Jacobian of Matrix Decomposition

Sungwoo Jeong

In this project we will discuss Jacobians of various matrix decompositions and their determinants. The notebook provides the validation for the theoretical determinants.

1 Independent variables for each types of matrices

First, it is very important to recognize the number of independent variables for each matrices. Other than orthogonal group it is obvious.
1) General $\mathbb{R}^{n\times n} \longrightarrow n^2$
2) Orthogonal group $O(n) \longrightarrow \frac{n(n-1)}{2}$
3) Symmetric $\longrightarrow \frac{n(n+1)}{2}$
4) Anti-Symmetric $\longrightarrow \frac{n(n-1)}{2}$
5) Triangular $\longrightarrow \frac{n(n+1)}{2}$
6) Diagonal $\longrightarrow n$

Except for orthogonal group and Anti-symmetric matrices, we will use a differentials of each entries to represent the differential of the whole matrix. For Anti-symmetric matrices, we will use the differentials of strictly lower triangular parts to represent the differential of the matrix. For orthogonal group, there are 3 ways to parametrize the differentials.

· $\frac{n(n-1)}{2}$ elements of $Q^T dQ$(Anti-symmetric) $\rightarrow (Q^T dQ)^\wedge$ gives Haar measure

· $\frac{n(n-1)}{2}$ Givens rotation angles (Difficulty of using Dual numbers)

· Householder vectors with first elements = 1 (lower triangular part of Julia QR type)
For theoretical proof, we will use the first case, but in the demo notebook we will mainly use the third case because it is easily obtained with Julia function (t→tril(qrfact(t).factors,-1))

2 LU Decomposition

We can easily obtain the Jacobian using Automatic differentiation. Also theoretical determinant of Jacobian is given as

$$\det(J_{LU}) = \prod_{i=1}^{n} U_{ii}^{n-i}$$

Figure Below is the shape of J_{LU}.

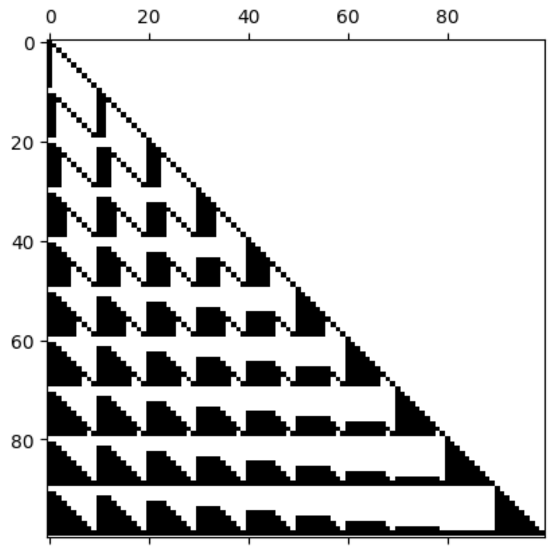

Figure 1: Lower triangular structure of Jacobian of LU Decomposition

3 Cholesky Decomposition

This is just a LU decomposition on Symmetric Positive definite matrices. In theory, we have $A_{ij} = \sum_{k=1}^{j} L_{ik}L_{jk}$ from $A = LL^T$. (Note that $i \leq j$). We observe that no L_{pq} are involved in A_{ij} if $p > i$ or $q > j$. So when we calculate the Jacobian matrix, $\frac{\partial A_{ij}}{\partial L_{pq}}$ with lexicographic order on (i, j) and (p, q) we have lower triangular matrix with diagonal elements $\frac{\partial A_{ij}}{\partial L_{ij}}$ Since we know that

$$\frac{\partial A_{ij}}{\partial L_{ij}} = \begin{cases} L_{jj} & (i > j) \\ 2L_{jj} & (i = j) \end{cases}$$

So we conclude that

$$\det J_{Chol} = \prod_{i \geq j} \frac{\partial A_{ij}}{\partial L_{ij}} = 2^n (\prod_{k=1}^{n} L_{kk})(\prod_{s=1}^{n-1} L_{ss}^{n-s}) = 2^n \prod_{i=1}^{n} L_{ii}^{n-i+1}$$

So (Lower)Triangular Jacobian matrix of LU and Cholesky leads to an element-to-element conversion from A to the factorization - which means we don't have to allocate any other matrices inside the algorithm.(i.e. Doolittle's LU decomposition or Crout's LU decomposition) We will consider a simple refinement for the note at the end of chapter 9 of textbook in next few sections.

