Jacobians of Matrix Transforms

김성민

서울대학교
통계학과, 베이즈통계 연구실

2021. 10. 28

CONTENTS

CONTENTS

Material

- http://web.mit.edu/18.325/www/handouts.html
  - Edelman (2005b)
  - Edelman (2005a)
  - Edelman (2005c)
- Muirhead (2009)

CONTENTS

Jacobian of $Y = BXA^T$

- $Y = X^{-1}$

$$\Rightarrow dY = -X^{-1} dX X^{-1}$$

- By using "Kronecker Product", we can instantly write down the Jacobian.

Operator Definition

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 down as follows:

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

- By using "Kronecker Product", we can instantly write down the Jacobian.

Matrix Definition

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

- Concretely, we have that

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

Properties

- $(A \otimes B)^T = A^T \otimes B^T$
- $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$
- $det(A \otimes B) = (detA)^m (detB)^n, \ A \in \mathbb{R}^{n,n}, \ B \in \mathbb{R}^{m,m}$
- $tr(A \otimes B) = tr(A)tr(B)$
- $A \otimes B$ is orthogonal if A and B is orthogonal
- $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$
- $A \otimes B$ and $B \otimes A$ have the same eigenvalues, and transposed eigenvectors.

Linear Subspace Kronecker Products

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 $X \in \mathcal{S}$.

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

Jacobians of $(A \otimes B)_{upper}$

- Special case : A is lower triangular, B is upper.
- $(A \otimes B)_{upper} X = BXA^T$
- 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 $Bv_i = \mu_i v_i$.
- Let $M_{ij} = v_i u_j^T$ for $i \leq j$, then $BM_{ij}A^T = \mu_i \lambda_j M_{ij}$
- $det(J) = \Pi_{i \leq j} \mu_i \lambda_j$

CONTENTS

Jacobians of Linear Functions, Powers and Inverses

- $Y = X^2$

$$\begin{aligned} dY &= XdX + dXX \\ &= XdXI + IdXX \\ &= (I \otimes X + X^T \otimes I)dX \end{aligned}$$

- u_i, v_j are eignevectors of X, X^T, repectively.
- Let $E_{ij} = u_i v_j^T$

$$\begin{aligned} (I \otimes X + X^T \otimes I)E_{ij} &= XE_{ij} + E_{ij}X \\ &= \lambda_i u_i v_j^T + u_i(\lambda_j v_j)^T \\ &= (\lambda_i + \lambda_j)E_{ij} \end{aligned}$$

- So that $det(J) = \Pi_{i,j}(\lambda_i + \lambda_j)$

Jacobians of Matrix Factorizations

- Gaussian Elimination : $A = LU$ (L : unit lower triangular, U : upper triangular)
- Gram-Schmidt : $A = QR$ (Q : Orthogonal, R : upper triangular)
- Eigenvalue Decomposition : $A = X\Lambda X^T$ (X : eigenvectors, Λ : eigenvalues)

Jacobian of Gauss Elimination

- The mapping $dU \to dUU^{-1}$ only affects the upper triangular part.
- The mapping $dL \to L^{-1}dL$ only affects the lower triangular part.

$$\begin{aligned} dA &= LdU + dLU \\ &= L(dUU^{-1} + L^{-1}dL)U \\ &= (U^T \otimes L)((U^T \otimes_{upper} I)^{-1}dU + (I \otimes_{lower} L)^{-1}dL) \\ &= (U^T \otimes L)\begin{pmatrix} U^T \otimes_{upper} I & \\ & I \otimes_{lower} L \end{pmatrix}^{-1}\begin{pmatrix} dU \\ dL \end{pmatrix} \end{aligned}$$

- Jacobian determinant of $U^T \otimes L$: Πu_{ii}^n
- $(U^T \otimes I)^{-1}$: Πu_{ii}^{-i}
- $I \otimes L$: 1

$$\therefore \quad det(J) = \Pi u_{ii}^{n-i}$$

Jacobian of Gram-Schmidt

$$
\begin{aligned}
dA &= QdR + dQR \\
&= Q(dRR^{-1} + Q^T dQ)R \\
&= (R^t \otimes Q)((R^t \otimes_{upper} I)^{-1} dR + Q^T dQ)
\end{aligned}
$$

- $det(J) = (\Pi r_{ii}^{n})(\Pi r_{ii}^{-i}) = \Pi r_{ii}^{n-i}$

CONTENTS

Wedge Product

$$
\begin{aligned}
&(2dx + x^2dy + 5dw + 2dz) \wedge (ydx - xdy) \\
&= (-2x - x^2y)dx \wedge dy + 5y(dw \wedge dx) - 5x(dw \wedge dy) \\
&- 2y(dx \wedge dz) + 2x(dy \wedge dz)
\end{aligned}
$$

- $(du \wedge dv) = -(dv \wedge du)$
- $du \wedge du = 0$

Wedge Product

$$F = \begin{pmatrix} 2 & x_2 \\ x_1^2 & -x_1 \\ 5 & 0 \\ 2 & 0 \end{pmatrix}$$

- $(2dx_1 + x_1^2 dx_2 + 5dx_3 + 2dx_4) \wedge (x_2 dx_1 - x_1 dx_2) = \wedge_{i=1}^{2} (F(x)^T dx)_i$
- $= \sum_{i_1 < i_2} det(F[(i_1, i_2), :]) dx_{i_1} dx_{i_2}$

Wedge Product

- We use the notation

$$(F(x)^T dx)^\wedge \equiv \wedge_{i=1}^{p} (F(x)^T dx)_i$$

- We extend $()^\wedge$ notation from vectors to matrices of differentials.

