ q_2^T\mathrm{d}q_1 & 0 & \dots & -q_p^T\mathrm{d}q_2 \\ & & \ddots & \\ q_p^T\mathrm{d}q_1 & q_p^T\mathrm{d}q_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ q_{n-1}^T\mathrm{d}q_1 & q_{n-1}^T\mathrm{d}q_2 & \dots & q_{n-1}^T\mathrm{d}q_p \\ q_n^T\mathrm{d}q_1 & q_n^T\mathrm{d}q_2 & \dots & q_n^T\mathrm{d}q_p \end{pmatrix} R$$

We first take the exterior product of the elements in $Q^T\mathrm{d}QR$ 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 $r_{11}^{n-1}\bigwedge_{j=2}^{n} q_j^T\mathrm{d}q_1$. (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 $r_{jj}^{n-j}\bigwedge_{i>j} q_i^T\mathrm{d}q_j$. The next step is to take the exterior product with the elements of $\mathrm{d}R + Q^T\mathrm{d}QR$ that are on or above the diagonal. This is easy since the terms in $Q^T\mathrm{d}QR$ make no further contribution. ☐

2.7 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.

The singular values may just as readily be defined for matrices $A \in \mathbb{R}^{n,p}$. 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 $\in \mathbb{R}^{n,p}$, Σ is diagonal with σ_i in the ith position, and V is a square orthogonal matrix $\in \mathbb{R}^{p,p}$. We will assume $n \geq p$, though there is an obvious modification for $n < p$. There are various expanded forms, for instance U could be square, and $\Sigma \in \mathbb{R}^{n,p}$, 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 $AV^T\Sigma^{-1}$.

Theorem 3. *The Jacobian of the singular value decomposition $A = U\Sigma V^T$ is*

$$\prod(\sigma_i^2 - \sigma_j^2)\prod \sigma_i^{n-m}(\mathrm{d}\Sigma)(H^T\mathrm{d}U)^\wedge(V^T\mathrm{d}V)^\wedge,$$

where H is an n by n orthogonal matrix whose first m columns are identical to U.

Proof. We first point out that the annoyance of defining H rather than U directly is due to the fact that U is rectangular. For most purposes you may "think" of H as representing U. ☐

Taking differentials

$$\mathrm{d}A = U\mathrm{d}\Sigma V^T + \mathrm{d}U\Sigma V^T + U\Sigma \mathrm{d}V^T$$

so that

$$\mathrm{d}X \equiv H^T\mathrm{d}AV = I_{n,p}\mathrm{d}\Sigma + H^T\mathrm{d}U\Sigma - I_{n,p}\Sigma V^T\mathrm{d}V\,.$$

We proceed to take the exterior product of the mn elements of the matrix of differentials $\mathrm{d}X$. The diagonal of $\mathrm{d}X$ consists only of the $\mathrm{d}\sigma_i$, so the exterior product of the diagonal elements is $(\mathrm{d}\Sigma)$. If $i < j \leq p$ then

$$\mathrm{d}X_{ij} = -\sigma_j u_j^T\mathrm{d}u_i - \sigma_i v_j^T\mathrm{d}v_i,$$

while

$$\mathrm{d}X_{ji} = \sigma_i u_j^T\mathrm{d}u_i - \sigma_j v_j^T\mathrm{d}v_i\,.$$

The exterior product of these two terms is

$$\mathrm{d}X_{ij}\wedge\mathrm{d}X_{ji} = (\sigma_i^2 - \sigma_j^2)(u_j^T\mathrm{d}u_i)\wedge(v_j^T\mathrm{d}v_i)\,.$$

Therefore the product of the off diagonal terms in the upper square part of $\mathrm{d}X$ is

$$\prod(\sigma_i^2 - \sigma_j)^2(U^T\mathrm{d}U)(V^T\mathrm{d}V)\,.$$

For $i > p$, we have

$$\mathrm{d}X_{ij} = \sigma_j h_i^T\mathrm{d}u_j,$$

and thus each σ_j appears an additional $m - n$ times in the Jacobian. We conclude that the exterior product of the portion below the top square is

$$\prod \sigma_i^{n-m}(H^T\mathrm{d}V)^\wedge\,.$$

Combining all the exterior products completes the proof.

2.8 Scalar Functions of Matrices

Let $f(x)$ be a scalar function of x. We may define $f(S)$ as $f(S) = Qf(\Lambda)Q^T$, where

$$f(\Lambda) = \begin{pmatrix} f(\lambda_1) & & \\ & \ddots & \\ & & f(\lambda_n) \end{pmatrix}.$$

Many students believe that one needs a Taylor series to define scalar functions of matrices, such as the matrix exponential, but it is sufficient to "apply" f to the eigenvalues.

Case I: Symmetric Matrices

$$\begin{aligned}(\mathrm{d}S)^\wedge &= \prod_{i<j}|\lambda_i - \lambda_j|(\mathrm{d}\Lambda)^\wedge(Q^T\mathrm{d}Q)^\wedge \text{ and}\\ (\mathrm{d}f(S))^\wedge &= \prod_{i<j}|f(\lambda_i) - f(\lambda_j)|(\mathrm{d}f(\Lambda))^\wedge(Q^T\mathrm{d}Q)^\wedge\\ &= \prod_{i<j}|f(\lambda_i) - f(\lambda_j)|\prod f'(\lambda_i)(\mathrm{d}\Lambda)^\wedge(Q^T\mathrm{d}Q)^\wedge\,.\end{aligned}$$

If follows immediately that

$$(\mathrm{d}f(S))^\wedge = \frac{\prod_{i<j}|f(\lambda_i)-f(\lambda_j)|}{\prod_{i<j}|\lambda_i-\lambda_j|}\prod f'(\lambda_i)(\mathrm{d}S)^\wedge\,.$$

We have not seen this formula or its analogs in the literature. It is a nice little formula in that it includes the discrete or continuous divided difference of (unequal or equal) parts of eigenvalues.

Case II: General Matrices

$$(\mathrm{d}f(A))^\wedge = \frac{\prod_{i\neq j}|f(\lambda_i)-f(\lambda_j)|}{\prod_{i\neq j}|\lambda_i-\lambda_j|}\prod f'(\lambda_i)(\mathrm{d}A)^\wedge$$

References

[1] Robb J. Muirhead. *Aspects of Multivariate Statistical Theory.* John Wiley & Sons, New York, 1982.