KTH ROYAL INSTITUTE OF TECHNOLOGY

Lecture 9

Magnus Jansson, May 23, 2016

Outline

- Matrix equations
- The Kronecker product
- Vectorization
- The Khatri-Rao product
- Differentiation

Parts of Chapter 4 in "Topics in Matrix Analysis," by R. A. Horn and C. R. Johnson + additional material, see references on the last slide.

Matrix Equations

Examples:

$$XA + A^*X = B$$
$$AX = B$$
$$AX = XA$$
$$AXB + CXD = E$$
$$AX + YB = C$$
$$X^2 = A$$
$$X^TAX + B^TX + X^TB = C$$

The Kronecker product

Let $A = [a_{ij}] \in M_{m,n}$ and $B = [b_{ij}] \in M_{p,q}$. The Kronecker product of A and B is defined as

$$A \otimes B \equiv \begin{bmatrix} a_{11}B & a_{12}B & \dots & a_{1n}B \\ a_{21}B & a_{22}B & \dots & a_{2n}B \\ \vdots & & \ddots & \vdots \\ a_{m1}B & a_{m2}B & \dots & a_{mn}B \end{bmatrix} \in M_{mp,nq}$$

Notice that $A \otimes B \neq B \otimes A$ in general.

The kth Kronecker power is defined as

$$A^{\otimes k} \equiv A \otimes A^{\otimes(k-1)}; \qquad A^{\otimes 1} \equiv A$$

Kronecker product: Some properties

For matrices A, B, C, D (of suitable dimensions) and scalar α we have:

$$\begin{aligned}
(\alpha A) \otimes B &= A \otimes (\alpha B) = \alpha(A \otimes B) \\
(A \otimes B)^T &= A^T \otimes B^T \\
(A \otimes B)^* &= A^* \otimes B^* \\
(A \otimes B) \otimes C &= A \otimes (B \otimes C) \\
(A + B) \otimes C &= (A \otimes C) + (B \otimes C) \\
A \otimes (B + C) &= (A \otimes B) + (A \otimes C) \\
(A \otimes B)(C \otimes D) &= AC \otimes BD \\
(A \otimes B)^{-1} &= A^{-1} \otimes B^{-1}
\end{aligned}$$

if the inverses exist.

The vec operator

Let $A = [a_{ij}] \in M_{m,n}$. Then the vector $\mathrm{vec}(A) \in \mathbf{C}^{mn}$ is defined as

$$\mathrm{vec}(A) = [a_{11}\ a_{21}\ \dots\ a_{m1}\ a_{12} \dots\ a_{m2}\ \dots\ a_{1n}\ \dots a_{mn}]^T$$

The vec operator: Properties

- It is simple to verify that

$$\begin{aligned}
\mathrm{tr}(AB) &= \mathrm{vec}^T(A^T)\,\mathrm{vec}(B) \\
&= \mathrm{vec}^T(B^T)\,\mathrm{vec}(A) \\
&= \mathrm{vec}^*(A^*)\,\mathrm{vec}(B)
\end{aligned}$$

The vec operator: Properties

- It is simple to verify that

$$\begin{aligned}
\mathrm{tr}(AB) &= \mathrm{vec}^T(A^T)\,\mathrm{vec}(B) \\
&= \mathrm{vec}^T(B^T)\,\mathrm{vec}(A) \\
&= \mathrm{vec}^*(A^*)\,\mathrm{vec}(B)
\end{aligned}$$

- A *very* useful result is

$$\mathrm{vec}(ABC) = (C^T \otimes A)\,\mathrm{vec}(B)$$

for any matrices A, B, C of appropriate dimensions.

Matrix equations, cont'd

$$\begin{aligned} XA + A^*X = B &\quad\Leftrightarrow\quad [(A^T \otimes I) + (I \otimes A^*)]\,\mathrm{vec}(X) = \mathrm{vec}(B) \\ AX = B &\quad\Leftrightarrow\quad (I \otimes A)\,\mathrm{vec}(X) = \mathrm{vec}(B) \\ AX = XA &\quad\Leftrightarrow\quad [(I \otimes A) - (A^T \otimes I)]\,\mathrm{vec}(X) = 0 \\ AXB + CXD = E &\quad\Leftrightarrow\quad [(B^T \otimes A) + (D^T \otimes C)]\,\mathrm{vec}(X) = \mathrm{vec}(E) \end{aligned}$$

Matrix equations, cont'd

More generally:

Lemma: Let $T : M_{m,n} \to M_{p,q}$ be a given linear transformation. There exists a unique matrix $K(T) \in M_{pq,mn}$ such that

$$\mathrm{vec}(T(X)) = K(T)\,\mathrm{vec}(X)$$

for all $X \in M_{m,n}$.

