18.338J/16.394J: The Mathematics of Infinite Random Matrices
Essentials of Finite Random Matrix Theory

Alan Edelman

Handout #6, Tuesday, September 28, 2004

This handout provides the essential elements needed to understand finite random matrix theory. A cursory observation should reveal that the tools for infinite random matrix theory are quite different from the tools for finite random matrix theory. Nonetheless, there are significantly more *published* applications that use finite random matrix theory as opposed to infinite random matrix theory. Our belief is that many of the results that have been historically derived using finite random matrix theory can be reformulated and answered using infinite random matrix theory. In this sense, it is worth recognizing that in many applications it is an integral of a function of the eigenvalues that is more important that the mere distribution of the eigenvalues. For finite random matrix theory, the tools that often come into play when setting up such integrals are the *Matrix Jacobians*, the *Joint Eigenvalue Densities* and the *Cauchy-Binet theorem*. We describe these in subsequent sections.

1 Matrix and Vector Differentiation

In this section, we concern ourselves with the differentiation of matrices. Differentiating matrix and vector functions is not significantly harder than differentiating scalar functions except that we need notation to keep track of the variables. We titled this section "matrix and vector" differentiation, but of course it is the function that we differentiate. The matrix or vector is just a notational package for the scalar functions involved. In the end, a derivative is nothing more than the "linearization" of all the involved functions. We find it useful to think of this linearization both symbolically (for manipulative purposes) as well as numerically (in the sense of small numerical perturbations). The differential notation idea captures these viewpoints very well.

We begin with the familiar product rule for scalars,

$$\mathrm{d}(uv) = u(\mathrm{d}v) + v(\mathrm{d}u),$$

from which we can derive that $\mathrm{d}(x^3) = 3x^2\mathrm{d}x$. We refer to $\mathrm{d}x$ as a differential.

We all unconsciously interpret the "$\mathrm{d}x$" symbolically as well as numerically. Sometimes it is nice to confirm on a computer that

$$\frac{(x+\epsilon)^3 - x^3}{\epsilon} \approx 3x^2\,. \tag{1}$$

I do this by taking x to be 1 or 2 or `randn(1)` and ϵ to be .001 or .0001.

The product rule holds for matrices as well:

$$\mathrm{d}(UV) = U(\mathrm{d}V) + (\mathrm{d}U)V\,.$$

In the examples we will see some symbolic and numerical interpretations.

Example 1: $Y = X^3$

We use the product rule to differentiate $Y(X) = X^3$ to obtain that

$$\mathrm{d}(X^3) = X^2(\mathrm{d}X) + X(\mathrm{d}X)X + (\mathrm{d}X)X^2\,.$$

When I introduce undergraduate students to matrix multiplication, I tell them that matrices are like scalars, except that they do not commute.

The numerical (or first order perturbation theory) interpretation applies, but it may seem less familiar at first. Numerically take `X=randn(n)` and `E=randn(n)` for $\epsilon = .001$ say, and then compute

$$\frac{(X+\epsilon E)^3 - X^3}{\epsilon} \approx X^2E + XEX + EX^2\,. \tag{2}$$

This is the matrix version of (1). Holding X fixed and allowing E to vary, the right-hand side is a linear function of E. There is no simpler form possible.

Symbolically (or numerically) one can take $\mathrm{d}X = E_{kl}$ which is the matrix that has a one in element (k,l) and 0 elsewhere. Then we can write down the matrix of partial derivatives:

$$\frac{\partial X^3}{\partial x_{kl}} = X^2(E_{kl}) + X(E_{kl})X + (E_{kl})X^2\,.$$

As we let k,l vary over all possible indices, we obtain all the information we need to compute the derivative in any general direction E.

In general, the directional derivative of $Y_{ij}(X)$ in the direction $\mathrm{d}X$ is given by $(\mathrm{d}Y)_{ij}$. For a particular matrix X, $\mathrm{d}Y(X)$ is a matrix of directional derivatives corresponding to a first order perturbation in the direction $E = \mathrm{d}X$. It is a matrix of linear functions corresponding to the linearization of $Y(X)$ about X.

Structured Perturbations

We sometimes restrict our E to be a structured perturbation. For example if X is triangular, symmetric, antisymmetric, or even sparse then often we wish to restrict E so that the pattern is maintained in the perturbed matrix as well. An important case occurs when X is orthogonal. We will see in an example below that we will want to restrict E so that X^TE is antisymmetric when X is orthogonal.

Example 2: $y = x^Tx$

Here y is a scalar and dot products commute so that $\mathrm{d}y = 2x^T\mathrm{d}x$. When $y = 1$, x is on the unit sphere. To stay on the sphere, we need $\mathrm{d}y = 0$ so that $x^T\mathrm{d}x = 0$, i.e., the tangent to the sphere is perpendicular to the sphere. Note the two uses of $\mathrm{d}y$. In the first case it is the change to the squared length of x. In the second case, by setting $\mathrm{d}y = 0$, we find perturbations to x which to first order do not change the length at all. Indeed if one computes $(x+\mathrm{d}x)^T(x+\mathrm{d}x)$ for a small finite $\mathrm{d}x$, one sees that if $x^T\mathrm{d}x = 0$ then the length changes only to second order. Geometrically, one can draw a tangent to a circle. The distance to the circle is second order in the distance along the tangent.

Example 3: $y = x^TAx$

Again y is scalar. We have $\mathrm{d}y = \mathrm{d}x^TAx + x^TA\mathrm{d}x$. If A is symmetric then $\mathrm{d}y = 2x^TA\mathrm{d}x$.

Example 4: $Y = X^{-1}$

We have that $XY = I$ so that $X(\mathrm{d}Y) + (\mathrm{d}X)Y = 0$ so that $\mathrm{d}Y = -X^{-1}\mathrm{d}XX^{-1}$.

We recommend that the reader compute $\epsilon^{-1}((X+\epsilon E)^{-1} - X^{-1})$ numerically and verify that it is equal to $-X^{-1}EX^{-1}$.

In other words,

$$(X+\epsilon E)^{-1} = X^{-1} - \epsilon X^{-1}EX^{-1} + O(\epsilon^2)\,.$$

Example 5: $I = Q^TQ$

If Q is orthogonal we have that $Q^T\mathrm{d}Q + \mathrm{d}Q^TQ = 0$ so that $Q^T\mathrm{d}Q$ is antisymmetric.

In general, $\mathrm{d}(Q^TQ) = Q^T\mathrm{d}Q + \mathrm{d}Q^TQ$, but with no orthogonality condition on Q, there is no anti-symmetry condition on $Q^T\mathrm{d}Q$.

If y is a scalar function of $x_1, x_2, \ldots, x_n$ then we have the "chain rule"

$$\mathrm{d}y = \frac{\partial y}{\partial x_1}\mathrm{d}x_1 + \frac{\partial y}{\partial x_2}\mathrm{d}x_2 + \ldots + \frac{\partial y}{\partial x_n}\mathrm{d}x_n\,.$$

Often we wish to apply the chain rule to every element of a vector or matrix.

Let $X = \begin{bmatrix} p & q \\ r & s \end{bmatrix}$ and $Y = X^2 = \begin{bmatrix} p^2+qr & pq+rs \\ pr+rs & qs+s^2 \end{bmatrix}$ then

$$\mathrm{d}Y = X\mathrm{d}X + \mathrm{d}XX\,. \tag{3}$$

2 Matrix Jacobians (getting started)

2.1 Definition

Let $x \in \mathbb{R}^n$ and $y = y(x) \in \mathbb{R}^n$ be a differentiable function of x. It is well known that the Jacobian matrix

$$J = \begin{pmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial y_n}{\partial x_1} & \cdots & \frac{\partial y_n}{\partial x_n} \end{pmatrix} = \left(\frac{\partial y_i}{\partial x_j} \right)_{i,j=1,2,\ldots,n}$$

evaluated at a point x approximates $y(x)$ by a linear function. Intuitively $y(x+\delta x) \approx y(x) + J\delta x$, i.e., J is the matrix that allows us to invoke perturbation theory. The function y may be viewed as performing a change of variables.

Furthermore (intuitively) if a little box of n dimensional volume ϵ surrounds x, then it is transformed by y into a parallelopiped of volume $|\det J|\epsilon$ around $y(x)$. Therefore *the Jacobian* $|\det J|$ is the magnification factor for volumes.

If we are integrating some function of $y \in \mathbb{R}^n$ as in $\int p(y)\mathrm{d}y$, (where $\mathrm{d}y = \mathrm{d}y_1 \ldots \mathrm{d}y_n$), then when we change variables from y to x where $y = y(x)$, then the integral becomes $\int p(y(x))\,|\det(\frac{\partial y_i}{\partial x_j})|\,\mathrm{d}x$. For many people this becomes a matter of notation, but one should understand intuitively that the Jacobian tells you how volume elements scale.

The determinant is 0 exactly where the change of variables breaks down. It is always instructive to see when this happens. Either there is no "x" locally for each "y" or there are many as in the example of polar coordinates at the origin.

Notation: throughout this book, J denotes the Jacobian matrix. (Sometimes called the derivative or simply the Jacobian in the literature.) We will consistently write $\det J$ for the Jacobian determinant (unfortunately also called the Jacobian in the literature.) When we say Jacobian, we will be talking about both.

2.2 Simple Examples (n=2)

We get our feet wet with some simple 2×2 examples. Every reader is familiar with changing scalar variables as in

$$\int f(x)\,\mathrm{d}x = \int f(y^2)(2y)\,\mathrm{d}y\,.$$

We want the reader to be just as comfortable when f is a scalar function of a matrix and we change $X = Y^2$:

$$\int f(X)(\mathrm{d}X) = \int f(Y^2)(\text{Jacobian})(\mathrm{d}Y)\,.$$

One can compute all of the 2 by 2 Jacobians that follow by hand, but in some cases it can be tedious and hard to get right on the first try. Code 8.1 in MATLAB takes away the drudgery and gives the right answer. Later we will learn fancy ways to get the answer without too much drudgery and also without the aid of a computer. We now consider the following examples:

2×2 Example 1: Matrix Square ($Y = X^2$)

2 × 2 Example 2: Matrix Cube ($Y = X^3$)
2 × 2 Example 3: Matrix Inverse ($Y = X^{-1}$)
2 × 2 Example 4: Linear Transformation ($Y = AX + B$)
2 × 2 Example 5: The LU Decomposition ($X = LU$)
2 × 2 Example 6: A Symmetric Decomposition ($S = DMD$)
2 × 2 Example 7: Traceless Symmetric = Polar Decomposition ($S = Q\Lambda Q^T$, $\operatorname{tr} S = 0$)
2 × 2 Example 8: The Symmetric Eigenvalue Problem ($S = Q\Lambda Q^T$)
2 × 2 Example 9: Symmetric Congruence ($Y = A^T S A$)

Discussion:

Example 1: Matrix Square ($Y = X^2$)

With $X = \begin{bmatrix} p & q \\ r & s \end{bmatrix}$ and $Y = X^2$ the Jacobian matrix of interest is

$$J = \begin{array}{c} \\ \\ \end{array}\begin{array}{cccc} \partial p & \partial r & \partial q & \partial s \end{array}$$

$$J = \begin{bmatrix} 2p & q & r & 0 \\ r & p+s & 0 & r \\ q & 0 & p+s & q \\ 0 & q & r & 2s \end{bmatrix} \begin{array}{c} \partial Y_{11} \\ \partial Y_{21} \\ \partial Y_{12} \\ \partial Y_{22} \end{array}$$

On this first example we label the columns and rows so that the elements correspond to the definition $J = \left(\frac{\partial Y_{ij}}{\partial X_{kl}}\right)$. Later we will omit the labels. We invite readers to compare with Equation (3). We see that the Jacobian matrix and the differential contain the same information. We can compute then

$$\det J = 4(p+s)^2(sp - qr) = 4(\operatorname{tr} X)^2 \det(X)\,.$$

Notice that breakdown occurs if X is singular or has trace zero.