4 QR Decomposition

We have the determinant of the Jacobian for QR decomposition,

$$\det(J_{QR}) = \prod_i r_{ii}^{n-i} (Q^T dQ)^{\wedge}$$

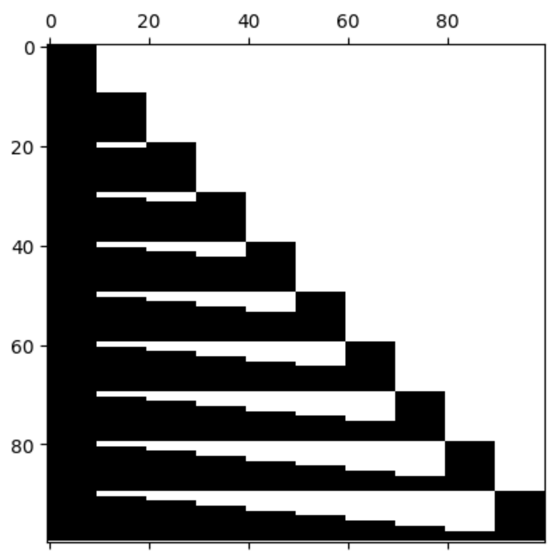

Figure 2: Lower block triangular structure of Jacobian of QR Decomposition

Even though QR decomposition has finite and (vectorwise) one-to-one conversion algorithm, it has a Jacobian matrix which is not triangular, but rather block triangular. So existence of finite algorithm for matrix decomposition is not equivalent to lower triangular Jacobian. There exists a finite algorithm provided that conversion of each diagonal block has its own finite algorithm. (i.e. Householder vector v is obtained by one-step algorithm, $v = sign(x_1)\|x\|e_1 + x$)

5 Schur Decomposition

Since the existence of finite algorithm of SVD or Eigenvalue related decompositions are equivalent to the solvability of polynomials, we don't have finite algorithms, and we expect fully dense Jacobians, which we can easily check in the demo.
In theory, differentiating $A = QTQ^T$ and multiplying Q^T, Q on each sides we get,

$$Q^T dAQ = (Q^T dQ)T + dT - T(Q^T dQ)$$

Here we will use kronecker products,

$$Q^T dAQ = (T^T \otimes_{\text{anti-sym}} I)(Q^T dQ) + dT - (I \otimes_{\text{anti-sym}} T)(Q^T dQ)$$

Making into matrix form,

$$Q^T dAQ = \begin{pmatrix} (T^T \otimes_{\text{anti-sym}} I) - (I \otimes_{\text{anti-sym}} T) & 0 \\ 0 & I \end{pmatrix} \begin{pmatrix} Q^T dQ \\ dT \end{pmatrix}$$

So we now know that $(dA)^\wedge = J(dT)^\wedge (Q^T dQ)^\wedge$ with

$$\det(J) = \det((T^T \otimes_{\text{anti-sym}} I) - (I \otimes_{\text{anti-sym}} T))$$

First, $(T^T \otimes I) - (I \otimes T)$ is block lower triangular with diagonal blocks $T_{ii}I - T$ (upper triangular). On Kronecker product operation on anti-symmetric matrices, we can only think of strictly lower triangular part of original matrices, so if we only extract indices from strictly lower triangular part from full kronecker product, we recover $\otimes_{\text{anti-sym}}$

In Julia notation,

$$(A \otimes B)[\mathcal{I}, \mathcal{I}] = A \otimes_{\text{anti-sym}} B, \qquad \mathcal{I} = \{\text{Strictly lower triangular linear indices}\}$$

Now we know that it is still block diagonal with diagonal blocks upper triangular, so we can just mutliply all diagonal entries to obtain the determinant of Jacobian. Recognizing that diagonal entries of $(T^T \otimes I) - (I \otimes T)$ are $T_{ii} - T_{jj}$ and we only count $i < j$ (strictly lower triangular) cases, to finally obtain

$$\det(J_{Schur}) = \prod_{i<j}(T_{ii} - T_{jj}) = \prod_{i<j}(\lambda_i - \lambda_j)$$

6 Singular Value Decomposition

In theory, differentiating $A = U\Sigma V^T$ and multiplying U^T, V on each sides we get,

$$U^T dAV = (U^T dU)\Sigma + d\Sigma - \Sigma(V^T dV)$$

Off-diagonal entries of RHS will be

$$\begin{cases} -\sigma_j u_j^T du_i - \sigma_i v_j^T dv_i & (i < j) \\ \sigma_j u_i^T du_j - \sigma_i v_i^T dv_j & (i > j) \end{cases}$$