Kronecker products: Further properties

Thm: Let $A \in M_m$ and $B \in M_n$. If (λ, x) is an eigenvalue/eigenvector pair of A and similarly (μ, y) an eigenvalue/vector pair of B, then $\lambda\mu$ is an eigenvalue of $A \otimes B$ with the corresponding eigenvector $x \otimes y$.

Furthermore, every eigenvalue arises in this way; that is, if $\sigma(A) = \{\lambda_1, \dots \lambda_m\}$ and $\sigma(B) = \{\mu_1, \dots \mu_n\}$, then

$$\sigma(A \otimes B) = \{\lambda_i \mu_j \;:\; i = 1, 2, \dots m;\; j = 1, 2, \dots, n\}$$

Notice also that $\sigma(A \otimes B) = \sigma(B \otimes A))$.

Kronecker products: Further properties cont'd

Cor: If $A \in M_m$ and $B \in M_n$ are positive (semi)definite, then $A \otimes B$ is also positive (semi)definite.

Kronecker products: Further properties cont'd

Cor: If $A \in M_m$ and $B \in M_n$ are positive (semi)definite, then $A \otimes B$ is also positive (semi)definite.

Cor: If $A \in M_m$ and $B \in M_n$, then

$$\operatorname{tr}(A \otimes B) = \operatorname{tr}(A)\operatorname{tr}(B)$$
$$\det(A \otimes B) = \det(A)^n \det(B)^m$$

SVD and the Kronecker product

Let $A \in M_{m,n}$ and $B \in M_{p,q}$ have the singular value decompositions $A = V_1 \Sigma_1 W_1^*$ and $B = V_2 \Sigma_2 W_2^*$, and assume $\operatorname{rank}(A) = r_1$ and $\operatorname{rank}(B) = r_2$. Then

$$A \otimes B = (V_1 \otimes V_2)(\Sigma_1 \otimes \Sigma_2)(W_1 \otimes W_2)^*$$

The nonzero singular values of $A \otimes B$ are the $r_1 r_2$ positive numbers $\{\sigma_i(A)\sigma_j(B) \; : \; i = 1, 2, \ldots, r_1; \; j = 1, 2, \ldots, r_2\}$ where $\sigma_i(A)$ is the ith singular value of A etc.. Hence, $\operatorname{rank}(A \otimes B) = \operatorname{rank}(B \otimes A) = r_1 r_2$.

Permutation equivalences

A trivial but sometimes useful observation is that

$$\operatorname{vec}(A^T) = P(m,n)\operatorname{vec}(A) \quad \forall \; A \in M_{m,n}$$

where $P(m,n) \in M_{mn}$ is a permutation matrix that only depends on the dimensions m and n $(P(m,n) = P^T(n,m) = P^{-1}(n,m))$.

Permutation equivalences

A trivial but sometimes useful observation is that

$$\operatorname{vec}(A^T) = P(m,n)\operatorname{vec}(A) \quad \forall \; A \in M_{m,n}$$

where $P(m,n) \in M_{mn}$ is a permutation matrix that only depends on the dimensions m and n $(P(m,n) = P^T(n,m) = P^{-1}(n,m))$.

From this it also follows that

$$B \otimes A = P^T(m,p)(A \otimes B)P(n,q)$$

for all $A \in M_{m,n}$ and $B \in M_{p,q}$.

Lyapunov equations and the Kronecker sum

Consider the matrix equation

$$AX + XB = C; \qquad A \in M_n,\ B \in M_m,\ C, X \in M_{n,m}$$

or in Kronecker form

$$[(I_m \otimes A) + (B^T \otimes I_n)]\,\text{vec}(X) = \text{vec}(C)$$

Lyapunov equations and the Kronecker sum

Consider the matrix equation

$$AX + XB = C; \qquad A \in M_n,\ B \in M_m,\ C, X \in M_{n,m}$$

or in Kronecker form

$$[(I_m \otimes A) + (B^T \otimes I_n)]\,\text{vec}(X) = \text{vec}(C)$$

Def: The matrix

$$(I_m \otimes A) + (B \otimes I_n)$$

is called the Kronecker sum of A and B.

Kronecker sum, cont'd

Thm: Let $A \in M_n$ and $B \in M_m$. If (λ, x) is an eigenvalue/eigenvector pair of A and similarly (μ, y) an eigenvalue/vector pair of B, then $\lambda + \mu$ is an eigenvalue of the Kronecker sum

$$(I_m \otimes A) + (B \otimes I_n)$$

with the corresponding eigenvector $y \otimes x$. Every eigenvalue of the Kronecker sum arises in this way.

Notice also that $I \otimes B$ and $A \otimes I$ commute.