Example 2: Matrix Cube ($Y = X^3$)

With $X = \begin{bmatrix} p & q \\ r & s \end{bmatrix}$ and $Y = X^3$

$$J = \begin{bmatrix} 3\,p^2 + 2\,qr & pq + q\,(p+s) & 2\,rp + sr & qr \\ 2\,rp + sr & p^2 + 2\,qr + (p+s)\,s & r^2 & rp + 2\,sr \\ 2\,pq + qs & q^2 & p\,(p+s) + 2\,qr + s^2 & pq + 2\,qs \\ qr & pq + 2\,qs & r\,(p+s) + sr & 2\,qr + 3\,s^2 \end{bmatrix},$$

so that

$$\det J = 9(sp - qr)^2(qr + p^2 + s^2 + sp)^2 = \frac{9}{4}(\det X)^2(\operatorname{tr} X^2 + (\operatorname{tr} X)^2)^2\,.$$

Breakdown occurs if X is singular or if the eigenvalue ratio is a complex cube root of unity.

Example 3: Matrix Inverse ($Y = X^{-1}$)

With $X = \begin{bmatrix} p & q \\ r & s \end{bmatrix}$ and $Y = X^{-1}$

$$J = \frac{1}{\det X^2} \times \begin{bmatrix} -s^2 & qs & sr & -qr \\ sr & -ps & -r^2 & rp \\ qs & -q^2 & -ps & pq \\ -qr & pq & rp & -p^2 \end{bmatrix},$$

so that

$$\det J = (\det X)^{-4}.$$

Breakdown occurs if X is singular.

Example 4: Linear Transformation ($Y = AX + B$)

$$J = \begin{bmatrix} a & b & 0 & 0 \\ c & d & 0 & 0 \\ 0 & 0 & a & b \\ 0 & 0 & c & d \end{bmatrix}$$

The Jacobian matrix has two copies of the constant matrix A so that $\det J = \det A^2 = (\det A)^2$. Breakdown occurs if A is singular.

Example 5: The LU Decomposition ($X = LU$)

The LU Decomposition computes a lower triangular L with ones on the diagonal and an upper triangular U such that $X = LU$.

For a general 2×2 matrix it takes the form

$$\begin{bmatrix} p & q \\ r & s \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \frac{r}{p} & 1 \end{bmatrix} \begin{bmatrix} p & q \\ 0 & \frac{ps-qr}{p} \end{bmatrix}$$

which exists when $p \neq 0$.

Notice the function of four variables might be written:

$$\begin{aligned} y_1 &= p \\ y_2 &= r/p \\ y_3 &= q \\ y_4 &= (ps - qr)/p = \det(X)/p \end{aligned}$$

The Jacobian matrix is itself lower triangular

$$J = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -\frac{r}{p^2} & p^{-1} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \frac{qr}{p^2} & -\frac{q}{p} & -\frac{r}{p} & 1 \end{bmatrix},$$

so that $\det J = 1/p$. Breakdown occurs when $p = 0$.

Example 6: A Symmetric Decomposition ($S = DMD$)

Any symmetric matrix $X = \begin{bmatrix} p & r \\ r & s \end{bmatrix}$ may be written as

$$X = DMD \quad \text{where} \quad D = \begin{bmatrix} \sqrt{p} & 0 \\ 0 & \sqrt{s} \end{bmatrix} \quad \text{and} \quad M = \begin{bmatrix} 1 & r/\sqrt{ps} \\ r/\sqrt{ps} & 1 \end{bmatrix}.$$

Since X is symmetric, there are three independent variables, and thus the Jacobian matrix is 3×3. The three independent elements in D and M may be thought of as functions of p, r, and s: namely

$$\begin{aligned} y_1 &= \sqrt{p} \\ y_2 &= \sqrt{s} \quad \text{and} \\ y_3 &= r/\sqrt{ps} \end{aligned}$$

The Jacobian matrix is

$$J = \frac{1}{2}\begin{bmatrix} \frac{1}{\sqrt{p}} & 0 & 0 \\ 0 & \frac{1}{\sqrt{s}} & 0 \\ -\frac{r/p}{\sqrt{ps}} & -\frac{r/s}{\sqrt{ps}} & \frac{2}{\sqrt{ps}} \end{bmatrix}$$

so that

$$\det J = \frac{1}{4ps}$$

Breakdown occurs if p or s is 0.

Example 7: Traceless Symmetric = Polar Decomposition ($S = Q\Lambda Q^T$, $\operatorname{tr} S = 0$)

The reader will recall the usual definition of polar coordinates. If (p, s) are Cartesian coordinates, then the angle is $\theta = \arctan(s/p)$ and the radius is $r = \sqrt{p^2 + s^2}$. If we take a symmetric traceless 2×2 matrix

$$S = \begin{bmatrix} p & s \\ s & -p \end{bmatrix},$$

and compute its eigenvalue and eigenvector decomposition, we find that the eigendecomposition is mathematically equivalent to the familiar transformation between Cartesian and polar coordinates. Indeed one of the eigenvectors of S is $(\cos\phi, \sin\phi)$, where $\phi = \theta/2$. The Jacobian matrix is

$$J = \begin{bmatrix} \frac{p}{\sqrt{p^2+s^2}} & \frac{s}{\sqrt{p^2+s^2}} \\ \frac{-s}{p^2+s^2} & \frac{p}{p^2+s^2} \end{bmatrix}$$

The Jacobian is the inverse of the radius. This corresponds to the familiar formula using the more usual notation $\mathrm{d}x\mathrm{d}y/r = \mathrm{d}r\mathrm{d}\theta$ so that $\det J = 1/r$. Breakdown occurs when $r = 0$.

Example 8: The Symmetric Eigenvalue Problem ($S = Q\Lambda Q^T$)

We compute the Jacobian for the general symmetric eigenvalue and eigenvector decomposition. Let

$$S = \begin{bmatrix} p & s \\ s & r \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \lambda_1 & \\ & \lambda_2 \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}^T .$$

We can compute the eigenvectors and eigenvalues of S directly in MATLAB and compute the Jacobian of the two eigenvalues and the eigenvector angle, but when we tried with the Maple toolbox we found that the symbolic toolbox did not handle "arctan" very well. Instead we found it easy to compute the Jacobian in the other direction.

We write $S = Q\Lambda Q^T$, where Q is 2×2 orthogonal and Λ is diagonal. The Jacobian is

$$J = \begin{bmatrix} -2\sin\theta\cos\theta\,(\lambda_1 - \lambda_2) & \cos^2\theta & \sin^2\theta \\ 2\sin\theta\cos\theta\,(\lambda_1 - \lambda_2) & \sin^2\theta & \cos^2\theta \\ (\cos^2\theta - \sin^2\theta)(\lambda_1 - \lambda_2) & \sin\theta\cos\theta & -\sin\theta\cos\theta \end{bmatrix}$$

so that

$$\det J = \lambda_1 - \lambda_2 .$$

Breakdown occurs if S is a multiple of the identity.

Example 9: Symmetric Congruence ($Y = A^T S A$)

Let $Y = A^T S A$, where Y and S are symmetric, but $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is arbitrary. The Jacobian matrix is

$$J = \begin{bmatrix} a^2 & c^2 & 2\,ca \\ b^2 & d^2 & 2\,db \\ ab & cd & cb + ad \end{bmatrix}$$

and $\det J = (\det A)^3$.

The cube on the determinant tends to surprise many people. Can you guess what it is for an $n \times n$ symmetric matrix ($Y = A^T S A$)? The answer ($\det J = (\det A)^{n+1}$)is in Example 3 of Section 3.

```
%jacobian2by2.m
%Code 8.1 of Random Eigenvalues by Alan Edelman

%Experiment:    Compute the Jacobian of a 2x2 matrix function
%Comment:       Symbolic tools are not perfect.  The author
%              exercised care in choosing  the variables.

syms p q r s a b c d t e1 e2
X=[p q ; r s]; A=[a b;c d];

%% Compute Jacobians

Y=X^2;                J=jacobian(Y(:),X(:)), JAC_square   =factor(det(J))
Y=X^3;                J=jacobian(Y(:),X(:)), JAC_cube     =factor(det(J))
Y=inv(X);             J=jacobian(Y(:),X(:)), JAC_inv      =factor(det(J))
Y=A*X;                J=jacobian(Y(:),X(:)), JAC_linear   =factor(det(J))
Y=[p q;r/p det(X)/p]; J=jacobian(Y(:),X(:)), JAC_lu       =factor(det(J))

x=[p s r];y=[sqrt(p) sqrt(s) r/(sqrt(p)*sqrt(s))];
                      J=jacobian(y,x),       JAC_DMD      =factor(det(J))

x=[p s]; y=[ sqrt(p^2+s^2) atan(s/p)];
                      J=jacobian(y,x),       JAC_notrace  =factor(det(J))

Q=[cos(t) -sin(t); sin(t) cos(t)];
D=[e1 0;0 e2];Y=Q*D*Q.';
y=[Y(1,1) Y(2,2) Y(1,2)]; x=[t e1 e2];
                      J=jacobian(y,x),       JAC_symeig   =simplify(det(J))
X=[p s;s r]; Y=A.'*X*A;
y=[Y(1,1) Y(2,2) Y(1,2)]; x=[p r s];
                      J=jacobian(y,x),       JAC_symcong  =factor(det(J))
```

Code 1

3 $Y = BXA^T$ and the Kronecker Product

3.1 Jacobian of $Y = BXA^T$ (Kronecker Product Approach)

There is a "nuts and bolts" approach to calculate some Jacobian determinants. A good example function is the matrix inverse $Y = X^{-1}$. We recall from Example 4 of Section 1 that

$$\mathrm{d}Y = -X^{-1}\mathrm{d}XX^{-1}.$$

In words, the perturbation $\mathrm{d}X$ is multiplied on the left and on the right by a fixed matrix. When this happens we are in a "Kronecker Product" situation, and can instantly write down the Jacobian.

We provide two definitions of the Kronecker Product for square matrices $A \in \mathbb{R}^{n,n}$ to $B \in \mathbb{R}^{m,m}$.

See [1] for a nice discussion of Kronecker products.

Operator Definition

$A \otimes B$ is the operator from $X \in \mathbb{R}^{m,n}$ to $Y \in \mathbb{R}^{m,n}$ where $Y = BXA^T$. We write

$$(A \otimes B)X = BXA^T.$$

Matrix Definition (Tensor Product)

$A \otimes B$ is the matrix

$$A \otimes B = \begin{pmatrix} a_{11}B & \ldots & a_{1m_2}B \\ \vdots & & \vdots \\ a_{m_1 1}B & \ldots & a_{m_1 m_2}B \end{pmatrix}. \tag{4}$$

The following theorem is important for applications.

Theorem 1. $\det(A \otimes B) = (\det A)^m (\det B)^n$

Application: If $Y = X^{-1}$ then $\mathrm{d}Y = -(X^{-T} \otimes X^{-1})\,\mathrm{d}X$ so that

$$|\det J| = |\det X|^{-2n}.$$

Notational Note: The correspondence between the operator definition and matrix definition is worth spelling out. It corresponds to the following identity in MATLAB

```
Y = B * X * A'
Y(:) = kron(A,B) * X(:)
% The second line does not change Y
```

Here `kron(A,B)` is exactly the matrix in Equation (4), and `X(:)` is the column vector consisting of the columns of X stacked on top of each other. (In computer science this is known as storing an array in "column major" order.) Many authors write `vec(X)`, where we use `X(:)`. Concretely, we have that

$$\mathrm{vec}(BXA^T) = (A \otimes B)\mathrm{vec}(X)$$

where $A \otimes B$ is as in (4).

