18.325: Finite Random Matrix Theory
Jacobians of Matrix Transforms (without wedge products)

Professor Alan Edelman

Handout #2, Tuesday, February 1, 2005

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 $X^T E$ is antisymmetric when X is orthogonal.

Example 2: $y = x^T x$

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: $y = x^T A x$

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

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: $I = Q^T Q$

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

In general, $\mathrm{d}(Q^T Q) = Q^T\mathrm{d}Q + \mathrm{d}Q^T Q$, 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

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

1 Matrix Jacobians (getting started)

1.1 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.

1.2 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 ($Y = A^TSA$)

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

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

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 ($Y = A^TSA$)

Let $Y = A^TSA$, 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 ($Y = A^TSA$)? The answer ($\det J = (\det A)^{n+1}$)is in Example 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

2 $Y = BXA^T$ and the Kronecker Product

2.1 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 that

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

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

$$A \otimes B = \begin{pmatrix} a_{11}B & \dots & a_{1m_2}B \\ \vdots & & \vdots \\ a_{m_11}B & \dots & a_{m_1m_2}B \end{pmatrix} . \qquad (4)$$

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 $Au_i = \lambda_i u_i$ $(i = 1, \dots, n)$ and $Bv_i = \mu_i v_i$ $(i = 1, \dots, m)$. 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

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

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:

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

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. $\mathrm{tr}(A \otimes B) = \mathrm{tr}(A)\mathrm{tr}(B)$

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 $\otimes_{\mathrm{sym}}$. 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

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

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

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

   but this matrix is anti-sym.

3. If $\mathcal{S}$ is the set of upper triangular matrices, then $m = n$ and $(A \otimes B)_{\mathrm{upper}} = \mathrm{upper}(BXA^T)$.

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 $\otimes_{\mathrm{sym}}$:

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

   If $B = A$ then

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

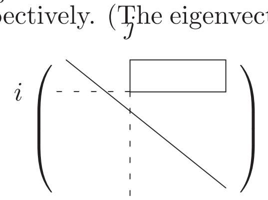
   by considering eigenmatrices $E = u_i u_j^T$ for $i \le j$.

6. Jacobian of $\otimes_{\mathrm{upper}}$:
   Special case: A lower triangular, B upper triangular so that

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

   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.

3 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.

3.0.1 General Matrices

Let $Y = X^2$ then

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

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

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

.

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.*

3.0.2 Symmetric Matrices

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

4 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:

Here is a table that will be expanded upon on later lectures.

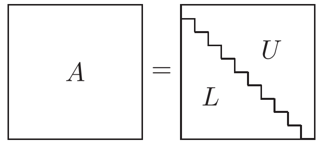
		parameter count
Gaussian Elimination:	$A = L \cdot U$ L ↑ unit lower triangular; U ↑ upper triangular	$n(n-1)/2 + n(n+1)/2$
Gram-Schmidt:	$A = Q \cdot R$ Q ↑ Orthogonal; R ↑ upper triangular	$n(n-1)/2 + n(n+1)/2$
Eigenvalue Decomposition:	$A = X \cdot \Lambda \cdot X^{-1}$ X ↑ eigenvectors; Λ ↑ eigenvalues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

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.

4.1 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*

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

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 $\det J = \prod u_{ii}^{n-i}$. □

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

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

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

and

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

Therefore

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

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 $p_1p_2\dots p_{n-1}$. 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.*

4.2 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* $\mathrm{d}A = Q\mathrm{d}R + \mathrm{d}QR = Q(\mathrm{d}RR^{-1} + Q^T\mathrm{dQ})R = (R^T \otimes Q)\left((R^T \otimes_{\text{upper}} I)^{-1}\mathrm{d}R + (Q^T\mathrm{dQ})\right)$. *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 lower(M) denotes the strictly lower part of M and upper(M) denotes the upper part of M, then

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

5 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:

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

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 chapter we will introduce wedge products and show that they are a convenient tool for handling curvilinear matrix coordinates.

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.