Lyapunov equation, cont'd

Returning to the Lyapunov type equation

$$AX + XB = C \qquad \Leftrightarrow \qquad [(I_m \otimes A) + (B^T \otimes I_n)]\,\text{vec}(X) = \text{vec}(C)$$

According to the previous result, this equation has a unique solution X if and only if $\lambda_i(A) + \mu_j(B) \neq 0$ for all i, j or equivalently $\sigma(A) \cap \sigma(-B) = 0$ (empty set).

Lyapunov equation, cont'd

Returning to the Lyapunov type equation

$$AX+XB = C \quad \Leftrightarrow \quad [(I_m \otimes A)+(B^T \otimes I_n)] \operatorname{vec}(X) = \operatorname{vec}(C)$$

According to the previous result, this equation has a unique solution X if and only if $\lambda_i(A) + \mu_j(B) \neq 0$ for all i, j or equivalently $\sigma(A) \cap \sigma(-B) = 0$ (empty set).

In the complex valued case we had the Lyapunov equation

$$XA + A^* X = C$$

It has a unique solution X if and only if $\sigma(-A^*) \cap \sigma(A) = 0$ or $\overline{\sigma(-A)} \cap \sigma(A) = 0$. This condition is certainly satisfied when A is stable (positive or negative).

The Khatri-Rao product

The Khatri-Rao product of $A \in M_{m,n}$ and $B \in M_{p,n}$ is defined as (the symbol may differ)

$$A \odot B = [a_1 \otimes b_1 \;\; a_2 \otimes b_2 \;\; \ldots a_n \otimes b_n\ ]$$

where a_i, b_i denote the ith column in A and B, respectively.

It is useful, for example, in matrix equations in which diagonal matrices are involved. Let A, B, C be matrices of appropriate dimensions and let B be a diagonal matrix. Then

$$\operatorname{vec}(ABC) = (C^T \otimes A) \operatorname{vec}(B) = (C^T \odot A) \operatorname{diag}(B)$$

where $\operatorname{diag}(B)$ denotes the column vector of the diagonal elements of B.

The Khatri-Rao product

The Khatri-Rao product of $A \in M_{m,n}$ and $B \in M_{p,n}$ is defined as (the symbol may differ)

$$A \odot B = [a_1 \otimes b_1 \;\; a_2 \otimes b_2 \;\; \ldots a_n \otimes b_n\ ]$$

where a_i, b_i denote the ith column in A and B, respectively.

17 / 31

The Hadamard product

Let $A = [a_{ij}] \in M_{m,n}$ and $B = [b_{ij}] \in M_{m,n}$. The Hadamard or Schur product of A and B is defined as

$$A \circ B \equiv [a_{ij} b_{ij}] \in M_{m,n}$$

More information and properties in Chapter 5 in "Topics in Matrix Analysis."

Derivatives: Some definitions

The derivative of a matrix $A(t) = [a_{ij}(t)]$ that depends on a (real) scalar t is defined as the matrix

$$\frac{dA(t)}{dt} = [\frac{da_{ij}(t)}{dt}]$$

Derivatives: Some definitions

We can also define the derivative of a vector $y \in \mathbf{R}^m$ with respect to a vector $x \in \mathbf{R}^n$ as follows

$$\frac{\partial y}{\partial x} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_2}{\partial x_1} & \cdots & \frac{\partial y_m}{\partial x_1} \\ \frac{\partial y_1}{\partial x_2} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_m}{\partial x_2} \\ \vdots & & \vdots & \\ \frac{\partial y_1}{\partial x_n} & \frac{\partial y_2}{\partial x_n} & \cdots & \frac{\partial y_m}{\partial x_n} \end{bmatrix}$$

(Please notice that other definitions are often used!)

Derivatives, cont'd

Clearly, if y is a scalar we get

$$\frac{\partial y}{\partial x} = \begin{bmatrix} \frac{\partial y}{\partial x_1} \\ \frac{\partial y}{\partial x_2} \\ \vdots \\ \frac{\partial y}{\partial x_n} \end{bmatrix}$$

or when x is a scalar

$$\frac{\partial y}{\partial x} = [\frac{\partial y_1}{\partial x} \quad \frac{\partial y_2}{\partial x} \quad \cdots \quad \frac{\partial y_m}{\partial x}]$$

Derivatives; simple results

Let $x \in \mathbf{R}^n$ and A a matrix independent of x.

$$\frac{\partial Ax}{\partial x} = A^T$$

$$\frac{\partial x^T A}{\partial x} = A$$

$$\frac{\partial x^T x}{\partial x} = 2x$$

$$\frac{\partial x^T Ax}{\partial x} = (A + A^T)x \quad (= 2Ax \text{ if } A \text{ is symmetric})$$

$$\frac{\partial}{\partial x}\left[\frac{\partial x^T Ax}{\partial x}\right] = (A + A^T) \quad (= 2A \text{ if } A \text{ is symmetric})$$

Chain rule for vectors

Assume $z[y(x)]$ where x, y, z are real vectors, then:

$$\frac{\partial z}{\partial x} = \frac{\partial y}{\partial x}\frac{\partial z}{\partial y}$$

Derivative of scalars with respect to a matrix

Let $X = [x_{ij}] \in M_{m,n}(\mathbf{R})$ and let $y = f(X)$ be a real valued scalar function of X.