Wedging gives

$$(-\sigma_j u_j^T du_i - \sigma_i v_j^T dv_i) \wedge (\sigma_j u_i^T du_j - \sigma_i v_i^T dv_j) = (\sigma_i^2 - \sigma_j^2)(u_j^T du_i) \wedge (v_j^T dv_i) = (\sigma_i^2 - \sigma_j^2)(U^T dU)_{ij} \wedge (V^T dV)_{ij}$$

We know that $(U^T dAV)^\wedge = (dA)^\wedge$ so finally we obtain

$$(dA)^\wedge = \prod_{i<j} (\sigma_i^2 - \sigma_j^2)(d\Sigma)^\wedge (U^T dU)^\wedge (V^T dV)^\wedge$$

$$\det(J_{SVD}) = \prod_{i<j} (\sigma_i^2 - \sigma_j^2)$$

7 Generalized Schur Decomposition (QZ decomposition)

Generalized Schur decomposition is a decomposition of two given square matrices A, B into

$$A = QSZ^T$$

$$B = QTZ^T$$

where Q, Z are real orthogonal matrices and S, T are (semi) upper triangular matrices with real entries. We again obtain differentials,

$$dA = (dQ)SZ^T + Q(dS)Z^T + QS(dZ)^T$$

$$dB = (dQ)TZ^T + Q(dT)Z^T + QT(dZ)^T$$

Of course we multiply Q^T and Z on left and right,

$$Q^T(dA)Z = (Q^T dQ)S + dS + S(Z^T dZ)^T$$

$$Q^T(dB)Z = (Q^T dQ)T + dT + T(Z^T dZ)^T$$

Defining $d\bar{Q} = Q^T dQ$, $d\bar{Z} = (Z^T dZ)^T$, $d\bar{A} = Q^T(dA)Z$, $d\bar{B} = Q^T(dB)Z$, where we know that $(dA)^\wedge = (d\bar{A})^\wedge, (dB)^\wedge = (d\bar{B})^\wedge$,

$$d\bar{A} = (d\bar{Q})S + dS + S(d\bar{Z})$$

$$d\bar{B} = (d\bar{Q})T + dT + T(d\bar{Z})$$

Using Kronecker product notations,

$$d\bar{A} = (S^T \otimes_{\text{anti-sym}} I)(d\bar{Q}) + (I \otimes_{\text{upper-tri}} I)dS + (I \otimes_{\text{anti-sym}} S)(d\bar{Z})$$

$$d\bar{B} = (T^T \otimes_{\text{anti-sym}} I)(d\bar{Q}) + (I \otimes_{\text{upper-tri}} I)dT + (I \otimes_{\text{anti-sym}} T)(d\bar{Z})$$

We will use strictly lower triangular part of anti-symmetric matrices, so that we can explicitly make $\otimes_{\text{anti-sym}}$ by deleting columns corresponding to diagonal entries and subtracting columns of upper triangular entries on columns of corresponding symmetric parts. In Julia notation, below is a snippet for producing $(X \otimes_{\text{anti-sym}} Y)$

```
XkY = kron(X,Y) #Usual kronecker product
for i in 1:size(XkY,1) j in 1:size(XkY,2)
    if i<j # if it is upper triangular index
        XkY[:,sub2ind(X,j,i)] -= XkY[:,sub2ind(X,i,j)] # Subtract symmetric part
        XkY[:,sub2ind(X,i,j)] .= zeros(XkY[:,1])       # We don't need the columns for upper triangular parts
    end
    if i==j  XkY[:,sub2ind(X,i,j)] .= zeros(XkY[:,1]) end # Also kill diagonal parts
end
SLT = vec(tril(ones(X)),-1) .==1
XkY_antisym = XkY[:,SLT] # Take colums of strictly lower triangular
```

Figure 3: Kronecker product acting on Anti-symmetric matrix

So making this into a big matrix form,

$$\begin{pmatrix} d\bar{A} \\ d\bar{B} \end{pmatrix} = \begin{pmatrix} S^T \otimes_{\text{anti-sym}} I & I \otimes_{\text{upper-tri}} I & I \otimes_{\text{anti-sym}} S & 0 \\ T^T \otimes_{\text{anti-sym}} I & 0 & I \otimes_{\text{anti-sym}} T & I \otimes_{\text{upper-tri}} I \end{pmatrix} \begin{pmatrix} d\bar{Q} \\ dS \\ d\bar{Z} \\ dT \end{pmatrix}$$

Carefully examining the big matrix(Jacobian) here, we realized that by using a column-major order on $d\bar{Q} + dS$ and $d\bar{Z} + dT$ we can nicely reorder the Jacobian matrix as block matrix of block lower triangular matrices. Below figure shows the effect of reordering the Jacobian matrix.