Proofs of Theorem 8.1:

Assume A and B are diagonalizable, with $Au_i = \lambda_i u_i$ $(i = 1, \ldots, n)$ and $Bv_i = \mu_i v_i$ $(i = 1, \ldots, m)$. Let $M_{ij} = v_i u_j^T$. The mn matrices M_{ij} form a basis for $\mathbb{R}^{mn}$ and they are eigenmatrices of our map since $BM_{ij}A^T = \mu_i \lambda_j M_{ij}$. The determinant is

$$\prod_{\substack{1 \le i \le n \\ 1 \le j \le m}} \mu_i \lambda_j \qquad \text{or} \qquad (\det A)^m (\det B)^n. \tag{5}$$

The assumption of diagonalizability is not important.

Also one can directly manipulate the matrix since obviously

$$A \otimes B = (A \otimes I)(I \otimes B)$$

as operators, and $\det I \otimes B = (\det B)^n$ and $\det A \otimes I$ which can be reshuffled into $\det I \otimes A = (\det A)^m$.

Other proofs may be obtained by working with the "LU" decomposition of A and B, the SVD of A and B, the QR decomposition, and many others.

Mathematical Note: That the operator $Y = BXA^T$ can be expressed as a matrix is a consequence of linearity:

$$\begin{array}{rrcl} & B(c_1X_1 + c_2X_2)A^T & = & c_1BX_1A^T + c_2BX_2A^T \\ \text{i.e.} & (A \otimes B)(c_1X_1 + c_2X_2) & = & c_1(A \otimes B)X_1 + c_2(A \otimes B)X_2 \end{array}$$

It is important to realize that a linear transformation from $\mathbb{R}^{m,n}$ to $\mathbb{R}^{m,n}$ is defined by an element of $\mathbb{R}^{mn,mn}$, i.e., by the m^2n^2 entries of an $mn \times mn$ matrix. The transformation defined by Kronecker products is an $m^2 + n^2$ subspace of this m^2n^2 dimensional space.

Some Kronecker product properties:

1. $(A \otimes B)^T = A^T \otimes B^T$
2. $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$
3. $\det(A \otimes B) = \det(A)^m \det(B)^n$, $A \in \mathbb{R}^{n,n}, B \in \mathbb{R}^{m,m}$
4. $\operatorname{tr}(A \otimes B) = \operatorname{tr}(A)\operatorname{tr}(B)$
5. $A \otimes B$ is orthogonal if A and B are orthogonal
6. $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$
7. If $Au = \lambda u$, and $Bv = \mu v$, then if $X = vu^T$, then $BXA^T = \lambda\mu X$, and also $AX^TB^T = \lambda\mu X^T$. Therefore $A \otimes B$ and $B \otimes A$ have the same eigenvalues, and transposed eigenvectors.

It is easy to see that property 5 holds, since if A and B are orthogonal $(A \otimes B)^T = A^T \otimes B^T = A^{-1} \otimes B^{-1} = (A \otimes B)^{-1}$.

Linear Subspace Kronecker Products

Some researchers consider the symmetric Kronecker product $\otimes_{\text{sym}}$. In fact it has become clear that one might consider an anti-symmetric, upper-triangular or even a Toeplitz Kronecker product. We formulate a general notion:

Definition: Let $\mathcal{S}$ denote a linear subspace of $\mathbb{R}^{mn}$ and $\pi_{\mathcal{S}}$ a projection onto $\mathcal{S}$. If $A \in \mathbb{R}^{n,n}$ and $B \in \mathbb{R}^{m,m}$ then we define $(A \otimes B)_{\mathcal{S}}X = \pi_{\mathcal{S}}(BXA^T)$ for $X \in \mathcal{S}$.

Comments

1. If $\mathcal{S}$ is the set of symmetric matrices then $m = n$ and

$$(A \otimes B)_{\text{sym}}X = \frac{BXA^T + AXB^T}{2}$$

2. If $\mathcal{S}$ is the set of anti-sym matrices, then $m = n$

$$(A \otimes B)_{\text{anti}}X = \frac{BXA^T + AXB^T}{2} \qquad \text{as well}$$

but this matrix is anti-sym.

3. If $\mathcal{S}$ is the set of upper triangular matrices, then $m = n$ and $(A \otimes B)_{\text{upper}} = \text{upper}(BXA^T)$.
4. We recall that $\pi_{\mathcal{S}}$ is a projection if $\pi_{\mathcal{S}}$ is linear and for all $X \in \mathbb{R}^{mn}$, $\pi_{\mathcal{S}} X \in \mathcal{S}$.
5. Jacobian of $\otimes_{\text{sym}}$:

$$(A \otimes B)_{\text{sym}} X = \frac{BXA^T + AXB^T}{2}$$

   If $B = A$ then

$$\det J = \prod_{i \le j} \lambda_i \lambda_j = (\det A)^{n+1}$$

   by considering eigenmatrices $E = u_i u_j^T$ for $i \le j$.

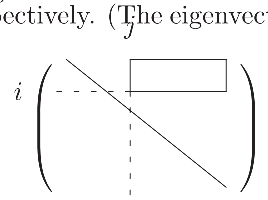
6. Jacobian of $\otimes_{\text{upper}}$:

   Special case: A lower triangular, B upper triangular so that

$$(A \otimes B)_{\text{upper}} X = BXA^T$$

   since BXA^T is upper triangular.

   The eigenvalues of A and B are $\lambda_i = A_{ii}$ and $\mu_j = B_{jj}$ respectively, where $Au_i = \lambda_i u_i$ and $Bv_i = \mu_i v_i$ $(i = 1, \ldots, n)$. The matrix $M_{ij} = v_i u_j^T$ for $i \le j$ is upper triangular since v_i and u_j are zero below the ith and above the jth component respectively. (The eigenvectors of a triangular matrix are triangular.)

   $M_{ij} =$  for $i \le j$.

   Since the M_{ij} are a basis for upper triangular matrices $BM_{ij}A^T = \mu_i \lambda_j M_{ij}$. We then have

$$\begin{aligned} \det J = \prod_{i \le j} \mu_i \lambda_j &= (\lambda_1 \lambda_2^2 \lambda_3^3 \ldots \lambda_n^n)(\mu_1^n \mu_2^{n-1} \mu_3^{n-2} \ldots \mu_n) \\ &= (A_{11} A_{22}^2 A_{33}^3 \ldots A_{nn}^n)(B_{11}^n B_{22}^{n-1} B_{33}^{n-2} \ldots B_{nn}) \,. \end{aligned}$$

   Note that J is singular if an only if A or B is.

7. Jacobian of $\otimes_{\text{Toeplitz}}$

   Let X be a Toeplitz matrix. We can define

$$(A \otimes_{\text{Toeplitz}} B) X = \text{Toeplitz}(BXA^T)$$

   where Toeplitz averages every diagonal.

4 Jacobians of Linear Functions, Powers and Inverses

The Jacobian of a linear map is just the determinant. This determinant is not always easily computed. The dimension of the underlying space of matrices plays a role. For example the Jacobian of $Y = 2X$ is 2^{n^2} for $X \in \mathbb{R}^{n \times n}$, $2^{\frac{n(n+1)}{2}}$ for upper triangular or symmetric X, $2^{\frac{n(n-1)}{2}}$ for antisymmetric X, and 2^n for diagonal X.

We will concentrate on general real matrices X and explore the symmetric case and triangular case as well when appropriate.

4.0.1 General Matrices

Let $Y = X^2$ then

$$\mathrm{d}Y = X\mathrm{d}X + \mathrm{d}X X \qquad \text{or} \qquad \mathrm{d}Y = I \otimes X + X^T \otimes I$$

Using the E_{ij} as before Equation (5), we see that the eigenvalues of $I \otimes X + X^T \otimes I$ are $\lambda_i + \lambda_j$ so that $\det J = \prod_{1 \le i,j \le n}(\lambda_i + \lambda_j) = 2^n \det X \prod_{i<j}(\lambda_i + \lambda_j)^2$. When $n = 2$, we obtain $4 \det X (\mathrm{tr} X)^2$ as seen in example 8.2.2.

For general powers, $Y = X^p$, $(p = 1, 2, 3, \ldots)$

$$\begin{aligned} \mathrm{d}Y &= \sum_{k=0}^{p-1} X^k \mathrm{d}X\, X^{p-1-k} \\ \text{and } \det J &= \prod_{1 \le i,j \le n} \left(\sum_{k=0}^{p-1} \lambda_i^k \lambda_j^{p-1-k} \right) = p^n (\det X)^{p-1} \prod_{i<j} \left[\frac{\lambda_i^p - \lambda_j^p}{\lambda_i - \lambda_j} \right]^2 . \end{aligned}$$

Of course by computing the Jacobian determinant explicitly in terms of the matrix elements, we can obtain an equivalent expression in terms of the elements of X rather than the λ_i.

Formally, if we plug in $p = -1$, we find that we can compute $\det J$ for $Y = X^{-1}$ yielding the answer given after Theorem 8.1. Indeed one can take any scalar function $y = f(x)$ such as $x^p, e^x, \log x, \sin x$, etc. The correct formula is

$$\det J = \det(f'(X)) \cdot \prod_{i<j} \left[\frac{f(\lambda_i) - f(\lambda_j)}{\lambda_i - \lambda_j} \right]^2 = \prod_{i=1}^{n} f'(\lambda_i) \prod_{i<j} \left[\frac{f(\lambda_i) - f(\lambda_j)}{\lambda_i - \lambda_j} \right]^2 .$$

We can prove these using the tools provided by wedge products.

Often one thinks that to define $Y = f(X)$ for matrices, one needs a power series, but this is not so. If $X = Z\Lambda Z^{-1}$, then we can define $f(X) = Zf(\Lambda)Z^{-1}$, where $f(\Lambda) = \mathrm{diag}(f(\lambda_1), \ldots, f(\lambda_n))$. The formula above is correct at any matrix X such that f is differentiable at the eigenvalues.

When X is of low rank, we can consider the Moore-Penrose Inverse of X. The Jacobian was calculated in [2] and [3]. There is a nonsymmetric and symmetric version.

Exercise. *Compute the Jacobian determinant for*

$$Y = X^{1/2} \quad \textit{and} \quad Y = X^{-1/2} .$$

Advanced exercise: *compute the Jacobian matrix for $Y = X^{1/2}$ as an inverse of a sum of two Kronecker products. Numerically, a "Sylvester equation" solver is needed. To obtain the Jacobian matrix for $Y = X^{-1/2}$ one can use the chain rule, given the answer to the previous problem. It is interesting that this is as hard as it is.*

4.0.2 Symmetric Matrices

If X is symmetric, we have that $\det J = (\det X)^{-(n+1)}$ for the map $Y = X^{-1}$.