Then we define

$$\frac{\partial y}{\partial X} = \begin{bmatrix} \frac{\partial y}{\partial x_{11}} & \frac{\partial y}{\partial x_{12}} & \cdots & \frac{\partial y}{\partial x_{1n}} \\ \frac{\partial y}{\partial x_{21}} & \frac{\partial y}{\partial x_{22}} & \cdots & \frac{\partial y}{\partial x_{2n}} \\ \vdots & & \vdots & \\ \frac{\partial y}{\partial x_{m1}} & \frac{\partial y}{\partial x_{m2}} & \cdots & \frac{\partial y}{\partial x_{mn}} \end{bmatrix} = [\frac{\partial y}{\partial x_{ij}}] = \sum_{i,j} E_{ij}\frac{\partial y}{\partial x_{ij}}$$

where $E_{ij} \in M_{m,n}$ is an elementary matrix with a 1 in the ijth position and zeros elsewhere.

Derivative of trace

Let $X \in M_n(\mathbf{R})$:

$$\frac{\partial\,\mathrm{tr}(X)}{\partial X} = I$$

Derivative of determinant

Let Y, X be square matrices:

$$\frac{\partial \det(Y(X))}{\partial X} = ?$$

Derivation: Notice first that

$$\det(Y) = \sum_r y_{rs}c_{rs}$$

where c_{rs} is the cofactor of y_{rs}.

$$\frac{\partial \det(Y(X))}{\partial x_{ij}} = \sum_{r,s}\frac{\partial \det(Y(X))}{\partial y_{rs}}\frac{\partial y_{rs}}{\partial x_{ij}}$$

$$= \sum_{r,s} c_{rs}\frac{\partial y_{rs}}{\partial x_{ij}} = \mathrm{tr}(\mathrm{Adj}(Y)\frac{\partial Y}{\partial x_{ij}})$$

Derivative of log det

Let $Y(X)$ be positive definite and recall $Y^{-1} = \text{Adj}(Y)/\det(Y)$.

From the above, we get

$$\begin{aligned}\frac{\partial \log\det(Y(X))}{\partial x_{ij}} &= \frac{1}{\det(Y(X))}\frac{\partial \det(Y(X))}{\partial x_{ij}} \\ &= \frac{1}{\det(Y)}\operatorname{tr}\left(\text{Adj}(Y)\frac{\partial Y}{\partial x_{ij}}\right) \\ &= \operatorname{tr}\left(Y^{-1}\frac{\partial Y}{\partial x_{ij}}\right)\end{aligned}$$

Other simple results for matrix expressions wrt elements:

$$\frac{\partial AXB}{\partial x_{ij}} = AE_{ij}B$$

$$\frac{\partial AX^{-1}B}{\partial x_{ij}} = -AX^{-1}E_{ij}X^{-1}B$$

Derivatives of matrices with respect to matrices

Some different possibilities:

$$\frac{\partial Y}{\partial X} = \text{a partitioned matrix with } ij\text{th block } \frac{\partial Y}{\partial x_{ij}}$$

$$\frac{\partial Y}{\partial X} = \text{a partitioned matrix with } ij\text{th block } \frac{\partial y_{ij}}{\partial X}$$

$$\frac{\partial Y}{\partial X} = \frac{\partial\, \text{vec}(Y)}{\partial\, \text{vec}(X)}$$

Second derivatives

The second derivative of a scalar function of a vector is a matrix called the Hessian and is defined as

$$\frac{\partial^2 f(x)}{\partial x \partial x^T} = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\ \vdots & & \vdots & \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_n \partial x_n} \end{bmatrix}$$

Some further reading

- Jan R. Magnus & Heinz Neudecker, Matrix differential calculus with applications in statistics and economics, Chichester : Wiley, 1999
- Alexander Graham, Kronecker products and matrix calculus : with applications, Chichester : Horwood, 1981 (vec, tr, Kronecker, differentiation
- J. Brewer, "Kronecker products and matrix calculus in system theory," Circuits and Systems, IEEE Transactions on, Vol.25, Iss.9, Sep 1978 Pages: 772- 781 (vec, Kronecker, Khatri-Rao)
- D. Brandwood, "A complex gradient operator and its application in adaptive array theory," IEE Proc., vol. 130, no. 1, pp. 11-16, Feb. 1983. (differentiation wrt complex parameters, signal processing/communications)
- Complex-Valued Matrix Derivatives – With Applications in Signal Processing and Communications by Are Hjørungnes, Cambridge U. Press