Now since we have block lower triangular matrices, note that the inverses of each blocks are also lower block triangular matrices, with each triangular blocks keep their shape. So we can only think about the diagonal parts when we think about the determinants, since determinants only involve diagonal terms.

So the determinant of Jacobian is

$$\det(J_{QZ}) = \det(J_{11}) \det(J_{22} - J_{21} J_{11}^{-1} J_{12})$$

where

$$J = \begin{pmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{pmatrix}$$

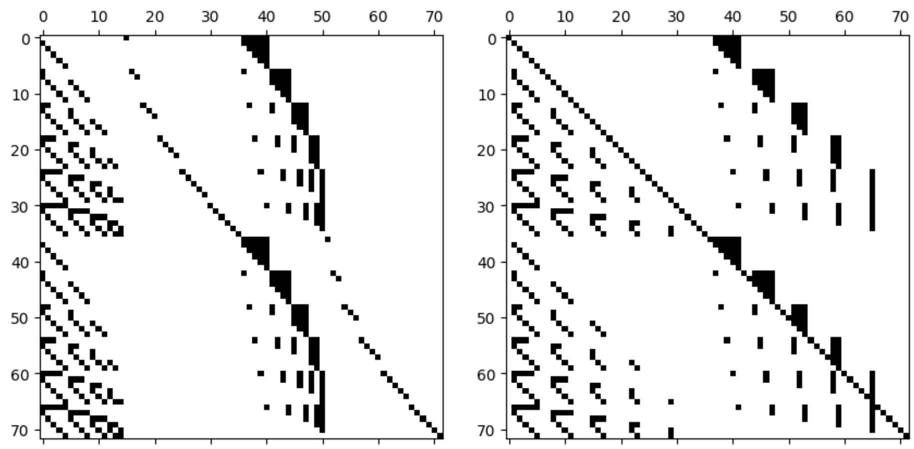

Figure 4: Jacobian of Generalized Schur without/with reordering

(i,j)th element on diagonals of each J_{pq} are as belows :

$$\begin{cases} S_{jj} & \text{on strictly lower triangular (i,j), } J_{11} \\ S_{ii} & \text{on strictly lower triangular (i,j), } J_{12} \\ T_{jj} & \text{on strictly lower triangular (i,j), } J_{21} \\ T_{ii} & \text{on strictly lower triangular (i,j), } J_{22} \end{cases}$$

with J_{11} and J_{22} has other diagonal entries 1, J_{21}, J_{12} with 0. Finally we can deduce from above equation that,

$$\det(J_{QZ}) = \prod_{i<j} (S_{ii}T_{jj} - S_{jj}T_{ii})$$

8 Generalized Singular Value Decomposition

Generalized Singular Value decomposition is a decomposition of two given matrices A, B into

$$A = USX^T$$

$$B = VCX^T$$

where U, V are real orthogonal matrices and C, S are diagonal matrices with entries corresponding to cosine and sine of some angles Θ. Also X is a nonsingular square matrix. We differentiate both sides and multiply U^T, V^T on the left, and X^{-T} on the right to obtain

$$U^T(dA)X^{-T} = (U^T dU)S + dS + S(X^{-1}dX)^T$$

$$V^T(dB)X^{-T} = (V^T dV)C + dC + C(X^{-1}dX)^T$$

Let $d\bar{A} = U^T(dA)X^{-T}$, $d\bar{B} = V^T(dB)X^{-T}$, $d\bar{U} = (U^T dU)$, $d\bar{V} = (V^T dV)$ and $d\bar{X} = (X^{-1}dX)^T$ so that $(d\bar{X})^\wedge = |X|^{-n}(dX)^\wedge$, we get

$$d\bar{A} = (d\bar{U})S + Cd\Theta + S(d\bar{X})$$

$$d\bar{B} = (d\bar{V})C - Sd\Theta + C(d\bar{X})$$

Making this into matrix form,

$$\begin{pmatrix} d\bar{A} \\ d\bar{B} \end{pmatrix} = \begin{pmatrix} S \otimes_{\text{anti-sym}} I & 0 & I \otimes_{\text{diag}} C & I \otimes S \\ 0 & C \otimes_{\text{anti-sym}} I & -I \otimes_{\text{diag}} S & I \otimes C \end{pmatrix} \begin{pmatrix} d\bar{U} \\ d\bar{V} \\ d\Theta \\ d\bar{X} \end{pmatrix}$$

Also examining the structure of the Jacobian matrix and reordering the matrix we get

$$\det(J_{GSVD}) = \prod_{i<j}(C_i^2 - C_j^2)|X|^n$$