5 Jacobians of Matrix Factorizations (without wedge products)

Let A be an $n \times n$ matrix. In elementary linear algebra, we learn about Gaussian elimination, Gram-Schmidt orthogonalization, and the eigendecomposition. All of these ideas may be written compactly as matrix factorizations, which in turn may be thought of as a change of variables:

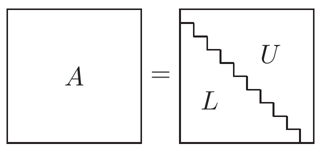
		parameter count
Gaussian Elimination:	$A = \underset{\substack{\uparrow \\ \text{unit lower} \\ \text{triangular}}}{L} \cdot \underset{\substack{\uparrow \\ \text{upper} \\ \text{triangular}}}{U}$	$n(n-1)/2 + n(n+1)/2$
Gram-Schmidt:	$A = \underset{\substack{\uparrow \\ \text{Orthogonal}}}{Q} \cdot \underset{\substack{\uparrow \\ \text{upper} \\ \text{triangular}}}{R}$	$n(n-1)/2 + n(n+1)/2$
Eigenvalue Decomposition:	$A = \underset{\substack{\uparrow \\ \text{eigenvectors}}}{X} \cdot \underset{\substack{\uparrow \\ \text{eigenvalues}}}{\Lambda} \cdot X^{-1}$	$\underset{\substack{\uparrow \\ \text{eigenvector}}}{(n^2-n)} + \underset{\substack{\uparrow \\ \text{eigenvalue}}}{n}$

Each of these factorizations is a change of variables. Somehow the n^2 parameters in A are transformed into n^2 other parameters, though it may not be immediately obvious what these parameters are (the $n(n-1)/2$ parameters in Q for example).

Our goal is to derive the Jacobians for those matrix factorizations.

5.1 Jacobian of Gaussian Elimination ($A = LU$)

In numerical linear algebra texts it is often observed that the memory used to store A on a computer may be overwritten with the n^2 variables in L and U. Graphically the $n \times n$ matrix A



Indeed the same is true for other factorizations. This is not just a happy coincidence, but deeply related to the fact that n^2 parameters ought not need more storage.

Theorem 2. *If $A = LU$, the Jacobian of the change of variables is*

$$\det J = u_{11}^{n-1} u_{22}^{n-2} \ldots u_{n-1,n-1} = \prod_{i=1}^{n} u_{ii}^{n-i}$$

Proof 1: Let $A = LU$, then using

$$\begin{aligned} \mathrm{d}A &= L\,\mathrm{d}U + \mathrm{d}L\,U \\ &= L(\mathrm{d}U U^{-1} + L^{-1}\mathrm{d}L)U \\ &= (U^T \otimes L)\left((U^T \otimes_{\text{upper}} I)^{-1}\mathrm{d}U + (I \otimes_{\text{lower}} L)^{-1}\mathrm{d}L\right) \end{aligned}$$

The mapping $\mathrm{d}U \to \mathrm{d}U U^{-1}$ only affects the upper triangular part. Similarly $\mathrm{d}L \to L^{-1}\mathrm{d}L$ for the lower triangular. We might think of the map in block format:

$$\mathrm{d}A = U^T \otimes L \begin{pmatrix} (U^T \otimes_{\text{upper}} I) & \\ & I \otimes_{\text{lower}} L \end{pmatrix}^{-1} \begin{pmatrix} \mathrm{d}U \\ \mathrm{d}L \end{pmatrix}$$

Since the Jacobian determinants of $U^T \otimes L$ is $\prod u_{ii}^n$ and of $(U^T \otimes I)^{-1}$ is $\prod u_{ii}^{-i}$ (and that of $I \otimes L$ is 1), we have that $\det J = \prod u_{ii}^{n-i}$. $\square$

That is the fancy slick proof. Here is the direct proof. It is of value to see both.

Proof 2: Let $M_{ij} = \begin{cases} l_{ij} & i > j \\ u_{ij} & i \leq j \end{cases}$ as in the diagram above. Since $A = LU$, we have that

$$A_{ij} \quad = \quad \sum_{k=1}^{i-1} M_{ik} M_{kj} + u_{ij} \quad \text{for} \quad i \leq j$$

and

$$A_{ij} \quad = \quad \sum_{k=1}^{j-1} M_{ik} M_{kj} + l_{ij} u_{jj} \quad \text{for} \quad i > j\,.$$

Therefore

$$\frac{\partial A_{ij}}{\partial M_{ij}} = \begin{cases} 1 & \text{if} \quad i \leq j \\ u_{jj} & \text{if} \quad i > j \end{cases} \quad . \tag{6}$$

Notice that A_{ij} never depends on M_{pq} when $p > i$ or $q > j$. Therefore if we order the variables first by row and next by column, we see that the Jacobian matrix is lower triangular with diagonal entries given in (6). Remembering that the determinant of a lower triangular matrix is the product of the diagonal entries, the theorem follows. □

Perhaps it should not have been surprising that the condition that A admits a unique LU decomposition may be thought of in terms of the Jacobian. Given a matrix A, let p_k denote the determinant of the upper left k by k principal minor. It may be readily verified that $u_{11} \dots u_{kk} = p_k$ and hence the Jacobian is $p_1 p_2 \dots p_{n-1}$. The condition that A admits a unique LU decomposition is well known to be that all the upper-left principal minors of dimension smaller than n are non-singular. Otherwise the matrix may have no LU decomposition. For example,

$$\begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \quad \text{has no } LU \text{ factorization.}$$

It may also have many LU decompositions as does the zero matrix when $n > 1$. This degeneracy explains the need for pivoting strategies (i.e., strategies that reorder the matrix rows and/or columns) for solving linear systems of equations even if the computation is done in exact arithmetic. Modern Gaussian elimination software for solving linear systems of equations in finite precision include pivoting strategies designed to avoid being near a matrix with such a degeneracy.

Exercise. *If A is symmetric positive definite, it may be factored $A = LL^T$. This is the famous Cholesky decomposition of A. How many independent variables are in A? In L? Prove that the Jacobian of the change of variables is*

$$\det J = 2^n \prod_{i=1}^{n} l_{ii}^{n+1-i}$$

Research Question. *It seems that the existence of a finite algorithm for a matrix factorization is linked to a triangular Jacobian matrix. After all, the latter implies only one new variable need be substituted at a time. This is the essential idea of Doolittle's or Crout's matrix factorization schemes for LU.*

5.2 Jacobian of Gram-Schmidt ($A = QR$)

It is well known that any nonsingular square matrix A may be factored uniquely into orthogonal times upper triangular, or $A = QR$ with $R_{ii} > 0$.

Theorem 3. *Given any $\mathrm{d}Q$ with $Q^T\mathrm{d}Q$ anti-symmetric, and any $\mathrm{d}R$, we can compute $\mathrm{d}A = Q\mathrm{d}R + \mathrm{d}QR = Q(\mathrm{d}RR^{-1} + Q^T\mathrm{d}Q)R = (R^T \otimes Q)\left((R^T \otimes_{\text{upper}} I)^{-1}\mathrm{d}R + (Q^T\mathrm{d}Q)\right)$. The linear maps from (lower $Q^T\mathrm{d}Q$), $\mathrm{d}R$ to $\mathrm{d}A$ then has determinant $(\prod_{i=1}^n r_{ii}^n)(\prod r_{ii}^{-i}) = \prod_{i=1}^n r_{ii}^{n-i}$.*

Note: It is straightforward to imagine computing the Jacobian numerically. We can make $\frac{n(n+1)}{2}$ perturbations to R, say by systematically adding .001 to each entry, and then compute $\mathrm{d}A$ via $\mathrm{d}A = QR - Q(R+\mathrm{d}R)$.

Similarly we can make $n(n-1)/2$ perturbation to Q by choosing two distinct columns of Q and multiplying the $n \times 2$ matrix formed from these columns by a matrix such as

$$\begin{bmatrix} \cos(.001) & \sin(.001) \\ -\sin(.001) & \cos(.001) \end{bmatrix}.$$

The $\mathrm{d}Q$ is the new Q minus the old one.

The n^2 resulting $\mathrm{d}A$'s may be squashed into an n^2 by n^2 matrix whose determinant could be computed numerically, confirming Theorem 3.

Note: It is easy to compute the inverse map, i.e., given $\mathrm{d}A$ compute $\mathrm{d}Q$ and $\mathrm{d}R$. Numerically we can ask for $[Q, R] = qr(A)$ and $[Q + \mathrm{d}Q, R + \mathrm{d}R] = qr(A + \mathrm{d}A)$.

Mathematically, $\mathrm{d}A = Q(\mathrm{d}RR^{-1} + Q^T\mathrm{d}Q)R$ so if $\mathrm{lower}(M)$ denotes the strictly lower part of M and $\mathrm{upper}(M)$ denotes the upper part of M, then

$$\begin{aligned} \mathrm{d}Q &= Q\left(\mathrm{lower}(Q^T\mathrm{d}AR^{-1}) - \mathrm{lower}(Q^T\mathrm{d}AR^{-1})^T\right) \\ \text{and} \quad \mathrm{d}R &= \left(\mathrm{upper}(Q^T\mathrm{d}AR^{-1}) + \mathrm{lower}(Q^T\mathrm{d}A\mathrm{d}R^{-1})^T\right) R \end{aligned}$$

6 Jacobians for Spherical Coordinates

The Jacobian matrix for the transformation between spherical coordinates and Cartesian coordinates has a sufficiently interesting structure that we include this case as our final example. The structure is known as a "lower Hessenberg" structure.

We recall that in $\mathbb{R}^n$, we define spherical coordinates $r, \theta_1, \theta_2, \ldots, \theta_{n-1}$ by

$$\begin{aligned} x_1 &= r\cos\theta_1 \\ x_2 &= r\sin\theta_1\cos\theta_2 \\ x_3 &= r\sin\theta_1\sin\theta_2\cos\theta_3 \\ &\vdots \\ x_{n-1} &= r\sin\theta_1\sin\theta_2\cdots\sin\theta_{n-2}\cos\theta_{n-1} \\ x_n &= r\sin\theta_1\sin\theta_2\ldots\sin\theta_{n-2}\sin\theta_{n-1} \end{aligned}$$

or for $j = 1, \ldots, n$, x_j may be written as

$$x_j = r\left[\prod_{i=1}^{j-1}\sin\theta_i\right]\cos\theta_j \qquad (\theta_n = 0).$$

We schematically place an $\times$ in the matrix below if x_j depends on the variable heading the columns. The dependency structure is then

$$\begin{array}{c} \\ x_1 \\ x_2 \\ \vdots \\ x_{n-2} \\ x_{n-1} \\ x_n \end{array} \begin{array}{c} \begin{array}{cccccc} r & \theta_1 & \theta_2 & \cdots & \theta_{n-2} & \theta_{n-1} \end{array} \\ \begin{pmatrix} \times & \times & & & & \\ \times & \times & \times & & & \\ \vdots & & & \ddots & & \\ \times & \times & \times & & \times & \\ \times & \times & \times & & \times & \times \\ \times & \times & \times & \cdots & \times & \times \end{pmatrix} \end{array}$$

Therefore the nonzero structure of the Jacobian matrix is represented by the pattern above.

Matrices that are nonzero only in the lower triangular part and on the superdiagonal are referred to as lower Hessenberg matrices.