$$(dM)^\wedge = \wedge_{i,j} dM_{ij}$$

Wedge Product to Square Matrix

- $A = lower(M) - lower(M)^T$: Anti-symmetric
- $R = upper(M) + lower(M)^T$: Upper Triangular
- $M = A + R$
- $(dM)^\wedge = (lower(dA) + dR)^\wedge = (dA)^\wedge (dR)^\wedge$

CONTENTS

Integration Using Differential Forms

- $y = y(x)$ is som function from $\mathbb{R}^n$ to $\mathbb{R}^n$

$$\int_{y(S)} f(y) dy_1 \wedge \ldots \wedge dy_n = \int_S f(y(x)) dx_1 \wedge \ldots \wedge dx_n$$

- Integration of surfaces of sphere $\Rightarrow$ $dx_1 \wedge \ldots \wedge dx_n = 0$

Plucker Coordinates

Definition 1.

$Pl(F)$ is the vector of $p \times p$ subdeterminants of F ($F \in \mathbb{R}^{n,p}$).

$$F = (f_{ij})_{\substack{i \leq n \\ j \leq p}} \xrightarrow{Pl} \left(det(f_{ij})_{\substack{i=i_1,\ldots,i_n \\ j=1,\ldots,p}} \right)_{i_1 < \ldots < i_p}$$

Definition 2.

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

Plucker Coordinates

Theorem 1.

$vol(F) = \Pi_{i=1}^{p} \sigma_i = det(F^T F)^{1/2} = \Pi_{i=1}^{p} r_{ii} = ||Pl(F)||$, 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.

Corollary 3.

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

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

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

$$\int f(x)d(surface) \approx \sum f(x^{(i)})Pl(F^{(i)})^T Pl(V^{(i)})$$

$$= \sum f(x^{(i)})det((F^{(i)})^T V^{(i)})$$

- Mathematicians write the continuous limit of the above equation as

$$\int f(x)d(surface) = \int f(x)(F^T dx)^{\wedge}$$

Volume Measurement

- Notice that $(F(x)^T dx)^\wedge$ formally compute $PI(F(x))$. Indeed

$$(F(x)^T dx)^\wedge = PI(F(x))^T \begin{pmatrix} \vdots \\ dx_{i1} \wedge \ldots \wedge dx_{ip} \\ \vdots \end{pmatrix}_{i_1 < \ldots < i_p}$$

Overview of special surfaces

We are very interested in the following three mathematical objects:

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

Example

- Integration over a sphere given q with $||q|| = 1$, let $H(q)$ be any $n \times n$ orthogonal matrix with first column q. (One way to construct $H(q)$ is $I - 2\frac{vv^T}{v^T v}$, where $v = e_1 - q$)
- The sphere is an $n-1$ dimensional surface in n dimensional space.
- Integration over the sphere is then,

$$\int_{\substack{x \in \mathbb{R}^n \\ ||x||=1}} f(x)dS = \int_{\substack{x \in \mathbb{R}^n \\ ||x||=1}} f(x) \wedge_{i=2}^n (H^T dx)_i$$

The Sphere

- $x = qr$, where $r = ||x||$ and $q = x/||x||$
- Then

$$Hdx = e_1 dr + Hdqr = \begin{pmatrix} dr \\ r(Hdq)_2 \\ \vdots \\ r(Hdq)_n \end{pmatrix}$$

- Thus

$$(dx)^\wedge = (Hdx)^\wedge = r^{n-1} dr \wedge_{i=2}^{n} (Hdq)_i$$

Surface Area of Sphere Computation

- We directly use the formula $(dx)^\wedge = r^{n-1}dr(Hdq)^\wedge$

$$(2\pi)^{n/2} = \int_{x\in\mathbb{R}^n} e^{-\frac{1}{2}||x||^2}dx = \int_{r=0}^{\infty} r^{n-1}e^{-\frac{1}{2}r^2}dr \int (Hdq)^\wedge$$
$$= 2^{\frac{n-2}{2}}\Gamma(\frac{n}{2}) \int (Hdq)^\wedge$$
$$\therefore \quad \int (Hdq)^\wedge = \frac{2\pi^{n/2}}{\Gamma(\frac{n}{2})} = A_n$$

- A_n is the surface of the sphere of radius 1.

The Stiefel Manifold

- $A = QR$, where $Q \in \mathbb{R}^{n,p}$ such that $Q^T Q = I_p$
- Consider the orthogonal matrix $H = [Q, h_{p+1}, \ldots, h_n]$, where $H \in \mathbb{R}^{n,n}$
- $dA = QdR + dQR$
- $H^T dA = H^T QdR + H^T dQR$
- $H^T QdR$: n by p upper triangular matrix
- $H^T dQR$: (rectangularly) antisymmetric

Haar Measure and Volume of the Stiefel Manifold

- The natural volume element on the Stiefel Manifold is as follows:

$$(H^T dQ) = \wedge_{j=1}^{p} \wedge_{i=j+1}^{n} h_i^T dh_j$$

- We may define

$$\mu(S) = \int_S (H^T dQ).$$

- This measure μ is known as Haar measure when $p = n$

References I

Edelman, A. (2005a). 18.325: Finite random matrix theory: Jacobians of matrix transforms (with wedge products).

Edelman, A. (2005b). 18.325: Finite random matrix theory: Jacobians of matrix transforms (without wedge products).

Edelman, A. (2005c). 18.325: Finite random matrix theory volumes and integration.

Muirhead, R. J. (2009). *Aspects of multivariate statistical theory*, Vol. 197, John Wiley & Sons.