Chapter 8

Jacobians of Matrix Transforms (without wedge products)

Chapters 8 and 9 provide the background to derive the matrix Jacobians such as those found in Chapter 42.

8.1 Matrix and Vector Differentiation

In this section, we concern ourselves with the differentiation of matrices.

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

$$\frac{(x+\epsilon)^3 - x^3}{\epsilon} \approx 3x^2\,. \tag{8.1}$$

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

$$\frac{(X+\epsilon E)^3 - X^3}{\epsilon} \approx X^2E + XEX + EX^2\,. \tag{8.2}$$

This is the matrix version of (8.1). 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:

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

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, $dY(X)$ 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.

Example 3: $y = x^T A x$

Again y is scalar. We have $\mathrm{d}y = \mathrm{d}x^T Ax + x^T A\mathrm{d}x$. 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 $-X^{-1}EX^{-1}$.

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.

Notation:

Let

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

We denote a vector of differentials by

$$\mathrm{d}x = \begin{pmatrix} \mathrm{d}x_1 \\ \vdots \\ \mathrm{d}x_n \end{pmatrix} .$$

Similarly if $A \in \mathbb{R}^{n,n}$, then we have the matrix of differentials

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

We are allowed to take any linear function of differentials with coefficients that depend on x. Thus we do not have any problem writing $2\mathrm{d}x_1 + 5\mathrm{d}x_2$ or $x_1^3 \mathrm{d}x_1 + x_2 \mathrm{d}x_2$. We also have no problem having differentials as elements of vectors and matrices. Indeed we already have when we used the notation $\mathrm{d}x, \mathrm{d}X, \mathrm{d}Y, \mathrm{d}Q$ in the previous examples.

Example 6: If y is a scalar function of $x_1, x_2, \ldots, x_n$ then we can write

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

We can always write $\mathrm{d}Y$ as a matrix of differentials involving the elements of $\mathrm{d}X$ and often the elements of X as well.

Example 7: Let $X = \begin{bmatrix} p & q \\ r & s \end{bmatrix}$ and $Y = X^2$ then

$$\mathrm{d}Y = \begin{bmatrix} 2\,p\mathrm{d}p + q\mathrm{d}r + r\mathrm{d}q & q\mathrm{d}p + (p+s)\,\mathrm{d}q + q\mathrm{d}s \\ r\mathrm{d}p + (p+s)\,\mathrm{d}r + r\mathrm{d}s & q\mathrm{d}r + r\mathrm{d}q + 2\,s\mathrm{d}s \end{bmatrix}. \tag{8.3}$$

8.2 Matrix Jacobians (getting started)

8.2.1 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 small volume of size $|\det J|\epsilon$ 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 $\int p(y(x))\left|\frac{\partial y_i}{\partial x_j}\right|\mathrm{d}x$. For many people this becomes a matter of notation, but one should understand intuitively that the Jacobian tells you how the little 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" change of variables or "many" but not one unique change of variables.

We note that in some literature "the Jacobian" sometimes denotes the Jacobian matrix. Other times, the matrix is what people refer to simply as "the derivative." We will stick to the convention that the Jacobian is the absolute determinant of the Jacobian matrix which we often denote J.

8.2.2 Simple Examples (n=2)

In 9.1 (later) we will show a fancy formalism of using wedge products for deriving Jacobians. Now we get our feet wet with some simple 2×2 examples.

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.

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

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

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 (8.3). 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

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

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 $Y = X^{-1}$

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

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 the determinant is $\det A^2 = (\det A)^2$. 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 $p \neq 0$.