It is fairly messy to write down the exact Jacobian matrix. Fortunately, it is unnecessary to do so. We obtain the LU factorization of the matrix by defining auxiliary variables y_i as follows:

$$\begin{array}{lclcl} y_1 & = & x_1^2 + \cdots\cdots\cdots + x_n^2 & = & r^2 \\ y_2 & = & x_2^2 + \cdots\cdots + x_n^2 & = & r^2 \sin^2\theta_1 \\ y_3 & = & x_3^2 + \cdots + x_n^2 & = & r^2 \sin^2\theta_1 \sin^2\theta_2 \\ \vdots & \vdots & \ddots \quad \vdots & \vdots & \vdots \\ y_n & = & x_n^2 & = & r^2 \sin^2\theta_1 \sin^2\theta_2 \cdots \sin^2\theta_{n-1} \end{array}$$

Differentiating and expressing the relationship between $\mathrm{d}y_i + \mathrm{d}x_j$ for $i, j = 1, \ldots, n$ in matrix form we obtain that the Jacobian matrix $J_{x\to y}$ from x to y is triangular:

$$2\begin{pmatrix} x_1 & x_2 & \cdots & x_n \\ & x_2 & \cdots & x_n \\ & & \ddots & \\ & & & x_n \end{pmatrix} \begin{pmatrix} \mathrm{d}x_1 \\ \mathrm{d}x_2 \\ \vdots \\ \mathrm{d}x_n \end{pmatrix} = \begin{pmatrix} \mathrm{d}y_1 \\ \mathrm{d}y_2 \\ \vdots \\ \mathrm{d}y_n \end{pmatrix}.$$

We recognize that we have an upper triangular matrix in front of the vector of Cartesian differentials and a lower triangular matrix in front of the vector of spherical coordinate differentials.

$$2\begin{pmatrix} 1 & 1 & \cdots & 1 \\ & 1 & \cdots & 1 \\ & & \ddots & \\ & & & 1 \end{pmatrix} \begin{pmatrix} x_1 & & & \\ & x_2 & & \\ & & \ddots & \\ & & & x_n \end{pmatrix} \begin{pmatrix} \mathrm{d}x_1 \\ \mathrm{d}x_2 \\ \vdots \\ \mathrm{d}x_n \end{pmatrix} = \begin{pmatrix} \mathrm{d}y_1 \\ \mathrm{d}y_2 \\ \vdots \\ \mathrm{d}y_n \end{pmatrix}$$

Similarly the Jacobian matrix $J_{\text{spherical}\to y}$ from spherical coordinates to y is triangular

$$\begin{pmatrix} 2r & & & & \\ \vdots & 2r\sin\theta_1 x_1 & & & \\ \vdots & \vdots & 2r\sin\theta_1\sin\theta_2 x_2 & & \\ \vdots & \vdots & \vdots & \ddots & \\ \vdots & \vdots & \vdots & & 2r\sin\theta_1\cdots\sin\theta_{n-1}x_{n-1} \end{pmatrix} \begin{pmatrix} \mathrm{d}r \\ \mathrm{d}\theta_1 \\ \mathrm{d}\theta_2 \\ \vdots \\ \mathrm{d}\theta_{n-1} \end{pmatrix} = \begin{pmatrix} \mathrm{d}y_1 \\ \mathrm{d}y_2 \\ \vdots \\ \mathrm{d}y_n \end{pmatrix}.$$

Therefore $J_{x\to r,\theta} = J^{-1}_{r,\theta\to y} J_{x\to y}$.

The LU (really $(L^{-1}U)^{-1}$) allows us to easily obtain the determinant as the ratio of the triangular determinants. The Jacobian determinant is then

$$\begin{aligned} \frac{\partial(x_1, \ldots, x_n)}{\partial(r, \theta_1, \ldots, \theta_n)} &= \frac{2^n r^n (\sin\theta_1)^{n-1}(\sin\theta_2)^{n-2}(\sin\theta_{n-1})x_1\cdots x_{n-1}}{2^n\, x_1\cdots x_n} \\ &= r^{n-1}(\sin\theta_1)^{n-2}\cdots(\sin\theta_{n-2}). \end{aligned}$$

In the next section we will introduce wedge products and show that they are a convenient tool for handling curvilinear matrix coordinates.

7 Wedge Products

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 Section 5 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 chapter 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.

8 The Mechanics of Wedging

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

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

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

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

Let us write down concretely the wedge product.

Rewriting the two examples above 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., "($\quad$)$^\wedge$" denotes wedge together all the elements of the vector inside the parentheses.

A notational note: We are inventing the "($\quad$)$^\wedge$" notation. Books such as [4] 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 "($\quad$)$^\wedge$" 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.

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

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 $(\mathrm{d}M_1)^\wedge \wedge (\mathrm{d}M_2)^\wedge$.

9 Jacobians with wedge products

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

Consider the 2×2 case of $Y = X^2$ as in Example 1 of Section 2.2.

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

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

Thus

$$\begin{array}{rcl} \mathrm{d}Y_{11} & = & 2p\mathrm{d}p + q\mathrm{d}r + r\mathrm{d}q \\ \mathrm{d}Y_{21} & = & r\mathrm{d}p + (p+s)\mathrm{d}r + r\mathrm{d}s \\ \mathrm{d}Y_{12} & = & q\mathrm{d}p + (p+s)\mathrm{d}q + q\mathrm{d}s \\ \mathrm{d}Y_{22} & = & q\mathrm{d}r + r\mathrm{d}q + 2s\mathrm{d}s\,. \end{array} \tag{7}$$

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:

$$J = \begin{array}{cc} \begin{array}{cccc} \partial p & \partial r & \partial q & \partial s \end{array} & \\ \begin{bmatrix} 2p & q & r & 0 \\ r & p+s & 0 & r \\ q & 0 & p+s & q \\ 0 & q & r & 2s \end{bmatrix} & \begin{array}{c} \partial Y_{11} \\ \partial Y_{21} \\ \partial Y_{12} \\ \partial Y_{22} \end{array} \end{array}$$

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

9.2 Square Matrix = Antisymmetric + Upper Triangular

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

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

where $\mathrm{lower}(M)$ is the strictly lower triangular part of M. Further let

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

where $\mathrm{upper}(M)$ includes the diagonal. We have the decomposition

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

Therefore

$$\mathrm{d}M = \begin{pmatrix} \mathrm{d}r_{11} & \mathrm{d}r_{12} - \mathrm{d}a_{21} & \mathrm{d}r_{13} - \mathrm{d}a_{31} & \cdots & \mathrm{d}r_{1n} - \mathrm{d}a_{n1} \\ \mathrm{d}a_{21} & \mathrm{d}r_{22} & \mathrm{d}r_{23} - \mathrm{d}a_{32} & & \mathrm{d}r_{2n} - \mathrm{d}a_{n2} \\ \mathrm{d}a_{31} & \mathrm{d}a_{32} & \mathrm{d}r_{33} & & \mathrm{d}r_{3n} - \mathrm{d}a_{n3} \\ \vdots & \vdots & \vdots & & \vdots \\ \mathrm{d}a_{n1} & \mathrm{d}a_{n2} & \mathrm{d}a_{n3} & & \mathrm{d}r_{nn} \end{pmatrix}. \tag{8}$$

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 $\mathrm{lower}(A)$ and would have written a block two by two matrix as in Section 5.2.

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

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

9.3 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.

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

9.4 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 $\left(\frac{\partial y_{ij}}{\partial x_{kl}}\right)$.

9.5 Eigenproblem Jacobians

Case I: S Symmetric

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

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

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 Section 4 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 not discuss this in length here but encourage you to ask us if you are curious.

10 Handbook of Jacobians

Some Jacobians that come up frequently are listed in this and the following sections.

Function	Sizes	Jacobian
$Y = BXA^T = (A \bigotimes B)X$ "Kronecker Product"	X $m \times n$, A $n \times n$, B $m \times m$	$(\mathrm{d}Y) = (\det A)^n (\det B)^m$
$Y = X^{-1}$ "inverse"	X $n \times n$	$(\mathrm{d}Y) = (\det X)^{-2n} (\mathrm{d}X)$
$Y = X^2$	X $n \times n$	$(\mathrm{d}Y) = \prod_{i,j} \lvert\lambda_i + \lambda_j\rvert (\mathrm{d}X)$
$Y = X^k$	X $n \times n$	$(\mathrm{d}Y) = \prod_{i \geq j} \left\lvert \sum_{l=0}^{k-1} \lambda_i^l \lambda_j^{n-1-l} \right\rvert (\mathrm{d}X)$
$Y = \frac{1}{2}(AXB + B^T XA)$ "Symmetric Kronecker"	X, A, B $n \times n$; A, B sym; $B = A^T$	$(\mathrm{d}Y) = \prod_{i \leq j} \lvert\lambda_i M_j + \lambda_j M_i\rvert (\mathrm{d}X)$; $(\mathrm{d}Y) = (\det A)^{n+1} (\mathrm{d}X)$

11 Real Case; General Matrices

In factorizations that involve a square $m \times m$ orthogonal matrix Q, and start from a thin, tall $m \times n$ matrix A ($m \geq n$ *always*), $H^T \mathrm{d}Q$ stands for the wedge product

$$H^T \mathrm{d}Q = \bigwedge_{i=1}^{n} \bigwedge_{j=i+1}^{n} q_i^T \mathrm{d}q_j \,.$$

Here H is a stand-in for the first n columns of Q. If the differential matrix and the matrix whose transpose we multiply by are one and the same, the wedging is done over all of the columns. For example,

$$V^T \mathrm{d}V = \bigwedge_{i=1}^{n} \bigwedge_{j=i+1}^{n} v_i^T \mathrm{d}v_j,$$

for V an $n \times n$ orthogonal matrix.

The Jacobian of the last factorization (non-symmetric eigenvalue) is computed only for the positive-measure set of matrices with real eigenvalues and with a full set of eigenvectors.

Factorization	Matrix sizes & properties	Parameter count	Jacobian
$A = LU$ lu	A, L, U $n \times n$; L, U^T lower triangular; $l_{ii} = 1$, $\forall i = 1, n$	$n^2 = \frac{n(n-1)}{2} + \frac{n(n+1)}{2}$	$(dA) = \prod_{i=1}^{n} u_{ii}^{n-i} (dL)(dU)$
$A = QR$ qr	A, R $m \times n$; Q $m \times m$; $Q^T Q = I_n$; R upper triangular	$mn = \left(mn - \frac{n(n+1)}{2}\right) + \frac{n(n+1)}{2}$	$(dA) = \prod_{i=1}^{n} r_{ii}^{m-i} (dR)(H^T dQ)$
$A = U\Sigma V^T$ svd	A, Σ $m \times n$; U $m \times m$, V $n \times n$; $U^T U = I_m$; $V^T V = I_n$	$mn = \left(mn - \frac{n(n+1)}{2}\right) + n + \frac{n(n-1)}{2}$	$(dA) = \prod_{i<j} (\sigma_i^2 - \sigma_j^2) \prod_{i=1}^{n} \sigma_i^{m-n} (d\Sigma)(H^T dU)(V^T dV)$
$A = QS$ polar	A, Q $m \times n$; S $n \times n$; $Q^T Q = I_n$; S positive definite	$mn = \left(mn - \frac{n(n+1)}{2}\right) + \frac{n(n+1)}{2}$	$(dA) = \prod_{i<j} (\sigma_i + \sigma_j)(dS)(Q^T dQ)$
$A = X\Lambda X^{-1}$ nonsymm eig	A, X, Λ $n \times n$; Λ diagonal	$n^2 = (n^2 - n) + n$	$(dA) = \prod_{i<j} (\lambda_i - \lambda_j)^2 (d\Lambda)(X^{-1} dX)$

12 Real Matrices; Symmetric "+" Cases

Here we gather factorizations that are symmetric or, in the case when this is required, positive definite. All matrices are square $m \times m$.

Factorization	"+"	Parameter count	Matrix properties	Jacobian
$A = LL'$ Choleski	Positive definite	$\frac{n(n+1)}{2} = \frac{n(n+1)}{2}$	L lower triangular	$(dA) = 2^n \prod_{i=1}^{n} l_{ii}^{n+1-i}(dL)$
$A = LDL'$ ldl	—	$\frac{n(n+1)}{2} = \frac{n(n-1)}{2} + n$	L lower triangular $l_{ii} = 1, \forall i = 1, n$ D diagonal	$(dA) = \prod_{i=1}^{n} d_i^{n-i}(dL)(dD)$
$A = Q\Lambda Q^T$ eig	—	$\frac{n(n+1)}{2} = \frac{n(n-1)}{2} + n$	$Q^TQ = I_n$ Λ diagonal	$(dA) = \prod_{i<j} \lvert\lambda_i - \lambda_j\rvert(d\Lambda)(Q^TdQ)$

13 Real Matrices; Orthogonal Case

This is the case of the CS decomposition. Note that the matrices U_1 and V_1 share $\frac{p(p-1)}{2}$ parameters; this is due to the fact that the product of the first p columns of U_1 and the first p rows of V_1^T is invariant and determined. This is equivalent to saying that we introduce an equivalence relation of the set of pairs (U_1, V_1) of orthogonal matrices of size k, by having

$$(U_1, V_1) \sim \left(U_1 \begin{bmatrix} Q & 0 \\ 0 & I_j \end{bmatrix}, V_1 \begin{bmatrix} Q & 0 \\ 0 & I_j \end{bmatrix} \right) ,$$

for any $p \times p$ orthogonal matrix Q.

Since it is rather hard to integrate over a manifold of pairs of orthogonal matrices with the equivalence relationship mentioned above, we will use a different way of thinking about this splitting of parameters. We can assume that U_1 contains $\frac{k(k-1)}{2}$ parameters (a full set) while V_1 only contains $\frac{k(k-1)}{2} - \frac{p(p-1)}{2}$ (having the first p columns already determined from the relationship with U_1). Alternatively, we could assign a full set of parameters to V_1 and think of U_1 as having the first r columns predetermined. In the following we choose to make the former assumption, and think of the undetermined part H of V_1^T as a point in the Stiefel manifold $V_{j,k}$.

Factorization	Matrix sizes & properties	Parameter count	Jacobian
$Q = \begin{bmatrix} U_1 & 0 \\ 0 & U_2 \end{bmatrix} \begin{bmatrix} I_p & 0 & 0 \\ 0 & C & S \\ 0 & S & -C \end{bmatrix} \begin{bmatrix} V_1 & 0 \\ 0 & V_2 \end{bmatrix}^T$ $n = k + j,\ p = k - j$ CS decomposition	Q $n \times n$ U_1, V_1 $k \times k$ U_2, V_2 $j \times j$ $Q^TQ = I_n$ U_1, V_1, U_2, V_2 orthogonal $C = \mathrm{diag}(\cos(\theta_i)),\ i = 1 \ldots n$ $S = \mathrm{diag}(\sin(\theta_i)),\ i = 1 \ldots n$	$U_1V_1^T$ has $k(k-1) - \frac{p(p-1)}{2}$ U_2, V_2 have $\frac{j(j-1)}{2}$ each C and S share j	$(dA) = \prod_{i<j} sin(\theta_j - \theta_i)sin(\theta_i + \theta_j)(d\theta)$ $(U_1^TdU_1)(U_2^TdU_2)(H^TdH)(V_2^TdV_2)$

14 Real Matrices; Symmetric Tridiagonal

The important factorization of this section is the eigenvalue decomposition of the symmetric tridiagonal.The tridiagonal is defined by its $2n - 1$ parameters, n on the diagonal and $n - 1$ on the upper diagonal:

$$T = \begin{pmatrix} a_n & b_{n-1} & & & \\ b_{n-1} & a_{n-1} & b_{n-2} & & \\ & \ddots & \ddots & \ddots & \\ & & b_2 & a_2 & b_1 \\ & & & b_1 & a_1 \end{pmatrix} .$$

The matrix of eigenvectors Q can be completely determined from knowledge of its first row $q = (q_1, \ldots, q_n)$ ($n-1$ parameters) and the eigenvalues on the diagonal of Λ (another n parameters). Here dq represents integration over the $(1, n)$ Stiefel manifold.

Factorization	Matrix sizes	Parameter count	Matrix properties	Jacobian
$T = Q\Lambda Q^T$ eig	$T, Q, \Lambda \;\; n \times n$	$2n-1 =$ $(n-1)+n$	$Q^TQ = I_n$ Λ diagonal	$(dT) = \dfrac{\prod_{i=1}^{n-1} b_i}{\prod_{i=1}^{n} q_i}(d\Lambda)(dq)$

Proof by MATLAB

At a basic level, as we discussed earlier, the Jacobian is the magnification in the volume element that occurs when a matrix is perturbed in "every conceivable direction". Hence the theoretical results can be verfied using actual MATLAB experiments. Consider the symmetric eigenvalue problem on the third row of the table in Section 12. The function `jsymeig.m` allows you to enter a real symmetric matrix A and computes the theoretical Jacobian and the numerical answer. You should easily be able to verify that both these answers match. We encourage you to verify the theoretical answers for other factorizations.

Once we have the Jacobians computing the eigenvalue densities is fairly straightforward. The densities for the Hermite, Laguerre and Jacobi ensembles is tabulated here as well.

```
function  j  =  jsymeig(a)
%JSYMEIG   Jacobian  for  the  symmetric  eigenvalue  problem
%        J  =  JSYMEIG(A)  returns  the  Numerical  and  Theoretical  Jacobian  for
%           the  symmmetric  eigenvalue  problem. The  numerical  answer  is
%           computed  by  perturbing  the  symmetric  matrix  A  in  the  N*(N+1)/2
%           independent  directions  and  computing  the  change  in  the
%           eigenvalues  and  eigenvectos.
%
%      A  is  a  REAL  N  x  N  symmetric  matrix
%      Example:  Valid  Inputs  for  A  are
%             1)  G  =  randn(5);  A  =  (G  +  G')
%             2)  G  =  randn(5,10);  A  =  G  *  G'
%
%      J  is  a  1  x  2  row  vector
%      J(1)  is  the  theoretical  Jacobian  computed  using  the  formula  in  the
%          third  row  of  the  table  in  Section  12  of  Handout  #  6
%
%
%      References:
%      [1]  Alan  Edelman,  Handout  6:  Essentials  of  Finite  Random  Matrix
%                       Theory,  Fall  2004,  Course  Notes  18.338.
%      [2]  Alan  Edelman,  Random  Matrix  Eigenvalues.
%
%      Alan  Edelman  and  Raj  Rao,  Sept.  2004.
%      $Revision:  1.1  $   $Date:  2004/09/28   17:11:18  $

format  long
[q,e]  =  eig(a);   %  Compute  eigenvalues  and  eigenvectors
e  =  diag(e);
epsilon  =  1e-7;   %  Size  of  perturbation
n  =  length(a);    %  Size  of  Matrix
jacmatrix  =  zeros(n*(n+1)/2);
k  =  0;  mask  =  triu(  ones(n),1);  mask  =  logical(mask(:));
for  i  =  1:n,
    for  j  =  i:n
        k  =  k+1;
        E  =  zeros(n);
        E(i,j)  =  1;  E(j,i)  =  1;
        aa  =  a  +  epsilon  *  E;
        [qq,ee]  =  eig(aa);
        de=  (diag(ee)-e)/epsilon;
        qdq  =  q'*(qq-q)/epsilon;  qdq  =  qdq(:);
        jacmatrix(1:n,k)  =  de;
        jacmatrix((n+1):end,k)  =  qdq(mask);
    end
end

%  Numerical  answer
j  =  1/det(jacmatrix);

%  Theoretical  answer
[e1,e2]  =  meshgrid(e,e);
z  =  abs(e1-e2);
j  =  abs([j  prod(  z(mask)  )  ]);
```

Code 2

Random Matrix Ensembles

Joint Eigenvalue Distributions: $c\,|\Delta|^{\beta} \prod_{i=1}^{n} V(\lambda_i)$

Technical name	Traditional name	β	Field	Property	Invariance	Connected Matrix Problem
Hermite ensembles	**Gaussian ensembles**					(EIG)
$V(\lambda) = e^{-\lambda^2/2}$	GOE	1	$\mathbb{R}$	Symmetric	$A \to Q^T A Q$	Symmetric Eigenvalue Problem
	GUE	2	$\mathbb{C}$	Hermitian	$A \to U^H A U$	
	GSE	4	$\mathbb{H}$	Self−Dual	$A \to S^D A S$	
Laguerre ensembles	**Wishart ensembles**					(SVD)
$V(\lambda) = \lambda^a e^{-\lambda/2}$	Wishart real	1	$\mathbb{R}$	Positive Semi−Definite	$A \to Q^T A Q$	Singular Value Decomposition
$a = \frac{\beta}{2}(n - m + 1) - 1$	Wishart complex	2	$\mathbb{C}$		$A \to U^H A U$	
	Wishart quaternion	4	$\mathbb{H}$		$A \to S^D A S$	
Jacobi ensembles	**MANOVA ensembles**					(QZ)
$V(\lambda) = \lambda^a (1-\lambda)^b$	MANOVA real	1	$\mathbb{R}$	Positive Semi−Definite	$X \to Q^T X Q_1$, $Y \to Q^T Y Q_1$	Generalized Symmetric Eigenvalue Problem
$a = \frac{\beta}{2}(n_1 - m + 1) - 1$	MANOVA complex	2	$\mathbb{C}$		$X \to U^H X U_1$, $Y \to U^H Y U_1$	
$b = \frac{\beta}{2}(n_2 - m + 1) - 1$	MANOVA quaternion	4	$\mathbb{H}$		$X \to S^D X S_1$, $Y \to S^D Y S_1$	

Type of ensemble	β	**MATLAB** code
Hermite	1	A = randn(n); A = (A+A') / 2;
	2	A = randn(n) + i * randn(n); A = (A+A') / 2;
	4	X = randn(n) + i * randn(n); Y = randn(n) + i * randn(n);
		A = [X Y; −conj(Y) conj(X)]; A = (A+A') / 2;
Laguerre	1	A = randn(m, n); A = A * A';
	2	A = randn(m, n) + i * randn(m, n); A = A * A';
	4	X = randn(m, n) + i * randn(m, n); Y = randn(m, n) + i * randn(m, n);
		A = [X Y; −conj(Y) conj(X)]; A = A * A';
Jacobi	1	X = randn(m, n_1); Y = randn(m, n_2); A = (X * X') / (X * X' + Y * Y');
	2	X = randn(m, n_1) + i * randn(m, n_1); Y = randn(m, n_2) + i * randn(m, n_2); A = (X * X') / (X * X' + Y * Y');
	4	X_1 = randn(m, n_1) + i * randn(m, n_1); X_2 = randn(m, n_1) + i * randn(m, n_1);
		Y_1 = randn(m, n_2) + i * randn(m, n_2); Y_2 = randn(m, n_2) + i * randn(m, n_2);
		X = [X_1 X_2; −conj(X_2) conj(X_1)]; Y = [Y_1 Y_2; −conj(Y_2) conj(Y_1)];
		A = (X * X') / (X * X' + Y * Y');

15 Volume and Integration

15.1 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 as we saw in 9.4 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 p orthogonal matrices Q ($Q^TQ = I_m$) which we shall introduce shortly. Exterior products will give us the correct volume element for integration.

If the x_i are Cartesian coordinates in n-dimensional Euclidean space, 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 $q^Tq = 1$). 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[1].

Integration:

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

Example. $\int_{\mathbb{R}^n} \exp(-||x||^2/2)(\mathrm{d}x) = (2\pi)^{n/2}$ and similarly $\int_{\mathbb{R}^{n,n}} \exp(-||x||_F^2/2)(\mathrm{d}A) = (2\pi)^{n^2/2}$. $||A||_F^2 = \mathrm{tr}(A^TA) = \sum_{i,j} a_{ij}^2 =$ "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., $\{x \in \mathbb{R}^n : ||x|| = 1\}$? $\bigwedge_{i=1} \mathrm{d}x_i = (\mathrm{d}x)^\wedge = 0$. 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.