The Jacobian matrix is itself lower triangular

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

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

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

The three independent elements in D and M may be thought of as functions p,r and s of X. 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 $\theta = \arctan(p/s)$ 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

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

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 8.3.1.

```
%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=[ atan(p/s) sqrt(p^2+s^2)];
                      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 8.1

8.3 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 anti-symmetric 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.

8.3.1 $Y = BXA^T$ and the Kronecker Product

The Kronecker Product

If $A \in \mathbb{R}^{m_1,m_2}$ and $B \in \mathbb{R}^{n_1,n_2}$, then the Kronecker product $A \otimes B \in \mathbb{R}^{m_1 n_1 \times m_2 n_2}$ is the $m_1 n_1$ by $m_2 n_2$ matrix that contains all products of elements of A with that of B:

$$A \otimes B = \begin{pmatrix} a_{11}B & \dots & a_{1m_2}B \\ \vdots & & \vdots \\ a_{m_1 1}B & \dots & a_{m_1 m_2}B \end{pmatrix}$$

There are many important properties of the Kronecker product. We recommend van Loan [278]. Probably most important property is that

$$\det(A \otimes B) = (\det A)^n (\det B)^m, \tag{8.4}$$

when A is $m \times m$ and B is $n \times n$. Also the linear operator from $\mathbb{R}^{m,n}$ to $\mathbb{R}^{m,n}$ sending X to BXA^T may be represented by $A \otimes B$, where $A \in \mathbb{R}^{n,n}, B \in \mathbb{R}^{m,m}$.

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.

It is easy to see that if Q_1 and Q_2 are orthogonal, then so is $Q = Q_1 \otimes Q_2$ because Q preserves Frobenius norms, i.e., $\|X\|_F = \|Q_2^T \times Q_1^T\|_F$.

The Symmetric Kronecker Product

If $X \in \mathbb{R}^{n,n}$ is symmetric, and $A, B \in \mathbb{R}^{n,m}$ we can consider the function $Y = \frac{1}{2}(BXA^T + AXB^T)$. We denote this operator $A \otimes_{\text{sym}} B$ and write $Y = (A \otimes_{\text{sym}} B)X$. It is a linear map on $\frac{n(n+1)}{2}$ dimensional space.

Kronecker Product Jacobian Computation

Example 1: General X

Consider the linear map $Y = BXA^T$ for $X \in \mathbb{R}^{m,n}$, where $A \in \mathbb{R}^{n,n}, B \in \mathbb{R}^{m,m}$. Assume A and B are diagonalizable, with $Au_i = \lambda_i u_i$ $(i = 1, \dots, m)$ and $Bv_i = \mu_i v_i$ $(i = 1, \dots, n)$. Let $E_{ij} = v_i u_j^T$. The mn matrices E_{ij} form a basis for $\mathbb{R}^{mn}$ and they are eigenvectors of our map since $BE_{ij}A^T = \mu_i \lambda_j E_{ij}$. The determinant is

$$\prod_{\substack{1 \le i \le n \\ 1 \le j \le m}} \mu_i \lambda_j \qquad \text{or} \qquad (\det A)^n (\det B)^m \,. \tag{8.5}$$

The assumption of diagonalizability is not important.

Example 2: Upper Triangular X

For upper triangular X we assume that $Y = BXA$, $BX \in \mathbb{R}^{n,n}$ upper triangular, $A \in \mathbb{R}^{n,n}$ lower triangular.

We order the eigenvalues so that $\lambda_i = A_{ii}$ and $\mu_j = B_{jj}$. Then $E_{ij} = y_i x_j^T$ for $i \le j$ is upper triangular since y_i and x_j are zero below the ith and above the jth component respectively. (The eigenvectors of a triangular matrix are triangular.)

We then have that

$$\det J = \prod_{i \le j} \lambda_i \mu_j = (\lambda_1^n \lambda_2^{n-1} \lambda_3^{n-2} \ldots \lambda_n)(\mu_1 \mu_2^2 \mu_3^3 \ldots \mu_n^n)\,.$$

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

Example 3: Symmetric X