15.2 Plucker Coordinates and Volume Measurement

Let $F \in \mathbb{R}^{n,p}$. We might think of the columns of F as the edges of a parallelopiped. By defining $\mathrm{Pl}(F)$ ("Plucker(F)"), we can obtain simple formulas for volumes.

Definition 1. $\mathrm{Pl}(F)$ *is the vector of* $p \times p$ *subdeterminants of* F *written in natural order.*

[1] 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 by taking the complex logarithm.

$$p=2: \qquad \begin{pmatrix} f_{11} & f_{12} \\ \vdots & \vdots \\ f_{n1} & f_{n2} \end{pmatrix} \xrightarrow{\text{Pl}} \begin{pmatrix} f_{11}f_{22}-f_{21}f_{12} \\ \vdots \\ f_{n-1,1}f_{n,2}-f_{n,1}f_{n-1,2} \end{pmatrix}$$

$$p=3: \qquad \begin{pmatrix} f_{11} & f_{12} & f_{13} \\ & \vdots & \\ f_{n1} & f_{n2} & f_{n3} \end{pmatrix} \xrightarrow{\text{Pl}} \begin{pmatrix} \begin{vmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{vmatrix} \\ \vdots \\ \begin{vmatrix} f_{n-2,1} & f_{n-2,2} & f_{n-2,3} \\ f_{n-1,1} & f_{n-1,2} & f_{n-1,3} \\ f_{n,1} & f_{n,2} & f_{n,3} \end{vmatrix} \end{pmatrix}$$

$$p \text{ general}: \qquad F=(f_{ij})_{\substack{1\le i\le n\\ 1\le j\le p}} \xrightarrow{\text{Pl}} \begin{pmatrix} \det (f_{ij})_{\substack{i=i_1,\ldots,i_p\\ j=1,\ldots,p}} \end{pmatrix}_{i_1<\ldots<i_p}$$

Definition 2. *Let* $\text{vol}(F)$ *denote the volume of the parallelopiped* $\{Fx : 0 \le x_i \le 1\}$, *i.e., the volume of the parallelopiped with edges equal to the columns of* F.

Theorem 4. $\text{vol}(F) = \prod_{i=1}^p \sigma_i = \det(F^TF)^{1/2} = \prod_{i=1}^p r_{ii} = \|\text{Pl}(F)\|$, *where the* σ_i *are the singular values of* F, *and the* r_{ii} *are the diagonal elements of* R *in* $F = YR$, *where* $Y \in \mathbb{R}^{n,p}$ *has orthonormal columns and* R *is upper triangular.*

Proof. Let $F = U\Sigma V^T$, where $U \in \mathbb{R}^{n,p}$ and $V \in \mathbb{R}^{p,p}$ has orthonormal columns. The matrix σ denotes a box with edge sides σ_i so has volume $\prod \sigma_i$. The volume is invariant under rotations. The other formulas follow trivially except perhaps the last which follows from the Cauchy-Binet theorem taking $A = F^T$ and $B = F$. $\square$

Theorem 5. (Cauchy-Binet) *Let* $C = AB$ *be a matrix product of any kind. Let* $M\left(\begin{smallmatrix} i_1 \ldots i_p \\ j_1 \ldots j_p \end{smallmatrix}\right)$ *denote the* $p \times p$ *minor*

$$\det(M_{i_k j_l})_{1\le k\le p, 1\le l\le p}.$$

In other words, it is the determinant of the submatrix of M *formed from rows* $i_1, \ldots, i_p$ *and columns* $j_1, \ldots, j_p$.

The Cauchy-Binet Theorem states that

$$C\begin{pmatrix} i_1,\ldots,i_p \\ k_1,\ldots,k_p \end{pmatrix} = \sum_{j_1<j_2<\cdots<j_p} A\begin{pmatrix} i_1,\ldots,i_p \\ j_1,\ldots,j_p \end{pmatrix} B\begin{pmatrix} j_1,\ldots,j_p \\ k_1,\ldots,k_p \end{pmatrix}.$$

Notice that when $p = 1$ *this is the familiar formula for matrix multiplication. When all matrices are* $p \times p$, *then the formula states that*

$$\det C = \det A \det B\,.$$

Corollary 6. *Let* $F \in \mathbb{R}^{n,p}$ *have orthonormal columns, i.e.,* $F^TF = I_p$. *Let* $X \in \mathbb{R}^{n,p}$. *If* $\text{span}(F) = \text{span}(X)$, *then* $\text{vol}(X) = \det(F^TX) = \text{Pl}(F)^T \cdot \text{Pl}(X)$.

Theorem 7. *Let* $F, V \in \mathbb{R}^{n,p}$ *be arbitrary. Then*

$$\text{Pl}(F)^T\text{Pl}(V) = \det(F^TX) = |\,\text{vol}(F)|\,|\,\text{vol}(V)| \prod_{i=1}^p \cos\theta_i\,,$$

where $\theta_1, \ldots, \theta_p$ *are the principal angles between span*(F) *and span*(V).

Remark 1. *If* S_p *is some* p*-dimensional surface it is convenient for* $F^{(i)}$ *to be a set of* p *orthonormal tangent vectors on the surface at some point* $x^{(i)}$ *and* $V^{(i)}$ *to be any "little" parallelopiped on the surface.*

If we decompose the surface into parallelopipeds we have

$$\mathrm{vol}(S_p) \approx \sum \mathrm{Pl}(F^{(i)})^T \mathrm{Pl}(V^{(i)})\,.$$

and

$$\int f(x)\mathrm{d}(\text{surface}) \approx \sum f(x^{(i)})\,\mathrm{Pl}(F^{(i)})^T \mathrm{Pl}(V^{(i)}) = \sum f(x^{(i)})\,\det((F^{(i)})^T V^{(i)})\,.$$

Mathematicians write the continuous limit of the above equation as

$$\int f(x)\mathrm{d}(\text{surface}) = \int f(x)(F^T\mathrm{d}x)^\wedge\,.$$

Notice that $(F(x)^T\mathrm{d}x)^\wedge$ formally computes $\mathrm{Pl}(F(x))$. Indeed

$$(F(x)^T\mathrm{d}x)^\wedge = \mathrm{Pl}(F(x))^T \begin{pmatrix} \vdots \\ \mathrm{d}x_{i_1} \wedge \ldots \wedge \mathrm{d}x_{i_p} \\ \vdots \end{pmatrix}_{i_1<\ldots<i_p}\,.$$

III. Generalized dot products algebraically: Linear functions $l(x)$ for $x \in \mathbb{R}^n$ may be written $l(x) = \sum f_i x_i$, i.e., $f^T x$.

The information that specifies the linear function is exactly the components of f.

Similarly, consider functions $l(V)$ for $V \in \mathbb{R}^{n,p}$ that have the form $l(V) = \det(F^T V)$, where $F \in \mathbb{R}^{n,p}$. The reader should check how much information is needed to "know" $l(V)$ uniquely. (It is "less than" all the elements of F.) In fact, if we define

$$E_{i_1 i_2 \ldots i_p} = \text{The } n \times p \text{ matrix consisting of columns } i_1, \ldots, i_p \text{ of the identity,}$$

i.e., $(E = \mathrm{eye}(n); E_{i,i_2,\ldots,i_p} = E(:,[i_1 \ldots i_p]))$ in pseudo-Matlab notation, then knowing $l(E_{i_1,\ldots,i_p})$ for all $i_1 < \ldots < i_p$ is equivalent to knowing l everywhere. This information is precisely all $p \times p$ subdeterminants of F, the Plucker coordinates.

IV. Generalized dot product geometrically. If F and $V \in \mathbb{R}^{n,p}$ we can generalize the familiar $f^T v = \|f\|\,\|v\| \cos\theta$.

It becomes

$$\mathrm{Pl}(F)^T \mathrm{Pl}(V) = \det(F^T V) = |\mathrm{vol}(F)|\,|\mathrm{vol}(V)| \prod_{i=1}^{p} \cos\theta_i\,.$$

Here $\mathrm{vol}(M)$ denotes the volume of the parallelopiped that is the image of the unit cube under M (Box with edges equal to columns of M). Everyone knows that there is an angle between any two vectors in $\mathbb{R}^n$. Less well known is that there are p principal angles between two p-dimensional hyperplanes in $\mathbb{R}^n$. The θ_i are principal angles between the two hyperplanes spanned by F and by V.

Easy special case: Suppose $F^T F = I_p$ and $\mathrm{span}(F) = \mathrm{span}(V)$. In other words, the columns of F are an orthonormal basis for the columns of V. In this case

$$\det(F^T V) = \mathrm{vol}(V)\,.$$

Other formulas for $\mathrm{vol}(V)$ which we mention for completeness is

$$\mathrm{vol}(V) = \prod_{i=1}^{p} \sigma_i(V)\,,$$

the product of the singular values of V. And

$$\mathrm{vol}(V) = \sqrt{\sum_{i_1<\ldots<i_p} V \begin{pmatrix} 1 \ldots p \\ i_1 \ldots i_p \end{pmatrix}^2}\,,$$

the sum of the $p \times p$ subdeterminants squared. In other words $\mathrm{vol}(V) = \|\mathrm{Plucker}(V)\|$.

1. Volume measuring function: (Volume element of Integration)

   We now more explicitly consider $V \in \mathbb{R}^{n,p}$ as a parallelopiped generated by its columns. For general $F \in \mathbb{R}^{n,p}$, $\det(F^T V)$ may be thought of as a "skewed" volume.

We now carefully interpret $\bigwedge_{i=1}^{p}(F^T \mathrm{d}x)_i$, where $F \in \mathbb{R}^{n,p}$ and $\mathrm{d}x = (\mathrm{d}x_1 \mathrm{d}x_2 \ldots \mathrm{d}x_n)^T$. We remind the reader that $(F^T \mathrm{d}x)_i = \sum_{j=1}^{n} F_{ji} \mathrm{d}x_j$.

By $\bigwedge_{i=1}^{p}(F^T \mathrm{d}x)_i$ we are thinking of the linear map that takes $V \in \mathbb{R}^{n,p}$ to the scalar $\det(F^T V)$. In our mind we imagine the following assumptions

- We have some p-dimensional surface $\mathcal{S}_p$ in $\mathbb{R}^n$.
- At some point P on $\mathcal{S}_p$, the columns of $X \in \mathbb{R}^{n,p}$ are tangent to $\mathcal{S}_p$. We think of X as generating a little parallelopiped on the surface of $\mathcal{S}_p$.
- We imagine the columns of F are an orthonormal basis for the tangents to $\mathcal{S}_p$ at P.
- Then $\bigwedge_{i=1}^{p}(F^T \mathrm{d}x)_i = (F^T \mathrm{d}x)^\wedge$ is a function that replaces parallelopipeds on $\mathcal{S}_p$ with its Euclidean volume.

Now we can do "the calculus thing." Suppose we want to compute

$$I = \int_{\mathcal{S}_p} f(x) \mathrm{d}(\text{surface}),$$

the integral of some scalar function f on the surface $\mathcal{S}_p$.

We discretize by circumscribing the surface with little parallelopipeds that we specify by using the matrix $V^{(i)} \in \mathbb{R}^{n,p}$. The word circumscribing indicates that the p columns of $V^{(i)}$ are tangent (or reasonably close to tangent) at $x^{(i)}$.

Let $F^{(i)}$ be an orthonormal basis for the tangents at $x^{(i)}$. Then I is discretized by

$$I \approx \sum_i f(x^{(i)}) \det({F^{(i)}}^T V^{(i)})$$

We then use the notation

$$I = \int_{\mathcal{S}_p} f(x) \bigwedge_{i=1}^{p}(F^T \mathrm{d}x)_i = \int_{\mathcal{S}_p} f(x)(F^T \mathrm{d}x)^\wedge$$

to indicate this continuous limit. Here f is a function of x.

The notation does not require that F be orthonormal or tangent vectors. These conditions guarantee that you get the correct Euclidean volume. Let them go and you obtain some warped or weighted volume. Linear combinations are allowed too.

The integral notation does require that you feed in tangent vectors if you discretize. Careful mathematics shows that for the cases we care about, no matter how one discretizes, the limit of small parallelopipeds gives a unique answer. (Analog of no matter how you take small rectangles to compute Riemann integrals, in the limit there is only one unique area under a curve.)

15.3 Overview of special surfaces

We are very interested in the following three mathematical objects:

1. The sphere $= \{x : \|x\| = 1\}$ in $\mathbb{R}^n$
2. The orthogonal group $O(n)$ of orthogonal matrices Q ($Q^T Q = I$) in $\mathbb{R}^{nn}$.
3. The Stiefel manifold of tall skinny matrices $Y \in \mathbb{R}^{np}$ with orthogonal columns ($Y^T Y = I_p$).

Of course the sphere and the orthogonal group are special cases of the Stiefel manifold with $p = 1$ and $p = n$ respectively.

We will derive and explain the meaning of the following volume forms on these objects

Sphere	$(H^T \mathrm{d}q)^\wedge_{i=2,\ldots,n}$	where H is a so-called Householder transform
Orthogonal Group	$(Q^T \mathrm{d}Q)^\wedge = (Q^T \mathrm{d}Q)^\wedge_{i>j}$	
Stiefel Manifold	$(Q^T \mathrm{d}Y)^\wedge_{i>j}$	where $Q^T Y = I_p$.

The concept of volume or surface area is so familiar that the reader might not see the need for a formalism to encode this concept.

Here are some examples.

Example 1: Integration about a circle Let $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$.

$$\int_{\substack{x \in \mathbb{R}^2 \\ \|x\|=1}} f(x)(-x_1 \mathrm{d}x_1 + x_2 \mathrm{d}x_2) = \int_0^{2\pi} f(\cos\theta, \sin\theta)\mathrm{d}\theta$$

Algebraic proof: $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}$ implies

$$-x_2 \mathrm{d}x_1 + x_1 \mathrm{d}x_2 = \mathrm{d}\theta = \begin{pmatrix} -x_2 \\ x_1 \end{pmatrix}^T \begin{pmatrix} \mathrm{d}x_1 \\ \mathrm{d}x_2 \end{pmatrix} = \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix}^T \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix} \mathrm{d}\theta\,.$$

Geometric Interpretation: Approximate the circle by a polygon with k sides.

$$\sum f(x^{(i)}) \begin{pmatrix} -x_2^{(i)} \\ x_1^{(i)} \end{pmatrix}^T (x^{(i)} - x^{(i-1)}) \approx \int_0^{2\pi} f(\cos\theta, \sin\theta)\mathrm{d}\theta\,.$$

Note that the dot product computes $\|v_i\|$.

Geometric interpretation of the wrong answer: Why not $\int_{\substack{x \in \mathbb{R}^2 \\ \|x\|=1}} f(x)\mathrm{d}x_1$?

This is approximated by

$$\sum f(x^{(i)}) \begin{pmatrix} 1 \\ 0 \end{pmatrix}^T v^{(i)}$$

which is the integral of f over the "shadow" of the circle on the x-axis.

Example 2: Integration over a sphere Given q with $\|q\| = 1$, let $H(q)$ be any $n \times n$ orthogonal matrix with first column q. (There are many ways to do this. One way is to construct the "Householder" reflector $H(q) = I - 2\frac{vv^T}{v^T v}$, where $v = e_1 - q$, and e_1 is the first column of I.)

The sphere is an $n - 1$ dimensional surface in n dimensional space.

Integration over the sphere is then

$$\int_{\substack{x \in \mathbb{R}^n \\ \|x\|=1}} f(x) \bigwedge_{i=2}^{n} (H^T \mathrm{d}x)_i = \int_{\substack{x \in \mathbb{R}^n \\ \|x\|=1}} f(x)\mathrm{d}S\,,$$

where $\mathrm{d}S$ is the surface "volume" on the sphere.

Consider the familiar sphere $n = 3$. Approximate the sphere by a polyhedron of parallelograms with k faces. A parallelogram may be algebraically encoded by the two vectors forming edges, into a matrix $V(i)$. We have $\sum f(x(v_i)) \cdot \det(H(q)^T V)$ approximates this integral.

Example 3: The orthogonal group Let $O(n)$ denote the set of orthogonal $n \times n$ matrices. This is an $\frac{n(n-1)}{2}$ dimensional set living in $\mathbb{R}^{n^2}$.

Tangents consist of matrices T such that Q^TT is anti-symmetric. We can take an orthogonal, but not orthonormal set to consist of matrices $Q(E_{ij} - E_{ji})$ for $i > j$ ($(E_{ij_{kl}} = 1)$ if $i = k$ and $j = l$; 0 otherwise). The matrix "F^TV" roughly amounts to taking twice the triangular part of $Q^T\mathrm{d}Q$.

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 $(Q^T\mathrm{d}Q)^\wedge = \mathrm{d}\theta$.

15.4 The Sphere

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, 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.

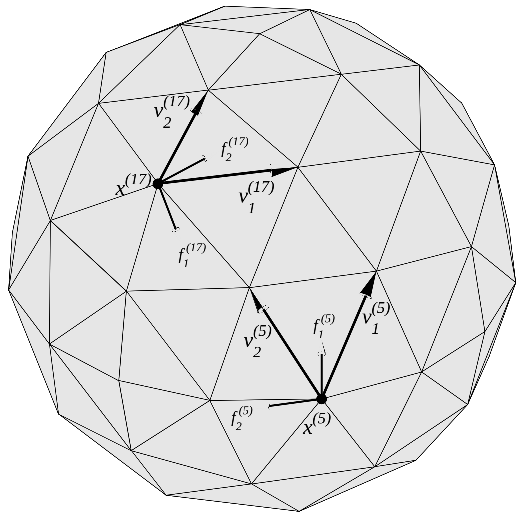
We can derive an expression for the surface element on a sphere. We introduce an orthogonal and symmetric matrix H such that $Hq = e_1$ and $He_1 = q$, where e_1 is the first column of the identity. Then

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

Thus

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

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 an H as described above so as to be the Householder reflector. Notice that H serves as a rotating coordinate system. It is critical that $H(:, 2:n)$ is an orthogonal basis of tangents. Choose $v = e_1 - q$ the external angle bisector and

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

See Figure 1 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.

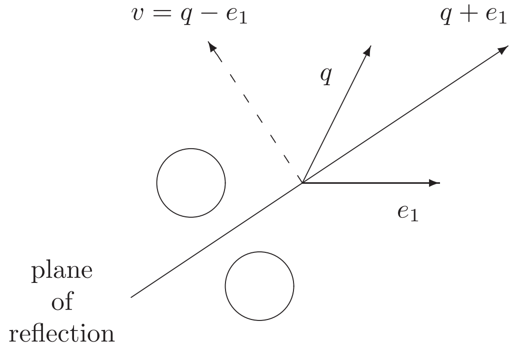


Figure 1:

I think many people do not really understand the meaning of $(Q^T\mathrm{d}Q)^\wedge$. I like to build a nice cube on the surface of the orthogonal group first. Then we will connect the notations. First of all $O(n)$ is an $\frac{n(n-1)}{2}$ dimensional surface in $\mathbb{R}^{n^2}$. At any point Q, tangents have the form $Q \cdot A$, where A is anti-symmetric.

The dot product of two "vectors" (matrices) X and Y is $X \cdot Y \equiv \mathrm{tr}(X^TY) = \sum X_{ij}Y_{ij}$. If Q is orthogonal $QX \cdot QY = X \cdot Y$.

If $Q = I$ then the matrices $A_{ij} = (E_{ij} - E_{ji})/\sqrt{2}$ for $i < j$ clearly form an $\frac{n(n-1)}{2}$ dimensional cube tangent to $O(n)$ at I, i.e., they form an orthonormal basis for the tangent space. Similarly the $n(n-1)/2$ matrices $Q(E_{ij} - E_{ji})/\sqrt{2}$ form such a basis at Q.

One can form an F whose columns are $\mathrm{vec}(Q(E_{ij} - E_{ji}))$ for $i < j$. The matrix would be n^2 by $n(n-1)/2$. Then $F^T(\mathrm{d}q_{ij})_{i<j}$ would be the Euclidean volume element. Mathematicians like to throw away the $\sqrt{2}$ so that the volume element is off by the factor $2^{n(n-1)/4}$. This does not seem to bother anyone. In non-vec format, this is $(Q^T\mathrm{d}Q)^\wedge$.

Application

Surface Area of Sphere Computation

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

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

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

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

15.5 The Stiefel Manifold

QR Decomposition

Let $A \in \mathbb{R}^{n,m}(m \le n)$. Taking $x = a_i$, we see $\exists H_1$ such that H_1A 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 $H_2H_1A = \begin{pmatrix} x & x \\ 0 & x \\ \vdots & \vdots \\ 0 & 0 \end{pmatrix}$. Continuing $H_m \cdots H_1A = \begin{pmatrix} 0 & & \\ & \ddots & \\ & & 0 \\ 0 & \cdots & 0 \end{pmatrix}$ 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 $Q \in \mathbb{R}^{n,p}$ such that $Q^TQ = I_p$ is denoted $V_{p,n}$ and is known as the Stiefel manifold. Considering the Householder construction, $\exists H_1, \cdots, H_p$ such that.

$$H_pH_{p-1}\cdots H_1Q = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}$$

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

Corollary 8. *The Stiefel manifold may be specified by $(n-1)+(n-2)+\cdots+(n-p) = pn - 1/2p(p+1)$ parameters. You may think of this as the pn arbitrary parameters in an $n \times p$ matrix reduced by the $p(p+1)/2$ conditions that $q_i^T q_j = \delta_{ij}$ for $i \geq j$. You might also think of this as*

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

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 p matrix A. If $Q \in \mathbb{R}^{p,n}$ is orthogonal, let H be an orthogonal p by p matrix such that $H^T Q$ is the first p columns of I. Actually H may be constructed by applying the Householder process to Q. Notice that Q is simply the first p columns of H.

As we proceed with the general case, notice how this generalizes the situation when $p = 1$. 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 p upper triangular matrix. While $H^T\mathrm{d}Q$ is (rectangularly) antisymmetric. ($h_i^T h_j = 0$ implies $h_i^T \mathrm{d}h_j = -h_j^T \mathrm{d}h_i$)

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 a 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

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

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.

16 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.

References

[1] Charles Van Loan. The ubiquitous Kronecker product. *J. Comp. and Applied Math.*, 123(200):85–100.

[2] Y. Zhang. The exact distribution of the Moore-Penrose inverse of X with a density. In P.R. Krishnaiah, editor, *Multivariate Analysis VI*, pages 633–635, New York, 1985. Elsevier.

[3] Heinz Neudecker and Shuangzhe Liu. The density of the Moore-Penrose inverse of a random matrix. *Multivariate Analysis*, 237/238:123–126, 1996.

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