For symmetric X, we might consider $Y = \frac{1}{2}(AXB + B^T XA)$ for general A and B. The Jacobian is not particularly pretty. If A and B are both symmetric and commute if $B = A^T$ the situation is nicer again.

Let $E_{ij} = \frac{1}{2}(x_i x_j^T + x_j x_i^T)$. If A and B are both symmetric, the Jacobian is $\prod_{i \le j}(\lambda_i \mu_j + \lambda_j \mu_i)$. If $B = A^T$ then the Jacobian is $\prod_{i \le j} \lambda_i \lambda_j = (\det A)^{n+1}$.

8.3.2 Jacobians of Powers and Inverses

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 we see that the eigenvalues of $J \otimes X + X^T \otimes J$ are $\lambda_i + \lambda_j$ so that $\det J = \prod_{i,j}(\lambda_i + \lambda_j)$.

If $Y = X^n$

$$\text{then} \quad \mathrm{d}Y = \sum_{k=0}^{n-1} X^k \mathrm{d}X X^{n-1-k}$$

and

$$\det J = \prod_{i>j} \left(\sum_{k=0}^{n-1} \lambda_i^k \lambda_j^{n-1-k} \right)$$

Symmetric Matrices

If $Y = X^{-1}$ then $\mathrm{d}Y = -X^{-1}\mathrm{d}XX^{-1}$ or $\mathrm{d}Y = -(X^{-T} \otimes X^{-1})\mathrm{d}X$, so that $\det J = \det(X)^{(-2n)}$. If X is symmetric, we have that $\det J = (\det X)^{-(n+1)}$ for the map $Y = X^{-1}$.

When X is of low rank, we can consider the Moore-Penrose Inverse of X. The Jacobian was calculated in [457] and [309]. There is a non-symmetric and symmetric version.

Exercise. *Compute the differentials and Jacobians for*

$$Y = X^{1/2} \text{ and } Y = X^{-1/2}$$

8.4 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 is expanded in Section 35.1.

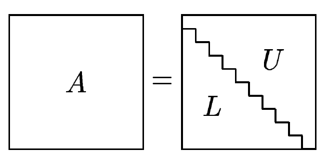
					parameter count
Gaussian Elimination:	$A =$	L	$\cdot$ U		$n(n-1)/2 + n(n+1)/2$
		$\uparrow$ unit lower triangular	$\uparrow$ upper triangular		
Gram-Schmidt:	$A =$	Q	$\cdot$ R		$n(n-1)/2 + n(n+1)/2$
		$\uparrow$ Orthogonal	$\uparrow$ upper triangular		
Eigenvalue Decomposition:	$A =$	X	$\cdot$ Λ	$\cdot$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

8.4.1 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 8.1. *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

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

or $\mathrm{d}A = (U^T \otimes L)((U^T \otimes I)^{-1}\mathrm{d}U + (I \otimes L)^{-1}\mathrm{d}L)$.

Using the triangular case of Section 8.3, the Jacobian of $(U^T \otimes I)^{-1}$ is $u_{11}^{-1}u_{22}^{-2}\dots u_{nn}^{-n}$ and that of $U^T \otimes L$ is $(\det U)^n = u_{11}^n u_{22}^n \dots u_{nn}^n$ putting the two together we obtain 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

$$\frac{\partial A_{ij}}{\partial M_{ij}} = \begin{cases} 1 & \text{if} \quad i \le j \\ u_{jj} & \text{if} \quad i > j \end{cases}\,. \tag{8.6}$$

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 (8.6). 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*

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

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.*

8.5 Jacobians for Spherical Coordinates

8.5.1 Spherical coordinates

The Jacobian matrix for the transformation from spherical coordinates to 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

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

Therefore the non-zero structure of the Jacobian matrix is represented by the pattern above.

Lower Hessenberg matrices that are non-zero only in the lower triangular part and on the superdiagonal are referred to as lower element 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:

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

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 $J_{x\to\text{spherical}} = J^{-1}_{\text{spherical}\to y}J_{x\to y}$.

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

$$\frac{2^n r^n(\sin\theta_1)^{n-1}(\sin\theta_2)^{n-2}(\sin\theta_{n-1})x_1\cdots x_{n-1}}{2^n\, x_1\cdots x_n} = r^{n-1}(\sin\theta_1)^{n-2}\cdots(\sin\theta_{n-2}).$$

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